Proceedings of

Gesture-based Interaction Design: Communication and Cognition

2014 CHI Workshop
Toronto, Canada

April 26, 2014

Gesture Interaction Design: Communication and Cognition
CHI Workshop W12
Schedule

The workshop has two kinds of presentations: paper session and poster session

The paper sessions allow 30 minutes per paper. The author presents highlights from the paper for 20 minutes, followed by 10 minutes for discussion.

The poster sessions start with a 3 min flash talk from each author that represents a paper followed by small group discussions around a poster or laptop display of the slides. The presentations will be given to the whole group, and the small group discussions will be around posters on the walls or at separate tables.

9:00-9:30	**Introductions and Overview**
9:30-10:00	**HCI and Cognition Studies**
	Gesture in the Crossroads of HCI and Creative Cognition: Mary Lou Maher, Tim C. Clausner, Alberto Gonzalez, Kazjon Grace
10:00-10:30	**Gesture and Cognition**
	Congruent Gestures can Promote Thought: Barbara Tversky, Azadeh Jamalian, Ayelet Segal, Valeria Giardino, Seokmin Kang
10:30-11:00	**Break**
11:00-11:30	**Mid-air Gesture Interaction Design**
	Towards Learnable Gestures for Exploring Hierarchical Information Spaces at a Large Public Display: Christopher Ackad, Judy Kay, Martin Tomitsch
11:30-12:00	**Touch Gesture Interaction Design**
	A Simple Universal Gesture Scheme for User Interfaces: Stuart K. Card
12:00-12:30	**Gestures and Learning**
	Using Embodied Cognition to Teach Reading Comprehension to DLLs: Andreea Danielescu, Erin Walker, Arthur Glenberg, M. Adelaida Restrepo, Ashley Adams
12:30-1:30	**Lunch and Poster Placement**
1:30-2:30	**Design Posters**
	Objects as Agents: how ergotic and epistemic gestures could benefit gesture-based interaction: Chris Baber

	A Cognitive Perspective on Gestures, Mani[ulations, and Space in Future Multi-Device Interaction; Hans-Christian Jetter
	Design of a Portable Gesture-Controlled Information Display: Sebastian Loehmann, Doris Hausen, Benjamin Bisinger, Leonhard Mertl
	Gesture Design and Freasibility in Emergency Response Environments: Francisco Marinho Rodrigues, Teddy Seyed, Apoorve Chokshi, Frank Maurer
	Embodying Diagramming through Pen + Touch Gestures: Andrew M. Webb and Andruid Kerne
	Tangible Meets Gestural: Gesture Based Interaction with Active Tokens: Ali Mazalek, Orit Shaer, Brygg Ullmer, Miriam K. Konkel
2:30-3:00	**Learning Posters**
	Animation Killed the Video Star: Voicu Popescu, Nicoletta Adamo-Villani, Meng-Lin Wu, Suren D. Rajasekaran, Martha W. Alibali, Mitchell Nathan, Susan Wagner Cook
	Help systems for gestural interfaces and their effect on collaboration and communication: Davy Vanacken, Anastasiia Beznosyk, Karin Coninx
3:00-3:30	**Break and Poster Placement**
3:30-4:30	**Gesture Recognition Posters**
	Towards Biomechanicly-Inspired Index of Expert Drawing Gestures Complexity: Myroslav Bachynskyi
	How Do Users Interact with an Error-prone In-air Gesture Recognizer? Ahmed Sabbir Arif, Wolfgang Stuerzlinger, Euclides Jose de Mendonca Filho, Alec Gordynski
	Challenges in Gesture Recognition for Authentication Systems: Gradeigh Clark, Janne Lindqvist
	Gesture and Rapport in Mediated Interactions: Andrea Stevenson Won
	Mining Expert-driven Models for Synthesis and Classification of Affective Motion: S. Ali Etemad and Ali Arya
4:30-5:00	**Closing Discussion**

Gesture in the Crossroads of HCI and Creative Cognition

Mary Lou Maher
UNC Charlotte
m.maher@uncc.edu

Timothy C. Clausner
University of Maryland
clausner@umd.edu

Berto Gonzalez, Kazjon Grace
UNC Charlotte
{agonza32, k.grace}@uncc.edu

ABSTRACT
Investigations of gesture in HCI research seek to enable usable and intuitive gesture-based interactive technology. Investigations of gesture in cognitive science seek to understand the role of gesture in thinking, language, and learning. The crossroads remains largely open, because designing for gesture interaction largely focuses on advancing device usability, and cognitive studies of gesture largely focus on advancing explanatory theory. The intersection is not a clearly defined research area, and methods that serve one focus do not necessarily serve the other. Moreover, gesture research in HCI and cognitive science each seeks to understand how gesture affects human performance, but neither discipline can predict how to do so. We approach this crossroads by focusing on research whose results contribute to an understanding of how gesture-based interaction changes cognition. We present a research program in which we look for evidence that an increase in gesturing with a tangible user interface while thinking about word combinations increases the creativity of the results. Broader implications of this research seek to cultivate the research area and engender new theory.

Author Keywords
Gesture; embodied cognition; design cognition; experimental design

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors; Design; Measurement.

INTRODUCTION

We are conducting research in the crossroads of HCI and cognitive science by focusing on how gesture-based interaction changes cognition. We selected tangible computing for this focus because it is a technology that enables gestures with handheld interactive objects. Here the term gesture, broadly-construed, encompasses body or hand movements including interactions with technology. Gesture is a kind action, which is linked with language, but has a separate ability to carry meaning not exclusively for communication; it is a way of thinking with our hands [15, 18]. Other actions are not intrinsically linked with language. Both gesture and action have wide ranging roles in HCI research. We adopt the term gesture in a broader sense, which includes hand movements when using tangible devices, because unlike prescribed actions afforded by some gesture interfaces, tangibles can be held in hand when performing gestures, actions, or both. Tangible user interfaces (TUI) are the coupling of physical objects and digital information, and eliminates the distinction between input and output devices [12,30]. For example, Figure 1 illustrates TUIs with children using Sifteo™ cubes. TUIs offer a dramatic shift from pointing and typing to grasping, holding, and arranging.

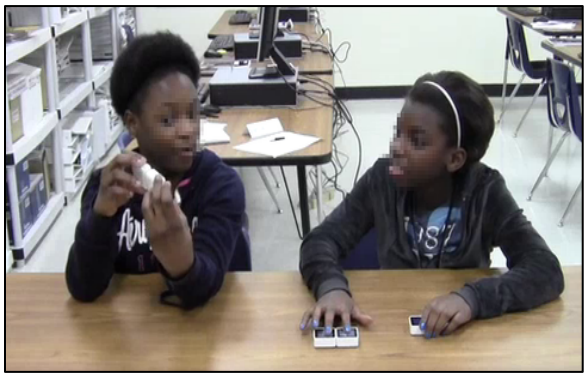


Figure 1 TUIs: Children using Sifteo™ cubes to make creative word combinations; each cube displayed one word.

HCI and design research link creativity with gesture

A special issue on the role of gesture in designing [31] provides an overview of how gesture is used in design and how TUIs affect and build on our use of gesture [17]. Marshall [24] developed a framework for research on TUI's and learning from a design research perspective. One study of the role of tangible interaction and gesture found that, for the same task, when participants used TUIs compared to keyboard and mouse, they focused on cognitive aspects of the task associated with creativity [9].

We propose that interfaces based on physical objects (i.e., TUIs) may offer more opportunities for epistemic (i.e., exploratory) actions than pragmatic (i.e., direct) actions. Epistemic actions are exploratory motor activity aimed at uncovering information that is hard to compute mentally. For example, novice chess players move pieces around to candidate positions to mentally explore possible moves and counter-moves. Epistemic actions offload some internal cognitive resources onto the external world by using physical actions. In contrast, pragmatic actions are motor activities directly aimed at a final goal. For example, experienced chess players are able to first set a mental goal then perform the minimal motor actions to reach the goal [12,14,20]. Direct mental goals typically do not require exploration. However, creative thinking is exploratory as its

Copyright is held by the author/owner(s).

Gesture-based Interaction Design: Communication and Cognition, CHI 2014 Workshop.

goals are insufficiently well-defined [28]. Using TUIs is correlated with changes in designer's thinking: Kim and Maher [19] showed an increase in epistemic actions and, through protocol analysis, observed an increase in cognitive processes typically associated with creative design.

Most studies on TUIs have been undertaken from a HCI technology viewpoint, which aims to describe fundamental technical issues and evaluate usability of prototypes. While many researchers have argued that TUIs improve spatial cognition, there has been insufficient empirical evidence to support the claim [13,21,22]. We adopt TUIs for the specific purpose of studying how bodily movement affects creative thinking, by intersecting HCI methods with empirical experimental science, and developing that crossroads approach is itself a dimension of our research.

Cognitive Science finds gesture affects cognition

Evidence from cognitive science finds actions with our hands affect thinking. Recent research on gesture and thought [1,3,4,5,6,8,9,10] has shown that gestures are an aid for thinking and not exclusively an aid for communication.

Gesturing aids learning. Goldin-Meadow et at. [16] found that children learned a strategy for solving math problems if they imitated a teacher's gestures, When instructed to imitate a teacher's gestures, Goldin-Meadow et al. [16] found that those children learned a strategy for solving math problems compared to children who did not gesture. For example, while teaching children to solve math problems such as "6 + 3 + 5 = _ + 5" the instructor made gestures indicating a grouping strategy. Placing a V-hand under the "6 + 3" then pointing to the blank indicated the strategy of grouping 6 and 3 then putting the sum in the blank. Observing and mimicking hand movements that reflected the grouping strategy led to the formation of knowledge about that strategy. The "V" gestures were metaphorical because they indicated mental groupings where no explicit groups were marked, and mental movement where no numbers physically moved across the equal sign. Goldin-Meadow and Beilock [15] summarized these and related findings as, "gesture influences thought by linking it to action", "producing gesture changes thought" and can "create new knowledge" (pp. 667-8). These effects may build on the role of gesturing in cognitive development. When children are learning to count, touching physical objects facilitates learning [1,5].

Gesturing aids creative problem solving. Kessell and Tversky [18] found that when people are solving and then explaining spatial insight problems, gesturing facilitates finding solutions. Similarly, gesturing helps people recall and maintain abstract conceptual themes: Preventing gesture altogether reduces the use of spatial metaphors in speech [4]. Gestures do not merely enable access to words, they enable mental access to metaphorical concepts. For example, people consistently made upwardly gestures while telling a story about feeling "up" (i.e., happy) even when they described downward spatial aspects of a scene [6]. When the storytelling task was paired with moving marbles upwards or downwards, storytelling was disrupted when the movements were inconsistent with the story's theme and recall was improved when making thematically consistent movements [6]. These results suggest that physical motions are linked to creativity, metaphor and abstract thinking.

Gesture and embodied cognition

The crossroads of gesture in HCI and cognitive science invites a confluence of theory, and theories of embodied cognition are most germane to our focus. Embodied cognition offers explanations of empirical results in which bodily states and actions affect thinking [2,29]. Even mental simulations of bodily positions can affect thinking. Palmer, Clausner and Kellman [7,26,27] designed air traffic displays with the aim of improving visual search by graphically encoding altitude in 2D displays as icon size and contrast. Two metaphors were tested: larger-darker is higher altitude, and smaller-lighter is higher altitude. Interpreting the altitude of aircraft icons depended on whether participants imagined displays as viewed from above or below. Participants were instructed in one of two conditions: bodily looking head up or head down at instructional displays. Afterward in the visual search task participants looked straight ahead at desktop displays but were asked to imagine them from the perspective in which they had been instructed. Search performance was better when imagining displays from above than from below. The results found an imagined perspective effect consistent with embodied theories of cognition. Moreover, theory served a dual role in the research. Display design was guided by theories of metaphor and visual perception; experimental results contributed to theories of embodied cognition. That is, the research was both theory-grounded and theory-building. HCI has widely embraced embodied cognition theories (as well as others, e.g., [17,25]) for guiding design research, while empirical science is largely theory-building.

In the crossroads of cognitive science and HCI research our investigation is both theory-grounded and theory-building. Whether our results are consistent with embodied cognition theories, or alternative theories is itself a matter of our investigation. To these ends, our research is enabling us to design methods of analysis and experiments to observe children using TUIs during a creative cognition task. We have chosen to study children because our research has a potentially broad impact on cognitive development and on promoting creative thinking with educational technology.

EXPLORATION IN AN HCI DESIGN PERSPECTIVE

We initially explored how children use TUIs by observing play and design sessions [23]. We applied protocol analysis and the KidsTeam co-design methodology which includes children as active partners in design research [12,13].

Seven children aged 7 to 12 years were assigned to work in three groups of 2, 2, and 3 children each. Each group played one of thee different Sifteo™ games, which varied in the kinds of cognitive skill required: a math puzzle, a spatial

tiles puzzle, and a resource sharing game. We video recorded the children using the tangible cubes.

The process of coding the video data revealed a fundamental methodological challenge arising from our joint perspective for both HCI and cognitive considerations on these data. There was a need to categorize multiple children interacting with multiple cubes, in multiple modes of communication, in multiple spatial locations. Each child talked, gestured, and played with the cubes, and interacted socially. Each cube displayed texts and pictures, emitted sounds, and sensed actions. Play resulted in actions on cubes that cubes were designed to sense, and ones they were not. Cubes stood in relationships with other cubes. Spatial relationships dynamically evolved in child-child, child-cube, and cube-cube combinations. Simultaneous actions compounded the challenge. For example, in one group a boy and girl played on the floor with three cubes. In the span of only six seconds a boy held a cube in his left hand while he pointed at a cube resting on the floor, touched its display, while a girl grasped a third cube with both hands, put it on the floor, neighbored one of its sides to the side of the cube the boy was touching, then rotated her cube 90º, and neighbored that side, lifted the cube off the floor, then touched its display. Theses actions included holding, pointing, touching, putting down, picking up, rotating, and interpersonal coordination.

In developing a coding scheme for analyzing, these data we considered both technology-centered and human-centered approaches, which yielded schemes of contrasting usefulness. A cube-centric perspective resulted in a combinatorial explosion of cube actions, per cube, per hand, per person, in combination with speech and other gestures. A behavioral perspective yielded a simple set of coding categories. We contribute this human-centric coding scheme for analyzing groups of people using tangible computing devices (Table 1). We use the term "action on cube" to distinguish gestures directed to a TUI from other gestures. This scheme is not limited to cube-sensed actions supported by a specific TUI design. The simplicity of the behavioral perspective is not an artifact of its generality, but because it represents a human-centered point of view on the interactions among humans and devices. This observation highlights the issues in an analysis of data at the intersection of HCI and cognitive science.

Table 1 Coding Scheme Categories

- Action on Cube
  - Cube-Sensed (e.g., press, neighbor, flip, shake, tilt)
  - Not Sensed (e.g., rotate, stack, pick up, put down, arrange, grasp)
- Gesture
  - Hand gesture (e.g. pointing)
  - Non-hand gesture (e.g. head nod)
- Audible Communication
  - Child speech
  - Child-directed sounds emitted by TUI
- Interpersonal action exchanged between children

The purpose of coding schemes and experimental designs in HCI research is to explore and enable technology design, and often introduce confounding variables without demanding rigid reductionism. Investigating cognitive effects of using technology, however, has a scientific basis in cognitive science, which demands more rigorous experiments. We are exploring this methodological crossroads.

EXPERIMENTAL DESIGN

Focusing on gesture with TUIs is enabling us to design novel experiments that meet another challenge of working in the crossroads of two areas: how to vary affordances while holding stimulus parameters constant. We met this challenge by rigorously designing experimental and control conditions that differ in selected physical and perceptual attributes, but not in the task-relevant information they display.

Two hypotheses are the focus of this experiment. 1. Tangible interaction increases the quantity and creative quality of new ideas. 2. Tangible interaction encourages spatial and metaphorical thinking. The creativity task we chose for testing these hypotheses was conceptual combination/blending [32]. Given a set of words children combined words and invented creative meanings. Stimuli were designed to promote creative thinking.

Word Stimuli: We designed two word sets consisting of six words each, matched on a variety of task-relevant psycholinguistic properties, e.g., all words were nouns about familiar objects. Within each set, there were no pairs of words that formed familiar compounds (e.g., cow, boy)

Display Design: We designed the TUI displays and poster paper stimuli to match on a variety of perceptual attributes: font size, cube/square size, and spatial layout (Figure 3).

Method and Procedure

Forty 6th grade children (aged 11-12) from a local middle school participated in our study. Children were randomly placed into 20 pairs. Working in pairs was intended to promote talking and gesturing, while still encouraging focus on the task. In the control condition, the children were seated at a table with a poster paper on which six stimulus words were printed (Figure 2 & 3). In the tangible condition, the children were seated at a table with six Sifteo cubes, each displaying a single word (Figure 1). Each pair of children participated in both conditions.

The experimental design was within-subjects and the stimuli were counterbalanced such that each group saw words displayed with similar visual form in two contrasting conditions. This allowed us to contrast tangible and non-tangible affordances for the same creative cognition task while keeping the stimuli as constant as possible.

Initial Observations

Data analysis is work-in-progress: we are applying our coding scheme in preparation for testing our hypotheses. We developed a set of observations that will guide our

coding and analysis of the protocol data in two categories: verbal responses and gesture responses.

Verbal Responses: Each pair of children generally repeated a three-stage response pattern: searching for word combinations, verbally responding by saying the words, then explaining their response by describing their creative idea. Though not instructed to do so, children tended to describe the meanings of their selected combinations with word-relation-word series in the form:

$$Word_1\text{-}Relation_1\text{-}Word_2, \ldots Relation_i\text{-}Word_{i+1} \ldots Relation_n\text{-}Word_n \text{ , where } 2 \leq n \leq 6$$

For example, a four-word selection, *bee-shirt-rope-desk* was explained as, "A bee that wears a shirt made of ropes who sits at a desk eating rice in a car." A particularly creative example was the two-word selection *shoe-cow* explained as "A cow that is obsessed with shoes."

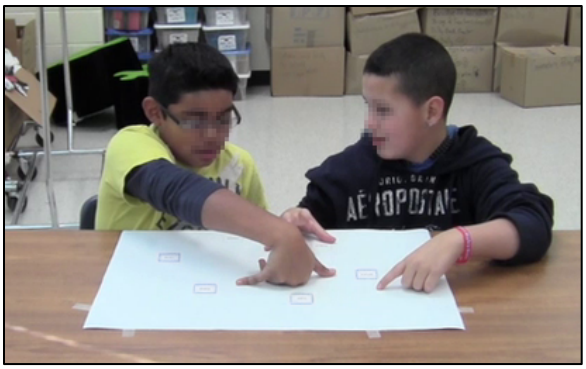


Figure 2 Control Condition: Children using words printed on a poster to make creative word combinations.

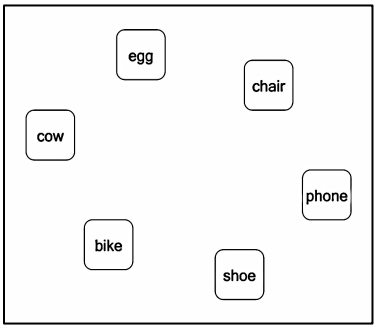


Figure 3 Poster word stimuli matched TUIs.

Gesture Responses: In the control (paper) condition gestures were largely restricted to pointing gestures. In contrast, in the TUI condition children arranged cubes into arrays of pairs and triples, while sorting through two-word and three-word combinations or linear arrays of six words in preparation of (or while) verbally reporting word combinations. TUIs promoted creative gestures, e.g., a child held three cubes representing his three-word response as he demonstrated that two of the cubes, one stacked on the other, represented the relation "happens at the same time". This use of space to express times is metaphorical.

Beyond applying our coding scheme to the experimental data for testing our hypotheses this research confronts evaluation of creativity, and investigates gestures as creative exploratory actions (i.e., epistemic gestures).

HYPOTHESES IN THE CROSSROADS

Our initial research both showcases the difficulty of confronting variables that confound experimental method when studying human interactions with technologies, and provides directions for refining our approach. At this stage in our methodological crossroads, there is significant value in hypothesis generation.

Empirical science typically proceeds by formulating hypotheses testable against expected results in un-confounded ideal conditions. HCI research typically proceeds toward developing design principles, undeterred by confounding variables. In the crossroads of HCI and cognitive science we are finding statements emerging from design-guided HCI research questions, and simultaneously refining them toward testable experimental hypotheses. We propose the following preliminary hypotheses about tangible interaction and creative cognition:

Tangible interaction...

1. facilitates epistemic gestures and actions.
2. encourages thinking about non-spatial / abstract concepts.
3. encourages greater bodily movement beyond that necessary to interact.
4. encourages more spatial and metaphorical thinking.
5. enhances creative cognition.
6. serves to offload cognition as tactile, visual, and spatial representation of working memory, and as externalization of cognitive processes.

SUMMARY AND FUTURE WORK

Our initial protocol study, the resulting coding scheme, and our experimental design represent progress thus far on cultivating a theory-grounded research area. We aim to engender new methods and build new theory in the crossroads of HCI and cognitive science. These initial studies are raising questions about confounding variables, measurement, and other methodological challenges unique to the intersection of HCI and cognitive science. The challenge is not merely a tug-of-war between research paradigms in technology design and experimental science, or where on a spectrum to select a method. Our preliminary results inform new directions for blending methodologies, from which we expect will emerge new methods, results, and theory. Our research program is revealing the essential value in bringing together HCI design and cognitive science for conducting research in the crossroads.

ACKNOWLEDGMENTS

This research was funded by NFS grant no. IIS-1218160 to M. Maher, T. Clausner, and A. Druin.

REFERENCES

1. Alibali, M.W. and DiRusso, A.A. The function of gesture in learning to count: more than keeping track. *Cognitive Development 14*, 1 (1999), 37-56.
2. Barsalou, L.W. Perceptual symbol systems. *Behavioral and brain sciences 22*, 04 (1999), 577–660.
3. Bødker, S., Ehn, P., Sjögren, D., and Sundblad, Y. Co-operative Design – perspectives on 20 years with 'the Scandinavian IT Design Model.'*Proceedings of NordiCHI 2000, Stockholm, October 2000*, (2000).
4. Bos, A.J. and Cienki, A. Inhibiting gesture reduces the amount of spatial metaphors in speech. *Gesture and Speech in Interaction*, (2011).
5. Carlson, R.A., Avraamides, M.N., Cary, M., and Strasberg, S. What do the hands externalize in simple arithmetic? *Journal of Experimental Psychology: Learning, Memory, and Cognition 33*, 4 (2007), 747.
6. Casasanto, D. and Dijkstra, K. Motor action and emotional memory. *Cognition 115*, 1 (2010), 179–185.
7. Clausner, T.C., Kellman, P.J., and Palmer, E.M. Conceptualization in Language and Its Relation to Perception. *Proceedings of the Fifty-first Annual Meeting of the Cognitive Science Society*, (2008).
8. Cook, S.W. and Goldin-Meadow, S. The role of gesture in learning: Do children use their hands to change their minds? *Journal of Cognition and Development 7*, 2 (2006), 211–232.
9. Day, S., Goldstone, R.L., and Hills, T. The effects of similarity and individual differences on comparison and transfer. *Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society*, (2010), 465–470.
10. Day, S.B. and Goldstone, R.L. Analogical transfer from a simulated physical system. *Journal of Experimental Psychology: Learning, Memory, and Cognition 37*, 3 (2011), 551.
11. Druin, A. Children as codesigners of new technologies: Valuing the imagination to transform what is possible. *New Directions for Youth Development 2010*, 128 (2010), 35–43.
12. Fitzmaurice, G.W. *Graspable user interfaces*. *Proceedings of CHI '95*, (1995), 442–449..
13. Fjeld, M., Bichsel, M., and Rauterberg, M. BUILD-IT: an intuitive design tool based on direct object manipulation. *Gesture and Sign Language in Human-Computer Interaction*, (1998), 297–308.
14. Gibson, J.J. Observations on active touch. *Psychological review 69*, 6 (1962), 477.
15. Goldin-Meadow, S. and Beilock, S.L. Action's influence on thought: The case of gesture. *Perspectives on Psychological Science 5*, 6 (2010), 664–674.
16. Goldin-Meadow, S., Cook, S.W., and Mitchell, Z.A. Gesturing gives children new ideas about math. *Psychological Science 20*, 3 (2009), 267–272.
17. Van den Hoven, E. and Mazalek, A. Grasping gestures: Gesturing with physical artifacts. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing 25*, 03 (2011), 255–271.
18. Kessell, A. and Tversky, B. Using diagrams and gestures to think and talk about insight problems. *Proceedings of the Meeting of the Cognitive Science Society*, (2006).
19. Kim, M.J. and Maher, M.L. The impact of tangible user interfaces on designers' spatial cognition. *Human–Computer Interaction 23*, 2 (2008), 101–137.
20. Kirsh, D. and Maglio, P. On distinguishing epistemic from pragmatic action. *Cognitive science 18*, 4 (1994), 513–549.
21. Lee, C.-H., Ma, Y.-P., and Jeng, T. A spatially-aware tangible interface for computer-aided design. *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, ACM (2003), 960–961.
22. Ma, Y., Lee, C.H., and Jeng, T. iNavigator: A spatially-aware tangible interface for interactive 3d visualization. *Proceedings of Computer Aided Architectural Design Research in Asia (CAADRIA2003)*, (2003), 963–973.
23. Maher, M.L., Clausner, T.C., Gonzalez, B., and Grace, K. (Submitted) A Protocol Analysis Methodology for Tangible Interaction to Enhance Creative Cognition. (2014).
24. Marshall, P. Do Tangible Interfaces Enhance Learning? *Proceedings of the 1st International Conference on Tangible and Embedded Interaction*, ACM (2007), 163–170.
25. McNeill, D. *Hand and mind: What gestures reveal about thought*. University of Chicago Press, 1992.
26. Palmer, E.M., Brown, C.M., Bates, C.F., Kellman, P.J., and Clausner, T.C. Perceptual affordances and imagined viewpoints modulate visual search in air traffic control displays. (2009), 1111–1115.
27. Palmer, E.M., Clausner, T.C., and Kellman, P.J. Enhancing air traffic displays via perceptual cues. *ACM Transactions on Applied Perception (TAP) 5*, 1 (2008), 4.
28. Rittel, H. and Webber, M. Wicked problems. *Man-made Futures*, (1974), 272–280.
29. Spivey, M.J. and Dale, R. On the continuity of mind: Toward a dynamical account of cognition. *Psychology of learning and motivation 45*, (2004), 87–142.
30. Ullmer, B. and Ishii, H. Human-Computer Interaction in the New Millenium. In J.M. Carroll, ed., Addison-Wesley, 2001, 579–601.
31. Visser, W. and Maher, M.L. The role of gesture in designing. *AI EDAM-Artificial Intelligence Engineering Design Analysis and Manufacturing 25*, 3 (2011), 213.
32. Wisniewski, E.J. and Gentner, D. On the combinatorial semantics of noun pairs: minor and major adjustments to meaning. *Understanding word and sentence, Amsterdam, North Holland*, (1991), 241–284.

Congruent Gestures can Promote Thought

Barbara Tversky
Azadeh Jamalian
Ayelet Segal
Columbia Teachers College
New York, NY
btversky@stanford.edu
azijamalian@gmail.com
segalayelet@gmail.com

Valeria Giardino
Institut Jean Nicod
Paris, France
Valeria.Giardino@ens.fr

Seokmin Kang
Arizona State University
Tempe, AZ
chandlerkang@gmail.com

ABSTRACT
Considerable research has indicated that gestures can facilitate thinking in those who view them as well as those who make them. Gestures can represent concrete and abstract information, both about structures and especially about actions, more directly than purely symbolic words. This research suggests that embedding gestures that are congruent with thought in touch interfaces can improve performance.

Author Keywords
Gesture; congruence; thinking; touch interface.

ACM Classification Keywords
H.5.m. Information interfaces and presentation.

INTRODUCTION
Gestures take many forms and serve many ends (e.g., Goldin-Meadow, 2003; Kendon, 1994; McNeill, 1985, 1992). *Emblems*, like "thumbs up" or "OK" are conventionalized and act like words. *Beats* serve to support the structure of the discourse, typically accompanying phrases like "on the other hand" or "first." The gestures of interest here are those that convey meanings, often called *representational* gestures. They include *deictic* gestures that point to actual or imagined objects and features in the world, *iconic* gestures that depict or enact features like shapes or actions, and *metaphoric* gestures that express abstract meanings and relations. Integrated strings of related representational gestures can create models or diagrams, for example, of environments or family trees or natural processes (e. g., Emmorey, Tversky, and Taylor, 2000; Engle, 1996; Kang, Tversky, and Black, 2012; Jamalian, Giardino, and Tversky, 2013; Tversky, Heiser, Lee, and Daniel, 2009). Such gestures are communicative, for self or others or both.

Gestures, then, can and do express thoughts and thinking. Moreover, because they can bear similarities to the meanings they represent, gestures can express the objects and processes of thoughts more directly than words, which typically bear only arbitrary, symbolic relations to meanings. Points to things in the world, for example, are universally understood and produced by children before they begin to speak. That gestures can express meanings quite directly seems to underlie their utility in promoting thinking, both for self (e. g., Krauss, 1998; Kessell and Tversky, 2006; Jamalian, et al., 2012) and for others (e.g., Goldin-Meadow and Beilock, 2010; Goldin-Meadow, Cook, and Mitchell, 2009; Kang, et al., 2012).

Gestures are actions, but they are representative actions, not instrumental actions. Gestures can represent actions, as well as objects, features, models, and processes. Just as imagery has been regarded as internalized perception (e. g., Kosslyn, 1980; Shepard and Podgorny, 1978), thinking can be regarded as internalized action (e. g., Bruner, 1966; Piaget, 1928; Vygotsky, 1962). That thinking is regarded as internalized action is evident in the ways we speak: the language we use to express thought is often the language we use to express actions (e.g., Lakoff and Johnson, 1980). Thoughts are reified as objects that can be acted on in myriad ways. We pull our ideas together or tear them apart. We mix them up. We put them on the table. We draw them out. We raise issues for discussion or we let them drop. Others crush them. The list of examples is endless; once attuned, the ear hears them everywhere. These expressions are not heard as colorful poetic figures of speech, but rather as direct and ordinary ways of talking. And many are accompanied by gestures that enact the metaphoric actions.

HOW MIGHT GESTURES AFFECT THOUGHT?
If thinking can be regarded as--at least in part--internalized action, then reexternalizing thinking as action should facilitate thinking, for self and for others. Gesture could augment thinking in many different ways. Gestures provide an additional code for memory, a motor code that can enrich the representation and provide additional retrieval cues. It is well documented that forming multiple codes of the same information facilitates memory (e.g., Paivio, 1986), and memory retrieval can be regarded as a kind or part of thinking Although much of the research showing facilitation of memory from dual codes was on verbal and pictorial codes, gestures, like pictorial codes, are often

The author(s) retain copyright, and ACM has an exclusive publication license.

depictive, and motor or enactive coding also facilitates memory (e.g. Engelkamp, 1998; Johnson, Foley, Suengas, and Raye, 1988). Because gesture can map meaning more directly than symbolic words, gesture can provide a more congruent mode of expression than speech. People often express thoughts in gesture that are not expressed in words, suggesting that those thoughts are more accessible to similar actions than to abstract words (e. g., Goldin-Meadow, 2003). Children, for example, often demonstrate rudimentary understandings in their gestures, for example, a pouring gesture while attempting to explain conservation, without being able to express them in speech (Broaders, Wagner Cook, Mitchell, and Goldin-Meadow, 2007; Church and Goldin-Meadow, 1986). Gestures often precede the related words in speech (e. g., Kita, 2000), another indication that gestural expression is often more accessible to speakers than words. The motoric aspects of gestural actions may have benefits to thinking in and of themselves. They constitute a visceral, active, embodied representation of mental actions (e.g., Cartmill, Beilock, and Goldin-Meadow, 2012; Hoestetter and Alibali, 2008). Gestural actions can represent, externalize, and embody mental actions, for example, putting, raising, combining, separating, aligning, moving, removing, rotating, and more. The actual actions can directly benefit the thinking of those who make them both by directly representing the meanings and by providing a motor code. Gestures are may provide an embodied motor code to those who observe them as well as those who make them by a brain processes known as "motor resonance" (e. g., Decety, Grezes, Costes, Perani, Jeannerod, Procyk, Grassi, and Fazio, 1997; Iacobini, Woods, Brass, Bekkering, Mazziotta, and Rizzolatti, 1999). Although there are a number of versions of motor resonance (e. g., Utihol, van Rooij, Bekkering, and Haselager, 2011, the general phenomenon is that viewing certain actions activates some areas of the brains that are associated with performing those actions, sometimes referred to as the human mirror system. Some of our recent research has shown how and why gesture affects our own thoughts and those of others, with implications for interface design.

SOME WAYS GESTURES CHANGE THOUGHT IN THOSE WHO SEE THEM

Gestures Change Thinking about Time

As noted, sequences of gestures can create virtual diagrams, drawings in the air. Diagrams, like gestures, whether created or viewed, can have large effects on thinking (e. g., Glenberg and Langston, 1992; Hegarty, 2011; Tversky, 2002; 2011). Diagrams use place in space and elements in space to express sets of ideas that are inherently spatial or metaphorically spatial. Diagrams of sets of ideas that are inherently spatial include maps, architectural plans, and some biological and mechanical structures. Diagrams of sets of ideas that are metaphorically spatial include time lines, calendars, graphs, and charts. Like diagrams of inherently spatial ideas, diagrams of metaphorically spatial ideas use elements and spatial relations in systematic ways (e. g., Tversky, 2011). So do gestures (Tversky, et al., 2009).

Ideas about time are frequently expressed in spatial terms, in language as well as diagrams, in children as well as adults (e. g., Clark, 1973; Boroditsky, 2000; McGlone and Harding, 1998; Tversky, Kugelmass, and Winter, 1991). One project has investigated ways that speakers' gestures can change observers' interpretations of several temporal concepts (Jamalian and Tversky, 2012). In a set experiments, participants heard the same verbal script but saw different gestures. In all cases, thinking was different for the different gestures. The first experiments examined thinking about cyclical processes. Previous work had shown that when people were asked to create diagrams for cyclical processes such as the seasons of the year, the activities of the day, the cell cycle, and the rock cycle, they typically drew linear displays rather than circular ones (Kessell and Tversky, submitted). There are a number of non-conflicting explanations for this: that time itself proceeds linearly and cannot turn back on itself; that processes are conceived to have beginnings, middles, and ends, that is, initial conditions, changes, and outcomes; that individual events do not repeat. Thinking about processes as cyclical then requires several conceptual leaps. It requires abstraction from individual events to classes of events and it requires ignoring actual time to focus on the repeating sequence of types of events that constitute the process. In the experiments using gesture, students sitting in a cafeteria were approached by an experimenter who asked them to think about a set of four events, such as those of the seed cycle or the cycle of a day. The description was accompanied by a set of four gestures either on a line, left to right from the participant's point of view, or clockwise in a circle. After listening to the set of events, participants were asked to put something down on paper to represent the events. Those who saw gestures arrayed along a line tended to draw lines and those who saw gestures arrayed along a circle tended to draw cycles. A follow-up experiment showed that the diagrams were not mere copies of the gestures, but that the gestures changed thought. Instead of drawing a diagram after the last event, participants were asked "what comes next?" Those who had seen gestures in arrayed in a circle tended to return to the first event of the cycle, for example, for the seed cycle that ended in a flower, they tended to return to the ungerminated seed. By contrast, those who had seen gestures arrayed along a line tended to say something completely different, for example, a bouquet was made.

The next experiment altered thinking about simultaneous as opposed to sequential processes. Participants heard a description of 4 steps to writing a paper. The description clearly stated that the middle two steps could be taken in either order. Half the participants saw gestures that emphasized each of the four steps and half saw gestures that

showed that the middle steps could be taken in either order. Later tests of understanding showed that those who had viewed the gestures showing that the order of the middle steps was optional understood that far better than those who hadn't. The final experiment in this series showed that gestures changed interpretations of an ambiguous question: "Next Wednesday's meeting has been moved forward two days; when is it?" Without accompanying gestures, half the people typically respond "Friday," and the other half "Monday." When the statement was accompanied by a gesture away from the body, most people answered "Friday," but when the statement was accompanied by gesture toward the body, most people answered "Monday" as if the gestures had moved the event forward or backward on a timeline extending from present to future away from the body. Interestingly, in all the studies, despite the fact that the gestures clearly influenced thought, many participants reported they hadn't noticed any gestures, suggesting that these gestures were mapped to meaning directly and effortlessly.

Gestures Change Thinking about Action

As noted, gestures are actions, suggesting that they may have a special role in conveying information about action, a concept difficult to convey in diagrams. In learning many complex systems, people typically have more trouble learning the actions of the systems than their structures ((e.g., Hmelo-Silver and Pfeffer, 2004; Tversky, Heiser, and Morrison, 2013). Kang, Black, and I (2012) thought that gestures might help. Students viewed videos of an explanation of how a car engine works. One video was accompanied with gestures that depicted the appearance and structure of the parts of the engine; another video was accompanied with gestures that showed the actions of the parts of the engine. Superimposed on the video was a schematic diagram showing the structure of the parts of the engine. After studying the video, participants were given three knowledge assessments. The first was a set of true-false questions about the structure and action of the engine. That test could be answered solely on the basis of the verbal description. After answering the questions, participants created visual explanations of the workings of the engine. They then explained the workings of the engine to a video camera so that someone watching the video could understand how the engine worked. Participants who saw the action gestures did better on the true-false test of action information than those who had viewed the structure gestures. They also showed more action effects in their visual explanations and used more action gestures in their videotaped explanations. Moreover, the gestures they used to convey the actions of the system were original; they were not mere copies of what they had seen.

Design Implications

Our work and that of many others, notably that of Goldin-Meadow, Alibali, and their collaborators, have demonstrated that a range of different kinds of gestures expressing different kinds of information can alter the thinking of those who view them. Gestures naturally, effortlessly, and significantly complement speech and diagrams in creating effective explanations. These findings have implications for interface design. Interfaces have long incorporated the equivalent of point gestures in numerous ways, highlights, arrows, bold face, blinking, and more. However, interfaces have rarely incorporated the equivalent of representational gestures. Incorporating representational gestures in interfaces is challenging because the semantics and pragmatics of representational gestures are more complex than those of deictic gestures. Careful thought as well as experimentation, such as the research presented and cited here, can give useful insights. Diagrams also give clues, as there are many visual-spatial parallels between the semantics and pragmatics of gestures and diagrams (e. g., Tversky, et al., 2009). For both gestures and diagrams, eliciting them in context has provided valuable information about the semantics and pragmatics of using space and elements in space to reflect and affect thought that can guide interface design.

SOME WAYS GESTURES CHANGE THOUGHT IN THOSE WHO MAKE THEM

It is common nowadays to see perfectly sane people walking alone down the street, oblivious to what is going on around them, gesturing dramatically. They are merely talking on cell phones. Their interlocuters cannot see the gestures, so those gestures can't be meant for them. The gestures seem to help the speakers. Indeed, experiments have shown that when people sit on their hands, they have trouble finding words (e.g., Krauss, 1998). But recent experiments have shown that gestures do far more for speakers than help them speak; they also help them think. Counting is facilitated by successive pointing, and if hands are occupied, heads and eyes point, though with less precision than fingers, showing that gestures of some sort are hard to suppress (e. g., Carlson, Avraamides, Cary, and Strasberg, S., 2007). Children who use two fingers on one hand to point to both sides of an equation solve elementary algebra problems better than children who don't (Goldin-Meadow, Cook, and Mitchell, 2009). So it seems that gestures that map to meaning not only express the meanings but also help the people who make them think about the meanings. A corollary to that proposition is that gestures that are congruent with meanings should facilitate thinking, but gestures that are incongruent with meanings might actually interfere with it.

Gesturing Helps Spatial Problem Solving

We have found that people alone in a room gesture when solving spatial problems or reading spatial descriptions, and that their gestures have consequences for problem solving and memory. In the first study, participants solved six spatial insight problems like Duncker's radiation problem (Kessell and Tversky, 2006). After solving each problem, they explained the solution to a video camera so that

someone watching the video could understand how to solve the problem. Naturally, when explaining their solutions, participants gestured for each problem and made far more gestures than when solving the problems. When solving, they gestured only for the two problems with high spatial memory loads. In a parallel experiment in which participants were given paper and pencil, they tended to diagram the same two problems while solving them. In solving the problems, both gestures and diagrams represented the spatial arrays of the problems. In explaining the solutions, the gestures represented both the spatial arrays and the problem solution. Participants provided with pencil and paper used primarily gestures in explaining problem solutions; they rarely used their diagrams. For the Six Glasses problem, participants were told that there was a row of six glasses, three full and three empty; they were told to change the configuration to empty-full-empty-full-empty-full by moving only one glass. Most participants gestured while solving the six glasses problem, and those who gestured were more likely to solve it than those who didn't. Most participants also gestured for the Boat problem, which was a red herring. Participants read that the boat ladder had X rungs that were Y feet apart, that the water level currently reached the third rung, and that the tide was coming in at so many feet per hour; they were asked which rung the water would reach when the tide came in. Most participants were fooled, and computed as if the boat were anchored to the bottom of the ocean and the tide would rise along the ladder. The participants who gestured solve the problem incorrectly, but correctly computed given their incorrect assumption. The small number who correctly realized that the boat floats so that water level would always be at the third rung didn't gesture. Thus in both cases, gestures for self facilitated finding a solution, but in the case of six glasses, the solution was correct, and in the case of the boat ladder, the solution was incorrect.

Gesturing Helps Memory for Environments

In a second study, participants read four descriptions of environments such as a town or a spa either from route or survey perspective (Jamalian, Giardino, and Tversky, 2013). Half the environments had four landmarks and half had eight. They studied the descriptions to prepare for later tests of memory for the environments, where the true-false statements were either taken directly from the descriptions or required inferences from information in the descriptions. More than 70% of participants gestured for at least one description as they studied the descriptions. Almost all of those who gestured while studying also gestured for at least one description while answering the true-false questions. For the most part, their gestures represented the spatial structure of the environments, paths through the environments, represented by lines, and places, represented by points to locations of landmarks. Their gestures did not represent the visual information in the descriptions, nor did they represent the perspective, route or survey, of the descriptions. Those who gestured while studying performed better on the memory tests than those who didn't, and those who gestured while studying performed better on the descriptions they gestured for than those they didn't gesture for. Similarly, gesturing at test improved performance over and above gesturing at study. Interestingly, participants rarely looked at their gestures, and when they did, it was brief glances. That is, although their gestures virtually drew bits and pieces of diagrams, they were motor diagrams, not visual ones. This is reminiscent of skilled musicians, who rarely need to look at their hands, the actions per se keep track of the locations of the keys or strings or holes. Our interpretation is that the gestural actions in space played a direct role in creating coherent and complete mental models of the environments, and that the actions per se represent the model of the environment. To demonstrate, we ask people to tell us where the *x* is on the keyboard; almost everyone moves their finger as if to type *x*.

Together, the findings suggest that spontaneous actions, in particular, gestures, can embody thought, that qualities of the gestures represent qualities of thought, and that representing thinking with gestures can facilitate thinking providing that the mappings between gesture and the required thought are congruent.

INCORPORATING CONGRUENT GESTURES IN TOUCH INTERFACES

Touch interfaces present an opportunity to harness the cognitive power of gestures to facilitate performance. The general idea is that gestures congruent with thought can facilitate thought. Cues for congruent gestures can come from gestures used in explanations. In one set of experiments, people explained their solutions to math problems. For discrete math problems, explainers spontaneously used discrete gestures; for continuous math problems, they used smooth, continuous gestures (Alibali, Bassok, Olseth, Syc, and Goldin-Meadow, 1999). In other words, discrete gestures are congruent with discrete thinking and continuous gestures with continuous thinking. We thought that embedding congruent gestures in a touch interface should facilitate math performance (Segal, Tversky, and Black, submitted). Young children performed a discrete task, addition, and a continuous task, number line estimation, with gestures that were either congruent or incongruent. In the addition task, children added two sums where each sum was the number of bricks in a pile. In both cases, gestures were discrete, but in the congruent case, they were mapped brick by brick and in the less congruent case, they were mapped to the sums of each column. In the number line estimation task, children decided where on a line from 0 to 100 a series of numbers were. In both cases, they made a mark on the number line. In the congruent case, they slide a marker from 0 to the desired point, so that both time and distance were congruent with the estimation. In the less congruent case, children marked the point

directly. For both addition and number line estimation, performance was better for the (more) congruent action.

IMPLICATIONS

Our work and that of many others, notably that of Goldin-Meadow, Alibali, and their collaborators, have demonstrated that a range of different kinds of gestures representing different kinds of information can alter the thinking of those who view them and those who make them. Gestures naturally, effortlessly, and significantly complement speech and diagrams in creating effective explanations. These findings have implications for interface design, both the graphics side and the action side. Interfaces have long incorporated the equivalent of deictic (pointing) gestures in numerous ways, highlights, arrows, bold face, blinking, and more. However, interfaces have rarely incorporated the equivalent of representational gestures. Incorporating representational gestures in interfaces is challenging because the semantics and pragmatics of representational gestures are more complex than those of deictic gestures. Designing congruent embodied interactions also presents challenges to uncover natural mappings of thinking to acting. Careful thought as well as experimentation, such as the research presented and cited here, can give useful insights. Diagrams also give clues, as there are many visual-spatial parallels between the semantics and pragmatics of gestures and diagrams (e. g., Tversky, et al., 2009). For both gestures and diagrams, eliciting them in context, turning users into designers (e. g., Tversky, 2011), has provided valuable information about the semantics and pragmatics of using space and elements in space to reflect and affect thought that can profitably guide interface design.

ACKNOWLEDGMENTS

The authors are grateful to the following grants for facilitating the research and/or preparing the manuscript: National Science Foundation HHC 0905417, IIS-0725223, IIS-0855995, and REC 0440103.

REFERENCES

1. Alibali, M. W., Bassok, M., Olseth, K. L., Syc, S. E., & Goldin-Meadow, S. (1999). Illuminating mental representations through speech and gesture. *Psychological Science, 10*, 327-333.

2. Andres, M., Seron, X., & Olivier, E. (2007). Contribution of hand motor circuits to counting. *Journal of Cognitive Neuroscience, 19*(4), 563-576.

3. Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition* (75), 1-28.

4. Broaders, S. C., Wagner Cook, S., Mitchell, Z. and Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning *Journal of Experimental Psychology: General, 136,* 539-550.

5. Bruner, J. S. (1966). On cognitive growth. In J. S. Bruner, R. R. Olver, & P. M. Greenfield (Eds.), *Studies in cognitive growth* (pp. 1–29). Oxford, England: Wiley.

6. Carlson, R. A., Avraamides, M. N., Cary, M., & Strasberg, S. (2007). What do the hands externalize in simple arithmetic? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 747-756.

7. Cartmill, E. A., Beilock, S. L., and Goldin-Meadow, S. (2012). A word in the hand: Human gesture links representations to actions. *Philosophical Transactions of the Royal Society,* B, 367, 129–143.

8. Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General, 137,* 706-723.

9. Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition, 23,* 43-71.

10. Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. Pp. 27–63. New York: Academic Press.

11. Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E.,Grassi, F., Fazio, F., 1997. Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain, 120,* 1763-1777.

12. Eisenegger, C., Herwig, U., & L. Jäncke, L. (2007). The involvement of primary motor cortex in mental rotation revealed by transcranial magnetic stimulation. *European Journal of Neuroscience, 4,* 1240-1244.

13. Emmorey, K., Tversky, B., & Taylor, H. (2000). Using space to describe space: Perspective in speech, sign, and gesture. *Spatial Cognition and Computation, 2,* 157-180.

14. Enfield, N. (2003). Producing and editing diagrams using co-speech gesture: Spatializing non-spatial relations in explanations of kinship in Laos. *Journal of Linguistic Anthropology, 13*, 17-50.

15. Engle, R. A. (1998). Not channels but composite signals: Speech, gesture, diagrams and object demonstrations are integrated in multimodal explanations. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

16. Engelkamp, J. (1998). *Memory for actions*. Hove, UK: Psychology Press.

17. Fitts, P. M., & Deininger, R. L. (1954). S-R compatibility: Correspondence among paired elements within stimulus and response codes. *Journal of Experimental Psychology, 48,* 483-492.

18.Ganis, G., Keenan, J. P., Kosslyn, S. M., & Pascual-Leone, A. (2000). Transcranial magnetic stimulation of primary motor cortex affects mental rotation. *Cerebral Cortex, 10*, 175-180.

19.Glenberg, A. M. & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language,* 31, 129-151.

20.Goldin-Meadow, S. (2003). *How our hands help us think*. Cambridge, MA: Harvard University Press.

21.Goldin-Meadow, S. and Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science, 5,* 5664-674.

22.Goldin-Meadow, S., Cook, S.W., & Mitchell, Z.A. (2009). Gesturing gives children new ideas about math. *Psychological Science,* 20, 267–272.

23.Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science, 3,* 446-474.

24.Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science, 1,* 127–138.

25.Hostetter, A. B. & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin and Review*, *15*, 495-514.

26.Iacoboni, M., Woods, R.P., Brass, M., Bekkering, H., Mazziotta, J.C.,Rizzolatti, G., (1999). Cortical mechanisms of human imitation. *Science, 286*, 2526-2528.

27.Iverson, J. M., & Goldin-Meadow, S. (1997). What's communication got to do with it? Gesture in children blind from birth. *Developmental Psychology, 33,* 453-467.

28.Jamalian, A. & Tversky, B. (2012). Gestures alter thinking about time. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society,* Pp. 551-557. Austin TX: Cognitive Science Society.

29.Jamalian, A., Giardino, V., and Tversky, B. (2013). Gestures for thinking. In M. Knauff, M. Pauen, N. Sabaenz, and I. Wachsmuth (Eds), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, Austin TX: Cognitive Science Society.

30.Johnson, M.K., Foley, M.A., Suengas, A.G., & Raye, C.L.(1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *JEP: General, 117,* 371-376.

31.Kang, S., Tversky, B. & Black, J. B. (2-12). From hands to minds: Gestures promote action understanding. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, Pp. 551-557. Austin TX: Cognitive Science Society.

32.Kang, S., Tversky, B. & Black, J. B. (2013) Gesture and speech in explanations to experts and novices. Submitted.

33.Kansaku, K., Carver, B., Johnson, A., Matsuda, K, Sadato, N., & Hallett, M. (2007). The role of the human ventral premotor cortex in counting successive stimuli. *Experimental Brain Research, 178*, 339-350.

34.Kendon, A. (2004). *Gesture: visible action as utterance.* Cambridge: Cambridge University Press.

*35.*Kessell, A. M. & Tversky, B. (2006). Using gestures and diagrams to think and talk about insight problems. In R. Sun and N. Miyake (Eds) *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. P. 2528.

36.Kessell, A. M. & Tversky, B. (submitted). Linear and circular thinking.

37.Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence*, 73, 31-68.

38.Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162 – 185). Cambridge, England: Cambridge University Press.

39.Kosslyn, S. M. (1980). *Image and mind.* Cam- bridge: Harvard University Press.

40.Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, *7*, 54–60.

41.Lakoff, G. and Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

42.McGlone, M.S., & Harding, J.L. (1998). Back (or forward?) to the future: The role of perspective in temporal language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1211-1223.

43.McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review, 92,* 350-371.

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

45.Manoach, D. S., Schlaug, G., Siewert, B., Darby, D. G., Bly, B. M., Benfield, A., Edelman, R. R., & Warach, S. (1997). Prefrontal cortex fMRI signal changes are correlated with working memory load. *Neuroreport, 8,* 545-549.

46.Paivio, A. (1986). *Mental representations*. New York: Oxford University Press.

47.Schwartz, D. (1999). Physical imagery: Kinematic vs. dynamic models. *Cognitive Psychology, 38,* 433-464.

48.Schwartz, D. L., & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fall-back. *Cognitive Science, 20,* 457-497.

49.Schwartz, D. L. and Black, T. (1999). Inferences through imagined actions: Knowing by simulated doing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 116-136.

50.Shepard, R. N., & Podgorny, P. (1978). Cognitive processes that resemble perceptual processes. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. *5,* pp. 189–237). Hillsdale, NJ: Lawrence Erlbaum.

51.Thomas, L. E. and Lleras, A. (2009). Swinging into thought: Directed movement guides insight in problem solving. *Psychonomic Bulletin and Review, 16*, 719-723.

52.Tversky, B. (2011). Visualizing thought. *Topics in Cognitive Science.* 3, 499-535.

53.Tversky, B., Heiser, J., Lee, P. and Daniel, M.P. (2009). Explanations in gesture, diagram, and word. In K. R. Coventry, T. Tenbrink, & J. A. Bateman (Editors), *Spatial language and dialogue*. Pp. 119-131. Oxford: Oxford University Press.

54.Tversky, B. Heiser, J. and Morrison, J. (2013). Space, time, and story. In. B. H. Ross, Editor, *The psychology of learning and motivation.* Pp. 47-76. Oxford: Elsevier.

55.Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology, 23*, 515–557.

56.Utihol, S., van Rooij, Il, Bekkering, H., and Haselager, P. (2011). Understanding motor resonance. *Social Neuroscience, 6*, 388-97.

57.Vygotsky, L. (1962). *Thought and language*. Cambridge: MIT Press.

58.Wexler, M., Kosslyn, S., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition, 68*, 77-94.

59.Wohlschläger, A., & Wohlschläger, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance, 24*(2), 397-314.

Towards Learnable Gestures for Exploring Hierarchical Information Spaces at a Large Public Display

Christopher Ackad
Faculty of Engineering and Information Technologies
The University of Sydney, NSW, 2006, Australia
christopher.ackad@sydney.edu.au

Judy Kay
Faculty of Engineering and Information Technologies
The University of Sydney, NSW, 2006, Australia
judy.kay@sydney.edu.au

Martin Tomitsch
Faculty of Architecture, Design and Planning
The University of Sydney, NSW, 2006, Australia
martin.tomitsch@sydney.edu.au

ABSTRACT
Large displays are fast entering the public spaces, and some are beginning to support interaction based on mid-air gestures. At present, there is no set of standard gestures for that interaction; so people need to draw on existing mental models to learn such interaction. As a step towards closing this gap, we present the design and evaluation of *learnable* gestures for exploring a hierarchical information space on a public display. We draw on a series of in-the-wild studies, observing how people learn our 4 gestures, combined with explanatory icons and a skeleton representation for feedback. We conclude that people readily learnt the pair for exploring a linear information space, while having more difficulty learning the gestures to move up and down the hierarchy.

Author Keywords
Gesture-based Interaction; Interactive Wall; Public Displays;

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
With falling display costs and the availability of large screens and high-visibility projection technologies, public displays have found their ways into a wide range of public spaces. Uses span train departure times, to the broadcast of sports events in pubs, and advertising in foyers. To date, most public screens have not been interactive; rare exceptions are some SMS-based audience participation [10]. This is set to change, with the arrival of cheap and readily available technologies, such as the Microsoft Kinect. This can recognise mid-air gestures, creating unprecedented opportunities to transform these large public screens into interactive platforms. Mid-air gestures are suited to large screen interaction, so that people can view the content from a distance. Mid-air gestures enable immediate use, unlike approaches that use a secondary display such as a mobile phone [2].

We designed the Media Ribbon public display [4, 1], to enable users to explore a potentially large hierarchical information space on the wall. Our goal was to enable people to use mid-air gestures to browse, find areas of interest, and then drill down to further information. This makes the Media Ribbon an ideal testbed for exploring learnable gesture sets. Our design aimed to enhance learnability, by ensuring: the gesture set is small, gestures are easy to describe/teach, perform, remember and the technology should reliably recognise them. Taking account of the public setting; we aimed for gestures people would find acceptable to perform in public.

BACKGROUND
There has been considerable work on attracting users to public displays [4, 8], and exploring social behaviours [11, 5] around these displays. For example, the Chained Displays interface [11] used an interactive space invaders style game to gain the interest of passers-by. The game used a number of discoverable poses to control the strength and frequency of the player's weapon. Similarly, the StrikeAPose interface [12] used a game to explore the learnability of a gesture through 3 different on-screen cues: a static instruction shown at the bottom of the display at all times, another where the instructions obstruct the display, and a third with the instruction integrated and displayed as a tool tip above the user's representation on the screen.

Hespanhol et al. [6] explored the learnability of four gesture primitives (Dwell, Lasso, Push, Grabb) without any user prompts. They conclude that the most intuitive of these was "Dwell", perhaps indicating a mental model based on using the hand like a mouse. To our knowledge, there are currently no standards or other primers for the gestures people would expect to use to navigate a hierarchical information space.

DESIGNING THE MEDIA RIBBON
We have deployed our "Media Ribbon" in a number of locations, from interior corridors during an Open Day event at our university to the exterior of buildings for our long term in-the-wild study. Our current installation is situated on the exterior wall of a glass building, adjacent to a theatre courtyard (Figure 1). The Kinect sensor reaches out to the pedestrian

Figure 1. The Media Ribbon running on the public display (left) facing a theatre courtyard (right).

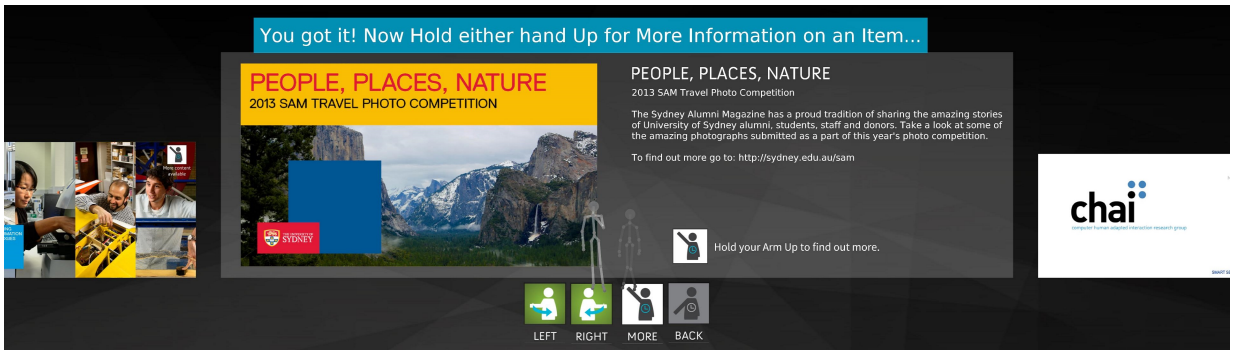


Figure 2. The Media Ribbon showing the tutorial on an expanded top-level item that has sub items.

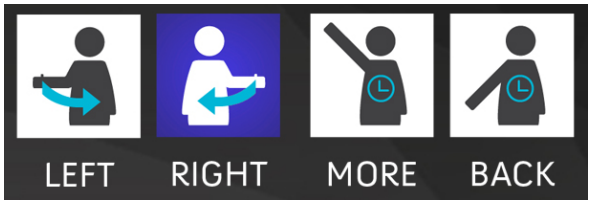


Figure 3. The on-screen help bar with the "Right" gesture highlighted.

walkway running between the two buildings, with the installation being clearly visible from within the courtyard. The installation consists of two high-intensity 1080p projectors, for rear projection on a film laminated onto the glass wall. A 2 centimetre gap separates the two projection panes. A single Kinect for Windows hangs in its own enclosure above and in front of the projection display.

The version of "Media Ribbon" shown in Figure 1 enables users to explore a hierarchical structured set of information, such as upcoming events, faculty news and a showcase of research projects. It is a wide interactive display, which first appears as a horizontal ribbon of media items (Figure 2). To aid engagement with the Media Ribbon, we followed the recommendations [7] to use a wide display, so that passers-by have time to react and engage while still in front of the display. It recognises four primary mid-air gestures (Left, Right, More and Back) which we now describe.

The central element of the Media Ribbon (Figure 2), is a ribbon showing the currently selected hierarchy level and all its items. The ribbon can be scrolled left or right. At any time the items closer to the centre are largest and as the user scrolls, items moving towards the centre become larger, and those moving from the centre become smaller. This means the central item is largest, and neighbouring items provide context. Once an item is centred, it automatically expands, displaying the multimedia content (an image or video) on the left and descriptive text on the right. This design approach was partly motivated by need to avoid having the focus content having a gap in the middle.

We provide three ways to help users learn the gestures. First is a real time representation of the user, as a skeleton on top of the content. Second is the four icons shown in Figure 3. We place this help bar below the content, right in the user's direct line of sight. Third, the help content gives real-time feedback on the user's interactions; when a gesture is recognised, its icon is highlighted, as shown for the "right" gesture in Figure 3. This less intrusive method of presenting the available gestures, was shown to be effective in a recent study [12].

To navigate through the Media Ribbon's hierarchical information structure, the user needs to swipe left or swipe right for horizontal navigation within the current level of the hierarchy. The "More" gesture calls for the user to hold one arm up; this moves the current item up, displaying *more* information about that topic. Items with sub-items have the same "More" icon and text telling the user to "Hold your Arm Up to find out more" as in Figure 2. The "Back" gesture is achieved by holding one arm at a 45 degree angle from the waist, similar to the Xbox 360 pause menu gesture (Figure 3). For media such as photos or videos, users can interact to "like" them, using the "More" gesture. For such items, the arm-up icon appears indicating the number of likes and whether this user has "liked" the item (see Figure 4). A user can undo their vote, by repeating the "More" gesture.

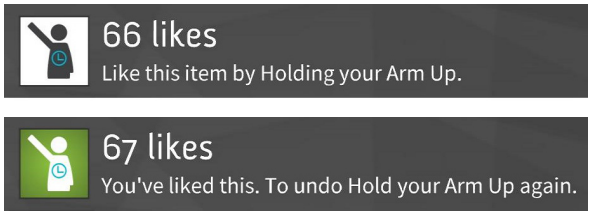


Figure 4. The Media Ribbon voting text-box, Top - Before a vote, Bottom - After a vote.

When a person first is detected by the display, they are presented with a welcome message at the top of the display, and then a tutorial. The gesture icons are initially greyed out. Text on top of the display suggests the user try the first gesture ('SWIPE your LEFT hand to navigate LEFT', see Figure 2). The tutorial walks the users through the four gestures: 'Left', 'Right', 'More', and 'Back'. The icon for the next gesture to try becomes highlighted in white. Each completed gesture shows as green. In Figure 2 the user has completed left and right gestures, is next encouraged to do the more-gesture (and yet to do the back-gesture). Once the tutorial is completed, the icons become white as in Figure 3, with the last recognised gesture in blue, giving real-time feedback, as in Figure 3 where the user has just done the right gesture.

STAGES IN DESIGNING THE GESTURES

"Media Ribbon" had to operate in a public space, often with many passers-by and some times when people lingered in the courtyard, such as during intermission or an outdoor music event in the area. Designing Media Ribbon's gesture vocabulary involved the following challenges:

1. It was crucial to make the gestures simple and memorable so that passers-by can learn and apply them easily.
2. Gestures needed to be quick to allow for fast and accurate navigation through the Media Ribbon hierarchies.
3. The gesture set had to be socially acceptable for use in a public space and within crowds.

The original design of the Media Ribbon, which we deployed for an Open Day, had a continuous series of images on a single level, requiring only the "Left" and "Right" gestures. Observations of people who used the gestures during the Open Day indicated these gestures were learnable without any intervention from the researchers. These results were also supported by the subsequent in-the-wild study [1].

We then explored the more ambitious goal of supporting browsing through a much larger hierarchically organised information space. For this, we added the "More" and "Back" gestures. Our first design for these were two-handed. The "More" gesture had the user move their hands to their chest and then forward. The matching "Back" gesture also started with hands near the chest, then moving down and sideways away from the body. These were inspired by sign language, which we hoped to be an appropriate metaphor and easy to learn. We were aware that sign language is fundamentally different from our gesture language in that people expect to invest time in learning sign language to use it for the long term; by contract our gesture language must be learnt quickly for immediate use.

The Open Day study showed the "More" and "Back" gestures were difficult for participants to learn; often one of the researchers had to intervene to explain and demonstrate the gestures. However, once they had learnt the gestures, users seemed to accept and understand how to use them for interacting with the display. We concluded that our animated instructions failed to describe the gestures well enough and that their placement at the top of the screen was a problem as most users ignored them entirely. This design also had the visual affordance of a dwell button, leading to some users to try to activate the button if they did recognise the icons. We next explored various expressive gestures and poses, such as an expressive shrug, gestures similar to the teapot [12] pose and handed gestures like those of the Chained Displays [11]. We felt that these more expressive gestures worked well in an entertainment contexts (such as games) but may not be appropriate for an information exploration context. We also needed to consider whether people would be likely to find gestures inappropriate for a public space [9] making them uncomfortable and reluctant to use them with the Media Ribbon.

Based on these findings and explorations we arrived at the current gestures described above We note that the arm-up "More" matches the familiar hands-up from the classroom, when asking a question. Another factor in our gesture design was to limit it to single arm gestures; we would expect that people often have one arm occupied, for example, carrying a drink or bag.

We then decided to add a voting gesture and reused the "More" gesture. Overloading the gesture is a clear risk for learning. But it also limits the gesture set that needs to be recognised by the system and learnt by the user. This is in line with work [3] indicating that such overloading, in distinct contexts has lower cognitive load than adding additional unique gestures.

DISCUSSION

We designed the "Media Ribbon" with just four gestures, aiming for a minimal but effective set. Our experience in the design process established the difficulty in communicating and reliably recognising more complex gestures. Both our studies [4, 1] and other work [8] indicate that people are likely to interact with these displays for very short periods, typically 30-60 seconds, making for very tight constraints in learning.

We found that our swiping gestures were easy to learn; perhaps due to the very direct mapping to the on-screen results. By contrast, out semaphoric gestures, "More" and "Back" seemed harder to learn, perhaps reflecting a mismatch between the action, description and result. In our theatre study [1], many users seemed to ignore the interactive tutorial. Our results indicate that the real-time feedback of the icons do facilitate learning the gestures, without an explicit tutorial. For our context, it also turned out to be quite important that gestures required only one hand [1].

Another important aspect of designing interactive displays is providing real-time feedback. Previous work in this area [8] has demonstrated that presenting passers-by with a silhouette

or video of themselves is more successful for attracting users than a written call to action. Our current "Media Ribbon" has a skeleton as the user representation. In future work, we plan to compare this against the effectiveness of a silhouette representation, both for learning and perceptions of the interface, particularly in terms of the sense of fun. Our skeletal representation seems important for feedback on users' interactions [4]. Equally, the changing icons of the the last recognised gesture seems important. We are also exploring additional means of feedback; as we can now recognise partial gestures, we can begin to respond and make the display make gradual changes so that a swipe left gesture will start moving the ribbon left a little early in the gesture and then continue as the user finishes the gesture; this approach may aid the learning of the up-down gestures as well. Of course, it also brings a risk of introducing confusion if very small arm movements that the user did not intended as a gesture produces an unexpected effect.

CONCLUSION

This paper has presented the design of the "Media Ribbon" mid-air gesture-controlled interface. We presented our design rational for the four gestures to navigate and interact with information hierarchies. Our in-the-wild studies provided insights into people's behaviours and expectations, beyond those possible in the lab.

Our work suggests that a small number of gestures is effective, perhaps because it reduces the cognitive load [3]. We observed that our manipulative gestures, involving swiping, were more intuitive than our semaphoric gestures, "More" and "Back". Providing real-time and responsive feedback is critical to reinforcing gestures and making users aware of the display's interactive nature. Overloading gestures helps to minimise the total number of gestures that need to be learnt and potentially reinforces them in users' memories from reuse. Environmental and social factors [9] affect the likelihood of people using the gesture vocabulary and it is important to consider even the very mundane matter of designing for use with one hand.

Our work provides a body of empirical, in-the-wild evidence about the design of learnable gestures for public display interaction for browsing through a hierarchical information space. Our mid-air gestures are quite different from most user's experience of interaction and cannot build upon existing mental models of mouse interaction.

REFERENCES

1. Ackad, C., Wasinger, R., Gluga, R., Kay, J., and Tomitsch, M. Measuring Interactivity at an Interactive Public Information Display. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration*, OzCHI '13, ACM (New York, NY, USA, 2013), 329–332.
2. Boring, S., Gehring, S., Wiethoff, A., Blöckner, A. M., Schöning, J., and Butz, A. Multi-user Interaction on Media Facades Through Live Video on Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, ACM (New York, NY, USA, 2011), 2721–2724.
3. Carrino, S., Caon, M., Khaled, O. A., Ingold, R., and Mugellini, E. Functional gestures for human-environment interaction. In *Human-Computer Interaction. Interaction Modalities and Techniques*. Springer, 2013, 167–176.
4. Grace, K., Wasinger, R., Ackad, C., Collins, A., Dawson, O., Gluga, R., Kay, J., and Tomitsch, M. Conveying Interactivity at an Interactive Public Information Display Categories and Subject Descriptors. In *Proceedings of the 2nd ACM International Symposium on Pervasive Displays*, ACM (2013), 19–24.
5. Hespanhol, L., Sogono, M. C., Wu, G., Saunders, R., and Tomitsch, M. Elastic Experiences: Designing Adaptive Interaction for Individuals and Crowds in the Public Space. In *Proceedings of the 23rd Australian Computer-Human Interaction Conference*, OzCHI '11, ACM (New York, NY, USA, 2011), 148–151.
6. Hespanhol, L., Tomitsch, M., Grace, K., Collins, A., and Kay, J. Investigating intuitiveness and effectiveness of gestures for free spatial interaction with large displays. In *Proceedings of the 2012 International Symposium on Pervasive Displays - PerDis '12*, ACM Press (New York, New York, USA, 2012), 1–6.
7. Michelis, D., and Müller, J. The Audience Funnel: Observations of Gesture Based Interaction With Multiple Large Displays in a City Center. *International Journal of Human-Computer Interaction 27*, 6 (2011), 562–579.
8. Müller, J., Walter, R., Bailly, G., Nischt, M., and Alt, F. Looking Glass: A Field Study on Noticing Interactivity of a Shop Window. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, ACM (New York, NY, USA, 2012), 297–306.
9. Rico, J., and Brewster, S. Usable gestures for mobile interfaces: evaluating social acceptability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2010), 887–896.
10. Schroeter, R., Foth, M., and Satchell, C. People, Content, Location: Sweet Spotting Urban Screens for Situated Engagement. In *Proceedings of the Designing Interactive Systems Conference*, DIS '12, ACM (New York, NY, USA, 2012), 146–155.
11. Ten Koppel, M., Bailly, G., Müller, J., and Walter, R. Chained displays: Configurations of Public Displays Can Be Used to Influence Actor-, Audience-, and Passer-By Behavior. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, CHI '12, ACM Press (New York, New York, USA, 2012), 317.
12. Walter, R., Bailly, G., and Müller, J. StrikeAPose: Revealing Mid-air Gestures on Public Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, ACM (New York, NY, USA, 2013), 841–850.

A Simple Universal Gesture Scheme for User Interfaces

Stuart K. Card
Dept. of Computer Science
Stanford, CA 94305 USA
scard at cs.stanford.edu

ABSTRACT
Gesture-based user-interfaces have become more important as computing has moved into mobile and embedded contexts, but their design is not without difficulty. On the machine side, they can be difficult to design, parse, and disambiguate. On the human side, they can be difficult to learn and remember. This paper documents a universal gesture scheme that avoids most of these problems. Though simple, the scheme enables the practical generation of moderately complex gesture user interfaces and serves as a concrete comparison and source of principles for discussing more subtle gesture designs.

Author Keywords
gestures; gesture-based UI; surface swipe gestures; mimetic gestures; description; depiction; deixis

ACM Classification Keywords
H.5.2. User Interfaces

GESTURE USER INTERFACES

Although research on gesture-based user interfaces is long-standing [7], it has recently acquired increased importance because of the development of mobile computation too small for traditional input devices and new sensors that make novel forms of gestures possible.

What is a gesture UI?

The study of gestures includes related, but sometimes conflicting, conceptions from psycholinguistics, cognitive neuroscience, biomechanics, and human-computer interaction. It is therefore useful at the start to make clear the particular notions being discussed. In this paper, the working definition of gesture is thus:

> *A gesture is the use of the body to express, communicate, or command action or meaning.*

We are concerned in this paper with gestures used directly as commands and defer consideration of gestures used to modulate speech in a multimodal context. On the machine side, the core problem for gestures as commands is to map a physical set of human body movements into a semantic set of application commands. This can be modeled as proceeding through a set of transformations from the physical properties of a gesture eventually to an application command [4] illustrated in Figure 1: A typical gesture recognition process begins with the physical sensing of the gesture. On an iPhone, for example, the screen has embedded in it two orthogonal sets of tiny wires, X and Y. When fingers are placed on the surface of the screen, the electrical charge present in the fingers changes the capacitance of the wires under the fingers. In a mutual capacitance device, such as the iPhone, the device cycles through the driving wires in the X-axis one at a time, and within each X value, it cycles through all of the Y values. In this manner, it is able to prevent ghosting and to sense the location of multiple fingers. The result is groups of pixels that have sensed the fingers. The next step is signal processing for noise reduction, clustering of the pixels, and movement detection resulting in recognition of the physical gesture. This recognized gesture must then be mapped into the command semantics of the application. It is this mapping that will concern us.

Why are gestures interesting?

Gestures are interesting for communicating with a machine because they are compact and because they provide an expressive representation for human-machine interaction. Gestures can be compact, because the same sensors may be reused for different gestures as in the iPhone screen just described. Gestures also combine command and argument in a movement-efficient way. The action (turning the page) and the argument (the particular page to be turned) are specified in the same physical motion. Biomechanically, gestures can be done with some fingers while holding the device with other fingers or, indeed, with head nods or some other exotic movement.

Gestures are also interesting because they provide a representation, beyond text and pictures, for expressing meaning [6], especially for expressing action. There are basically three ways of communicating an action: (1) *Description* is symbolic communication. (2) *Depiction* is acting out the information to be communicated. And (3) *Deixis* is communication by pointing.

Copyright is held by the author/owner(s).
Submitted to Gesture-based Interaction Design: Communication and Cognition, CHI 2014 Workshop..

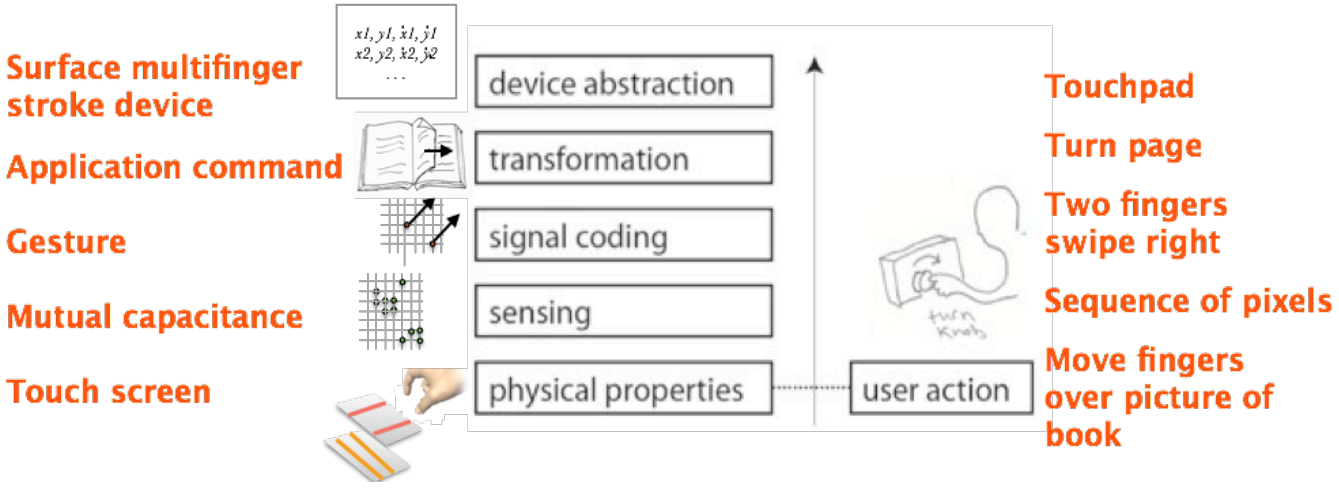


Figure 1. Gesture levels of abstraction.

Gestures allow us to recast communication from description into depiction and deixis. Instead of *telling* the system what to do by typing symbols like `Turn-page-Left`, or selecting the Turn-page-Left icon, we can simply *show* the system by doing SWIPE-LEFT with our finger across the page to be turned; that is to say, we perform a simplified version in the computer interface world of the action we would perform in the physical world. Instead of telling the system how high we want a quadcopter to fly, we can show it by holding a hand at that height. These are depiction. Deixis is the use of pointing. If we want quadcopter #2, one of a set of four, to take off, we might point to it, it lights up to acknowledge the selection, then we might make a lifting gesture, causing it to fly. Both depiction and deixis enrich our ability to represent and express our command goals and parameters.

Key problems for gesture UI's

In order for gestures to work well, there are at least five problems that need to be addressed. On the machine side, the problems are

- **Parsing** gestures into meanings with low error rates, despite variability within and between users;
- **Disambiguating** among gestures, as measured by low confusion rates.

On the human side, the problems are

- **Learning** gestures easily;
- **Remembering** gestures over time;
- **Ergonomic** design for low biomechanical strain or injury.

A SIMPLE UNIVERSAL GESTURE SCHEME (SUGS)

As an illustration of how these problems may be overcome, we now describe a simple "universal" gesture scheme (SUGS), which is illustrated schematically in Figure 2. In the scheme, the finger or mouse or arm can SWIPE or drag in only four directions, LEFT (←), RIGHT (→), UP (↑), or DOWN (↓). If the SWIPES are fast enough, they are called FLICKS (↞, ↠, ↟, ↡). SWIPES have an argument of where the finger goes down and where it comes up, which in some applications is used to drag an object to a specified place. FLICKS don't pay attention to where the fingers come up and can be used to throw an object to some automatically selected place. In the scheme, the finger can also CLICK (☟) or DOUBLE-CLICK (☟☟) a button or surface; or RUBOUT (※) by scratching on the screen. It can also ROTATE about a circle to the LEFT (↳) or RIGHT (↲) direction in a sort of screwing motion.

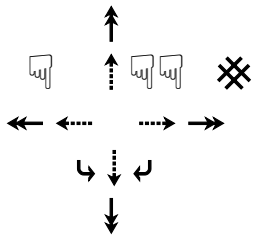
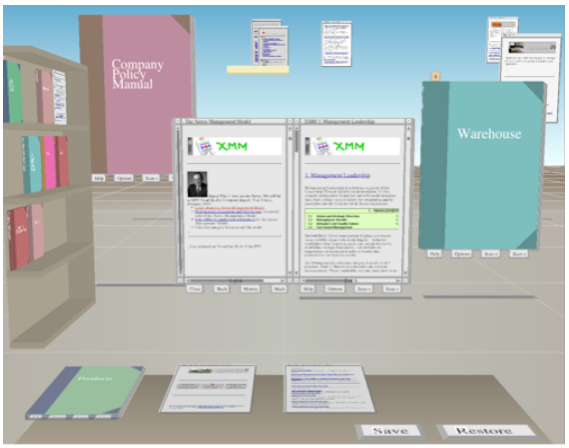


Figure 2. SUGS gesture set.

The SUGS gestures, simple as they are, are surprisingly versatile for mapping to application semantics. Figure 3 takes the schematic gestures in Figure 2 and illustrates the SUGS gesture-mapping table in pseudocode (Figure 3b) for an application called the Web Forager [2], a 3D Web workspace, browsing, and bookmarking system. CLICK turns pages. SWIPE-LEFT moves the book to a specified spot in the bookcase to the left. SWIPE–UP moves the

(a)



(b)

#	Symbol	Arg1	GES-TURE	Arg2	APPLICATION SEMANTICS
1	←---	<Obj>	LEFT-SWIPE	<Shelf place>	Move (obj:<Webbook>, to:<Shelf.n>)
2	↞	<Obj>	LEFT-FLICK	---	Move (obj:<Webbook>)
3	---→	<Obj>	RIGHT-SWIPE	---	---
4	↠	<Obj>	RIGHT-FLICK	---	---
5	↑	<Obj>	UP-SWIPE	<3D Place>	moveTo3dWorkspace (obj:<Webbook>, to:<Position.p>)
6	↟	<Obj>	UP-FLICK	---	moveTo3dWorkspace (obj:<Webbook>)
7	↓	<Obj>	DOWN-SWIPE	<Desk Place>	moveToDesk (obj:<Webbook>, to:<Desktop.p>)
8	↡	<Obj>	DOWN-FLICK	---	moveToDesk (obj:<Webbook>)
9	☟	<Obj>	CLICK	---	
10	☟☟	<Obj>	DOUBLE-CLICK	---	
11	※	<Obj>	RUBOUT	---	
12	↳	<Obj>	SPIRAL-LEFT	---	
13	↲	<Obj>	SPIRAL-RIGHT	---	

Figure 3. The Web Forager. (a) screenshot of application, (b) SUGS gesture mapping table.

(a)

(b) GRASP

#	Symbol	Arg1	GES-TURE	Arg2	APPLICATION SEMANTICS
2	↞	<2Page>	LEFT-FLICK	---	turnPageL(obj:<2Page>)
4	↠	<2Page>	RIGHT-FLICK	---	turnPageR(obj:<2Page>)
6	↟	<2Page>	UP-FLICK	---	copyPageToWorkspace (obj:<2Page>)
12	↳	<2Page>	SPIRAL-LEFT	---	zoomOut(obj:<2Page>)
13	↲	<2Page>	SPIRAL-RIGHT	---	zoomIn(obj:<2Page>)
6	↑	<Page>	UP-SWIPE	<Page>	newPile(<Page1>, <Page2>)
9	☟	<LPage>	CLICK	---	turnPageL(obj:<2Page>)
9	☟☟	<RPage>	CLICK	---	turnPageR(obj:<2Page>)

(c) WRITE

#	Sym-bol	Arg1	GES-TURE	Arg2	APPLICATION SEMANTICS
1	←---	<Word>	LEFT-SWIPE	---	Move (obj:<Webbook> to:<Shelf.n>)
2	↞	<Word>	LEFT-FLICK	---	DoubleUnder-line(obj:<Word>)
4	↠	<Word>	RIGHT-FLICK	---	SingleUnder-line(obj:<Word>)
8	↡	<Mar-gin>	DOWN-FLICK	---	drawVerticalParLine(obj: <ParagraphMargin>)
	*	<Mar-gin>	ASTERISK	---	drawAsterisk (obj:<Paragraph>, <Margin>)

Figure 4. 3Book. (a) screenshot of application, (b, c) SUGS gesture mapping tables.

book to a storage place in the 3D space. SWIPE-DOWN moves the book to a desktop. The FLICK version of each of

these gestures moves an object to a workspace around the book. Notice that not all SUGS gestures need to be mapped, and so there are blanks in the table.

CHARACTERIZING SUGS GESTURE DESIGNS

We are now in a position to characterize SUGS and why it and other gesture sets that follow similar principles might be effective.

SUGS gestures are simple. The simplicity of the SUGS gestures helps to solve the machine difficulties with gestures, since such gestures are easy to parse. Furthermore, with only four directions to disambiguate, there is little ambiguity or confusion among gestures.

The size of the SUGS gesture set is small. There are only 13 gestures and these are mostly generated from an even smaller set. This contributes to making the set easy to learn and to remember.

SUGS gestures are "universal". Instead of generating separate gestures for each application, we use the same set (or we use it as the base set with additions, if necessary). This reusability is the sense in which we mean that the SUGS gestures are universal over some modest domain rather than purpose-designed for a specific application. Figure 4 shows the SUGS gesture mapping of the same gestures as before, but for a different application, 3Book [2], a 3D book application. The gestures for 3Book are divided into Grasp and Write subsets. Grasp gestures move the book itself or its parts; Write gestures make marks on the book. The user shifts from Grasp-mode to Write-mode by holding down the ALT key. In Grasp-mode, SWIPE-LEFT or SWIPE-RIGHT turns pages (as does single clicks). SWIPE-Up slides pages out of the book into a surrounding workspace. ROTATE-LEFT zooms the text larger, ROTATE-RIGHT zooms the text smaller (the mimetic association is turning a screw to get closer or farther way). Write gestures enable a simple annotation system. SWIPE-LEFT draws a double underline under text, SWIPE-RIGHT draws a single underline, SWIPE-DOWN draws a vertical line beside the paragraph in the margin. In Figures 4b and 4c, unused SUGS gestures are omitted from the table. This application illustrates how we can augment the basic SUGS gesture set with application-specific gestures when justified. In this case, we add non-SUGS gestures to allow the user to draw LEFT and RIGHT SQUARE BRACKETS in the text and an ASTERISK in the margin. The SUGS gesture RUBOUT erases the markings. The point is, in a new application, the gestures are already known and needn't be relearned; it is only a matter of generating their associations.

SUGS gestures are mimetic. This is key. By *mimetic* we mean that association between each SUGS gesture and the semantic primitives of the application are natural and often guessable, that they are cognitively congruent in Tversky's sense [6] with what the action would be in the physical world. There are no symbolic, arbitrary associations. Instead, actions are rendered by acting out their stylized depiction and deixis. However, stroking left on either page of a double page layout might turn the page left as a memorable shortcut to the user. In this, we follow the de-tuning strategy of Boxer [3]. Boxer mapped application semantics onto boxes within boxes, even if the fit of the box metaphor was sometimes not precise. Accepting this "de-tuning" between the visual representation and the application semantics allowed the system to represent more concepts more flexibly than had the mapping been more precise. Mimetic gestures should be easy to learn and easy to remember [6]. They allow us to escape from the difficult domain of recall memory back into the easier domain of recognition memory.

The meaning of SUGS gestures is relative. SUGS gestures are interpreted relative to the virtual objects near them, not absolutely. A LEFT-SWIPE gesture over a page might turn the page, but a LEFT-SWIPE over the spine of a book, might turn the whole book in 3D.

The <arg1> column of the gesture-mapping tables Figure 4b and 4c shows the object to which that mapping entry applies in 3Book. Technically in human-computer interaction, the general version of this concept is called *input-on-output*, so we will refer to *gesture-on-output*. Gesture-on-output allows us to scale up the gesture interface to a larger number of commands, effectively using the combinatorics of the reuses of the universal gestures to chase the increasing functionality of applications without inventing new gestures.

CONCLUSION

The SUGS set of gestures spotlight an interesting part of the gesture design space. For one thing, they are designed together as a set that is easily parsable, unambiguous, yet expressive. To a surprising extent, the SUGS gestures form a universal set of gestures applicable across multiple applications or can at least form the core of such systems. Because the gesture set is small and common across applications, and especially since it is mimetic in the sense that it gives an analogical depiction of the gesture's meaning that is cognitively and biomechanically congruent with the motor actions to accomplish the action expressed, its ease of discovering, learning, and remembering are all enhanced. Because the SUGS gestures are relative to the user interface objects spatially associated with them, the same small set of gestures can scale in an application to express many meanings. We have focused on simple stroke gesture interfaces, but generalization of SUGS to embedded computing, tangible user interfaces, and other areas is straightforward.

References and Citations

1. Card, S. K., Hong, L., Mackinlay, J. D. (2004). 3Book: A 3D electronic smart Book. *AVI 2004:* 303-307.

2. Card, S. K., Robertson, G. G., and York, W. (1996). The WebBook and the Web Forager. An information workspace for the World-Wide Web. *CHI '96,* video.

3. diSessa. A.A. and Abelson, H. (1986). Boxer: A Reconstructible Computational Medium. *CACM 29*, 9, 1986. 859-868.

4. Follmer, S., Hartmann, B, and Hanrahan, P. 2009. Input Devices are like Onions: A Layered Framework for Guiding Device Designers. In Workshop of CHI.

5. Robertson, G. G., Henderson, D. A., Jr., Card, S. K. (1991). Buttons as first class objects on an X desktop. *ACM UIST '91*.

6. Tversky, B. et al (2014). Congruent gestures can promote thought. *CHI '14 Workshop on Gestures*.

7. Zhai, S. et al (2011). *Foundations and Trends in Human-Computer Interaction 5*(2):97-2005.

Using Embodied Cognition to Teach Reading Comprehension to DLLs

Andreea Danielescu, Erin Walker
School of Computing, Informatics, and Decision Systems Engineering
Arizona State University
{lavinia.danielescu, erin.a.walker}@asu.edu

Arthur Glenberg
Psychology
Arizona State University
aglenber@asu.edu

M. Adelaida Restrepo, Ashley Adams
Speech and Hearing Sciences
Arizona State University
{Laida.Restrepo, Ashley.M.Adams }@asu.edu

ABSTRACT
Dual Language Learners (DLLs) struggle with reading comprehension even more than native English speakers. Recent work in the cognitive sciences has demonstrated that reading comprehension requires the construction of mental models that link, or *ground*, the text to the reader's sensorimotor experiences. We propose the use of *Moved by Reading (MbR)*, developed to help children ground words in physical experiences, as an effective intervention for DLLs to improve their reading comprehension skills. By providing students with the ability to "act out" stories as they read them in an iPad application, our aim is to increase the likelihood that students can connect physical action to language.

Author Keywords
Touch interaction, Reading comprehension, embodied cognition.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Dual Language Learners (DLLs) in the United States struggle with reading comprehension significantly more than native English speakers, performing well below average [7]. In 2011, DLLs scored 33 points below the national average on the 4th grade NAEP reading comprehension test -- 69% of DLLs scored below basic, 31% at basic or above, and 7% at proficient (compared to only 28% below basic for non-DLLs nationally).

Copyright is held by the author/owner(s).

Submitted to Gesture-based Interaction Design: Communication and Cognition, CHI 2014 Workshop.

As Lakoff and Johnson argued, language is inherently embodied [5]. That is, language is understood when the words and phrases can be mapped or grounded in the reader's experiences, and those experiences are themselves encoded in neural systems of perception, action, and emotion. By leveraging the embodied properties of reading, we may be able to improve the reading comprehension skills of DLLs, who may struggle more than native speakers with connecting abstract English words to their experiences.

Moved by Reading (MbR) is an embodied reading comprehension intervention that uses action to help children ground nouns, verbs, and syntax in objects and movement. It has already been shown to help monolingual, English-speaking children to develop skill in grounding when reading simple narratives and mathematical story problems [5, 6]. *MbR* could also have a substantial effect on the grounding abilities of DLLs by helping them ground the stories they read in physical actions, and there is some early evidence to suggest that this is the case. Marley, Levin, and Glenberg [5] report that *MbR* helps listening comprehension for Native American children with learning disabilities. Marley, Levin, and Glenberg [5] also found benefits of *MbR* for nondisabled, third-grade Native American children living on a reservation.

EMBODIED COGNITION AND LANGUAGE
Our perceptions and the way we experience the world are influenced by our language and by our relationship to the world through our bodies. Glenberg argues that even very abstract concepts are represented in the physical body, through activation of motor circuits [4, 1]. Furthermore, language and action are intimately connected. Lakoff and Johnson argue that language is inherently embodied, from the metaphors we construct to the way we perceive the world around us [7]. For example, when children learn to count, they may gesture to different points in space to help them [5].

Language comprehension has also been shown to interfere with our abilities to detect or perform actions. For example, Meteyard, Bahrami, & Vigliocco [14] demonstrated that listening to words implying motion up or down interfered

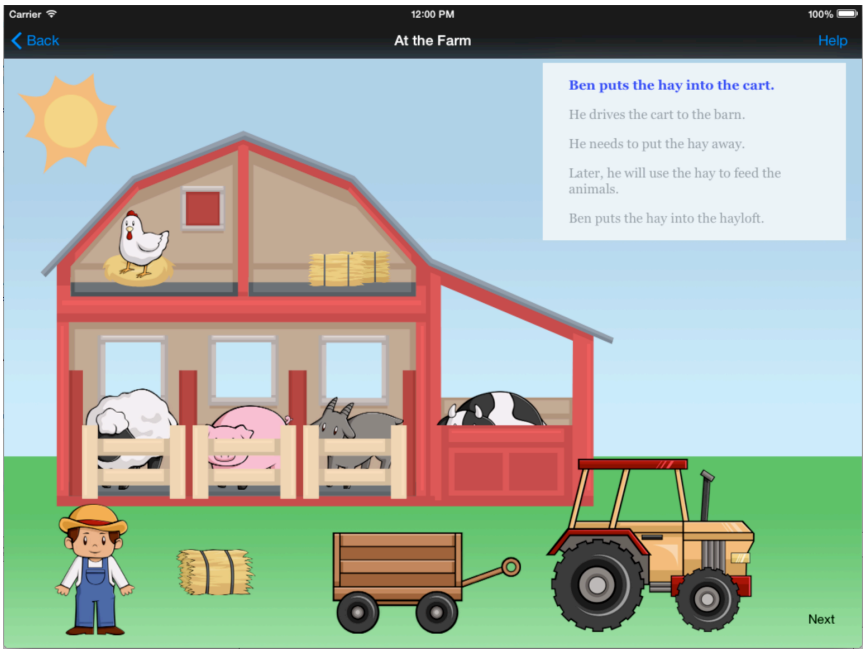


Figure 1: *Move by Reading* iPad application

with the detection of such motion in a display. Additional behavioral work (e.g. [8]) has demonstrated that understanding sentences such as "Close the drawer" (an action that typically requires moving the hand away from the body) interferes with literally moving the hand toward the body, whereas understanding sentences such as "Open the drawer" (an action that moves the hand toward the body) interferes with literally moving the hand away from the body. That is, understanding of these sentences seems to call on neural systems that control literal action.

MbR intervention is based on behavioral and neuroscience work demonstrating the link between language comprehension and neural systems of action control. For example, Hauk, Johnsrude, & Pulvermüller [9] found that listening to action verbs such as "lick," "pick," and "kick" produced activity in somatotopically mapped areas of motor cortex, whereas Aziz-Zadeh, Wilson, Rizzolatti, & Iacoboni [2] found activity in the premotor mirror neuron system during language comprehension. The use of these neural systems of action, perception, and emotion to simulate language is what we mean by grounding.

Embodied cognition, therefore, describes how attaching words to one's experiences of perception, action and emotion strengthens language comprehension. Without grounding, the words read would be meaningless.

MOVED BY READING

In *Moved by Reading,* students "act out" the stories as they read them. *MbR* has already been shown to help monolingual English speakers to develop grounding skills when reading simple narratives and mathematical story problems [5, 6]. Preliminary studies with DLL students have also shown promising results [1]. The goal of *MbR* is to scaffold reading through physical manipulation to help students develop grounding skills, which can then be used even when students cannot physically manipulate the characters in the story. To accomplish this goal, physical manipulation activities are followed by imagine manipulation activities, in which students are taught to imagine the manipulation while reading instead.

MbR has been implemented in a web-based application where students engage in a standard desktop interaction with a mouse. Action sentences – sentences that required the child to perform physical manipulations of the images on screen – are marked. As children correctly perform the actions corresponding to the current sentence, the story automatically moves on to the next sentence. Children can only perform physical manipulations in order as they relate to the current sentence.

For example, one sentence in a story about farm life is "Ben hooks the cart to the tractor." The student must first pick up Ben, the farmer, and drag him over to the cart. They cannot interact with any other image at this time. When the student drops Ben near the cart, Ben and the cart become connected to signify Ben picking up the cart. Then, the student must move Ben to the tractor. When Ben is dropped near the tractor, the cart is disconnected from Ben and is connected to the tractor instead. At this point, the students have completed the actions necessary to "act out" the current sentence, and the application moves on to the next sentence.

MOVING TO AN EXPLORATORY INTERACTION

We have re-implemented the application for touch interaction on an iPad. We constrained the current possible

Figure 2: Example of the hotspots that are displayed to the user. Blue hotspots signify that a connection would occur at that location between the two objects. Red hotspots are possible connections that could occur if the two hotspots were closer together.

gestures to moving, grouping, and ungrouping the objects that are part of the story the students are trying to understand. To move an object, students touch it with a finger and drag it around on the screen. When the finger is picked up off the surface, the object is no longer moving. The grouping gesture can be used for a large variety of verbs, including "pick up" (e.g. "the farmer picks up the hay"), "visit", "stand on", or "get in" (e.g. "the farmer gets into the tractor"). This gesture is similar to the move gesture, but it requires objects to be moved and dropped so that they are in close proximity or overlapping to each other. The final gesture, ungroup, is used for verbs such as "put down" or "get out of." To perform this gesture, a student must touch the images that should be ungrouped with two fingers, and move those fingers in opposite directions, similar to a zoom out gesture used on most touch devices. The decision to use only a few simple gestures was made in order to strike a balance between the types of verbs that we can currently support and the number of gestures students need to learn. We eventually intend to extend our system so we can broaden the types of stories we have students physically manipulate (e.g., how does the interaction scale change when we need to support other verbs such as "throw" or "pump"?).

It is important to note that the iPad application includes an exploratory component not present in the desktop-based version of *MbR*. In the iPad version, students can pick up and drag many of the objects that are not part of the background at any point in time throughout the activity. For example, the child may choose to manipulate the farmer, the hay, or even one of the pens in the story shown in Figure 1. This exploratory component will eventually tie into an intelligent tutoring system, with the goal of allowing students to explore connections between objects in the space and providing the possibility to make mistakes and learn from them. For example, one common mistake for DLL students when acting out the sentence "Ben puts the hay in the cart" is to try to pick up the hay and put it directly into the cart, without interacting with the Subject of this sentence – Ben. By allowing them to pick up the hay or Ben at any point during the interaction, we can more easily

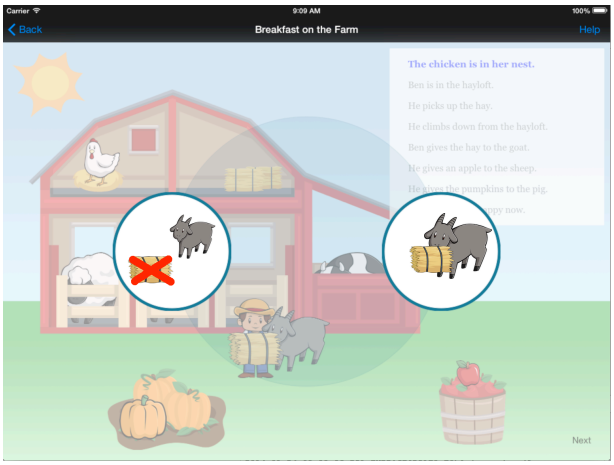


Figure 3: Example of the menu showed to the user to disambiguate possible interactions.

determine that a student is making this mistake, and develop an intervention that will help explain the misconception.

Given this exploratory component, designing a gestural touch interaction presented some challenges. The largest challenge was striking a balance between the number of gestures that were necessary to support different action verbs, and constraining the environment enough so that students are not constantly being prompted to disambiguate which action they wanted to perform. For example, the farmer may want to stand on the hay or pick up the hay. If the farmer is already holding the hay, he may want to place the hay in the cart, or put it down on the ground, or alternately he may even want to get in the cart with the hay.

We are investigating two different possibilities to identify intent in the application. One such interaction uses hotspots, visualized by small circles, to display possible groupings to the student (see Figure 2 for an example). As shown, by dragging the farmer over to the hay, the student will see a red hotspot at the farmer's feet and towards the center of his body, near his hands. Similarly, a hotspot will show up towards the top of the hay and towards the center of the hay. When two hotspots are in close proximity, the relevant hotspots signifying a possible interaction between two objects will turn blue. In Figure 2, the farmer's hands are in close proximity to the center of the hay, signifying the action of the farmer picking up the hay. This reduces the number of possible actions the students need to identify, but may still produce ambiguous situations, in which one hotspot is used for multiple action verbs. If an ambiguity occurs, a radial menu is displayed showing the end result of performing each action, as is shown in Figure 3. In this example, the farmer is giving the hay to the goat. There are two possible results of this action – the goat picks up the hay, or the goat eats the hay. The student can then touch whichever end result they were aiming for. By using touch-based gestures to act out sentences in the stories, students

may more deeply encode how the story ideas map to their own understanding of embodied actions and relationships.

A second possible interaction is the use of radial menus at all times. This would require students to explicitly recognize the action they wished to perform, which may be beneficial for learning as they view the image and map it to the sentence they have just read. This interaction style would also provide us with a simpler way to incorporate different verbs, such as throw and pump, which may be more difficult to accommodate with the interaction that uses hotspots. However, it would decrease the physical movement that students are performing themselves by spatially arranging the objects in the space, potentially limiting an embodied encoding of the action.

EVALUATION

Our intent is to evaluate the effectiveness of the two interactions we are exploring, and to compare these two interactions with a control condition. The aim is to identify whether asking students to spatially arrange the images through touch gestures improves learning gains, or if displaying a menu with pictorial representations would suffice. Additionally, we intend on evaluating whether either of the two solutions that allows for exploration of both correct and incorrect possibilities provides additional learning benefits over the physical manipulation in which the student must make the correct movement.

To investigate the usability and potential learning benefits of each kind of interaction, we are piloting two of the three conditions (the menu based condition and the control condition). Each condition will have three participants (age range of 3-5). At the beginning of the session, students will be provided with a brief introduction to all of the characters and objects on the farm. Additionally, students will go through a practice story with the experimenter in which they will be shown the gestures they can use and asked to mimic them.

Afterwards, students will go through two additional stories and will be asked to verbally let the experimenter know when they think they've performed the necessary manipulation. During the manipulation, if the child struggles for too long, or incorrectly performs the manipulation repeatedly, the experimenter will make a note of this and perform the correct manipulation for the student, so that the student can move on to the next sentence.

After each story, students will be asked questions to evaluate their comprehension level for that story. Here is an example of a question that may be asked by the experimenter: *In the beginning of the story, where did Ben put the hay?* If the child fails to answer this question, a more constrained question will be asked. For example: *In the beginning of the story, did Ben put the hay in the cart or in the tractor?* The number of correctly and incorrectly answered questions will be compared across conditions to determine learning gains, and interaction with our prototype will be video recorded. We hope to present preliminary results of the pilot at the workshop in April.

Ultimately, we intend on running a study with an additional 45 students (age range of 3-5), split across all three conditions. Each child will participate in a total of three sessions. Students will be given a pre- and post-test for reading comprehension in which the experimenter will read a story to the student and ask them to answer questions about it. The pre- and post-tests will use stories that are not used during their interaction with the iPad. Additionally, a post-test to evaluate the student's engagement will be given. Reading comprehension pre-tests will be given prior to the start of the study. Session one will consist of an introduction to the characters, a practice story and two additional stories. Session two will consist of three stories, and session three will include two more stories, the reading comprehension post-test, and a post-test to measure engagement. The number of correctly and incorrectly answered questions will be compared across conditions to determine learning gains.

CONCLUSION

Dual Language Learners struggle considerably with reading comprehension. *MbR* can be an effective intervention to help DLLs ground words in physical experiences to improve their reading comprehension skills. *MbR* has been shown to help monolingual English speakers with reading comprehension, and our studies so far have shown promising results for DLLs as well for a desktop and mouse version of the application. By moving to a touch interaction we aim to further help students ground nouns and syntax in action. We also aim to provide students with the possibility to explore the scene and allow them to make mistakes, which could further lead to learning gains.

ACKNOWLEDGMENTS

This study is supported by the National Science Foundation grant number 1324807. The opinions expressed in this paper are those of the authors and are not endorsed by the National Science Foundation.

REFERENCES

1. Adams, A., Restrepo, M. A., & Glenberg, A.M. (November, 2013). An English-Only & Bilingual Version of the Moved by Reading Intervention in an ELL Population. Poster presented at the meeting of the American Speech-Language-Hearing Association, Chicago, IL.
2. Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology*, *16*(18), 1818-1823.
3. Cattaneo, L., Barchiesi, G., Tabarelli, D., Arfeller, C., Sato, M. and Glenberg, A.M. One's motor performance predictably modulates the understanding of others'

actions through adaptation of premotor visuo-motor neurons. Social Cognitive and Affective Neuroscience 6 (2011), 301-310.

4. Cook, S. W., Mitchell, Z. and Goldin-Meadow, S. Gesturing makes learning last. Cognition 106 (2008), 1047-1058.
5. Glenberg, A.M. What Memory Is For. Behavioral and Brain Sciences 20, 1 (1997), 1-55.
6. Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and Imagined Activity Can Enhance Young Children's Reading Comprehension. Journal of Educational Psychology, 96(3), 424.
7. Glenberg, A.M. "How reading comprehension is embodied and why that matters." International Electronic Journal of Elementary Education 4.1 (2011): 5-18.
8. Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action.*Psychonomic bulletin & review*, *9*(3), 558-565.
9. Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301-307.
10. Planty, M., Hussar, W. J., and Snyder, T. D. (2009). Condition of Education 2009. National Center for Education Statistics.
11. Lakoff, G. and Johnson, M. (1980). Metaphors we live by. Chicago: University of Chicago Press.
12. Marley, S. C., Levin, J. R., and Glenberg, A. M. (2007). Improving Native American children's listening comprehension through concrete representations. Contemporary Educational Psychology, 32, 537-550.
13. Marley, S. C., Levin, J. R., and Glenberg, A. M. (2010). What cognitive benefits does an activity-based reading strategy afford young Native American readers? Journal of Experimental Education, 78, 395-417.
14. Meteyard, L., Bahrami, B. & Vigliocco, G. (2007) Motion Detection and Motion Words: Language Affects Low-Level Visual Perception. Psychological Science, 18(11), 1007-1013.

Objects as Agents: how ergotic and epistemic gestures could benefit gesture-based interaction

Chris Baber
University of Birmingham
School of Electronic, Electrical
and Computer Engineering
c.baber@bham.ac.uk
0044 121 414 3965

ABSTRACT
I am interested in the ways in which people perform everyday tasks, the ways in which computers recognize the physical activities that people perform in their everyday or working tasks, the ways in which the results of this recognition could be presented back to the person, and how this presentation could influence subsequent performance. The first of these interests encompasses the theoretical position of Distributed Cognition. This involves not only questions of how thinking can be performed through physical manipulation of objects in the world and how the physical manipulation of the world can be performed to better support thinking but also how the agency of objects in the world interact with the intention of their users. The question that I'd like to raise in this talk is why does it make sense to speak of agency in objects and what might this mean for the design of gesture-based HCI?
.

Author Keywords
Ergotic gesture; Epistemic gesture; Distributed Cognition; Activity recognition.

ACM Classification Keywords
H.1.2 User/machine systems; H.5.2 User interfaces

INTRODUCTION
The notion of gesture that used in this talk comes from [1]'s three part definition: ergotic (arising from physical activity on objects in the world); epistemic (arising from explorations of the environment); semiotic (conveying meaning). Following the general belief that gestures convey meaning, the question becomes how can 'ergotic' and 'epistemic' gestures be meaningful and, if so, to whom or what do they convey meaning?
In the 'internet of things' (where sensors are everywhere and share their impressions of the behavior of people who interact with them), it is plausible that a physical action can communicate particular meanings. For example, the simple act of opening a door has the potential to be gestural (because the action is performed by a specific person who could be identified as allowed to open the door; because the action could indicate a sequence of subsequent events such as the person entering a building in order to go to a specific office etc.).

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single-spaced in TimesNewRoman 8 point font. Please do not change or modify the size of this text box.

OBJECTS AS AGENTS
I want to start this section with a claim that at first glance might seem wilfully surreal: when you use a teaspoon to put coffee granules into a mug you are engaging in collaboration between you, the mug and the spoon. This is not simply to say that you are acting on the spoon and mug but that they are equal partners in this interaction. The reason why this might seem surreal is that it is implying some agency on the part of spoon and mug. Of course, one can imbue the spoon and mug with some degree of 'agency' by fitting them with an appropriate collection of sensors, processors and actuators so that they can make responses to the actions being performed on them. However, I want to ask why it might not be surreal to consider the conventional spoon and mug as 'agents' in the coffee spooning task.

It is a truism to claim that the design of an object reflects a particular set of socio-cultural expectations of how that object ought to be used. Following Malafouris' [2] conception of 'material engagement', one can elaborate this to imply that the interactivity between a person and the object in pursuit of a goal is not simply a matter of the person imposing meaning on the object but a reciprocal process. The object constrains, or at least influences, the actions that the person performs, the person adapts her actions to the changing state of the object, and the overall pursuit of a goal arises from these relationships. For Malafouris [2] this means that "Agency is the relational and emergent product of material engagement." (p.148), and it makes sense for both the object and the human to swap agency as the process unfolds.

To develop this point further, consider the question of how people perceive agency in the behavior of objects.

Michotte [3] asked people to report their observations when they saw a moving object (the 'launcher') hitting a stationary object (the 'target') which then started moving. If the target moved in the same direction and with the same velocity as the launcher, people spoke of the launcher *causing* the target to move (providing the time between contact and motion was negligible). Small changes in latency, velocity, direction affect reports of causality, suggesting that this is stimulus driven rather than application of theories of mechanics [4]. The launcher effect suggests that people interpret the behavior of objects *as if* they were capable of autonomy and *as if* they possessed sufficient agency to act. The explanations do not seem to involve predictive theories of causality so much as ad hoc responses to changes in state in which objects 'cause' events to occur.

In the coffee spooning task, the notion of agency of the spoon or mug might become more relevant if there was an unanticipated change in their state. For example, if the spoon collides with the rim of the coffee jar and the granules spill, one might speak of the spoon hitting the jar and spilling the granules (rather than a lack of control by the person holding the spoon). For me this raises the question of what type of event might invoke agency. By 'event', I am following Gibson [5] in viewing this as a discontinuity in the stream of information (to by the person performing the task and anyone / anything observing this action). This suggests that there are boundaries in the stream of information (or a sequence of actions) that can be responded to. These boundaries can be characterized by changes in the state of objects and by 'attractor' states in human activity. The notion of an attractor state comes from Dynamical Systems theory which posits that the number of degrees of freedom in a biomechanical system can be reduced through the development of coordinative structures, leading to preferred states through which movement sequences pass. For well-practiced movement sequences, attractors provide consistent and repeatable patterns of action. In contrast, the space between attractors can be navigated in a more flexible and adaptive manner, by the expert. Combining these notions gives an idea that different goals will involve different 'events', and different levels of expertise in the performer will lead to different 'attractors'. Such an approach to making sense of activity means that it is not sufficient simply to seek to detect specific actions but becomes crucial to more clearly delineate the ways in which activity can be considered as a Dynamical System (comprising the person and the objects

Assume that the spoon and mug are equipped with sensors that are monitoring a person's behavior, and that the person is engaging in relearning an Activity of Daily Living (ADL) that they have lost as the result of injury or illness. In this instance, there might be some benefit in providing the person with some form of feedback in terms of how they are performing the task, e.g., in terms of proximity of spoon to mug or in terms of stability of the spoon as it transports the coffee granules. Such feedback could be integrated into the spoon itself, e.g., in terms of lights in the handle changing color when the action deviates from a 'norm', or could be integrated into the workplace, e.g., through images projected onto a work-surface or displayed on a screen. In either case, the nature of the task could easily change from manipulating objects in order to perform a task into manipulating objects in order to satisfy the requirements of the recognition algorithms. In other words, the 'agency' of the physical objects, in terms of setting constraints on the performance of an ergotic task, shifts to setting constraints on the performance of a semantic task.

Figure 1. Detecting objects in a tea-making task

Figure 1 shows a table with cup, milk jug, kettle and bowls for sugar and teabags. This is for a project that we are working on that is exploring ways of supporting ADL. Each object has an instrumented coaster attached to it to detect movement and changes in weight (with the addition or removal of the object's contents). A Wii sensor unit is positioned above the table to capture hand movements. Thus, the task of making a cup of tea can be captured and the sequence of object interactions timed and recorded. In this example, the interesting features of the person's activity are related not only to which objects they interact with but also to their pauses, hesitations and changes in their intentions. A challenge for this work is how we might begin to interpret these non-specific actions and to treat them not as the 'noise' that happens between task-relevant movements but as indices of intent which could be recognized in terms of confusion, distraction or requests for assistance.

MEANING IN ERGOTIC GESTURE

The ergotic gesture involves the expenditure of kinetic energy, e.g., pushing, pulling, lifting, cutting, grasping etc. [1]. While each of these actions (with appropriate sensing capabilities) could be captured and recognized by a computer, the question remains as to what their 'gestural' valence might be.

Figure 2. Sawing a disc with a jeweler's piercing saw

I've been putting sensors on to the tools that jewelers use (figure 1). The aim has been to collect data pertaining to the practice and development of skill in jewelry making. The focus of the work is to understand how the craftworker is "…engaged in continual dialogue with materials…" ([7], p.125). Accelerometers and stain gauges on the handles of piercing saws allow us to compare different techniques in sawing metals. Good practice, when sawing a disc from a strip of metal, would be to keep the orientation of the saw as consistent as possible and to rotate the metal as it is worked (figure 2). In terms of the ergotic gestures in this task, one could seek to define the 'attractors' to which the skillful actor moves via the coordinative structures of well-practiced performance, and couple this with the state of the metal being worked (as it is cut and rotated during the task). Capturing the consistency of the actor's movement and the smoothness of the movement of the metal would allow one to appreciate the dynamics of the overall performance in achieving the goal of producing a disc. These events form the vocabulary of the dialogue between craftworker and material. This raises the question of epistemic gestures.

MEANING IN EPISTEMIC GESTURE

Lederman and Klatzy [8] identified a range of canonical actions that people perform when exploring objects. These actions could inform the design of a vocabulary to support recognition of gestures during exploratory activity, using finger or a hand-held stylus or probe [9]. The implication was that performance could be described in terms of the relationship between the effector, the nature of the textured surface, and the manner in which the effector moved across (and thus sampled) this surface. Exploratory action is directed towards supporting perception arising from that action, i.e., tactile exploration of a surface in order to report tactile perceptions of that surface. In other situations, the physical activity might be performed to support other forms of perception, e.g., brushing dust onto a glass surface in order to reveal fingerprints during crime scene examination (figure 3).

Figure 3. Dusting for fingerprints

In this instance, the role of the physical action (of brushing dust and rotating the bottle) is performed to support visual perception. The manner in which the physical actions are performed reflect understanding of the nature of the prints to be revealed and the type of visual exploration which should be performed. This reflects the distinction between epistemic and pragmatic gestures raised by Kirsh and Maglio [10], i.e., epistemic actions change the world in order to simplify a problem solving task ("actions that an agent performs to change his or her own computational state" ([10], p.514), rather than to implement a plan (pragmatic actions). While one might view the majority of human activity has comprising an epistemic component [11], it is the interactivity in the support of task performance which is relevant to the Tetris studies of Kirsh and Maglio [10]. The challenge becomes less of detecting specific actions and more of determining *which* of these actions constitute epistemic gesture (in terms of presenting the bottle to enable visual search) and which of these actions constitute pragmatic actions (in terms of identifying usable prints).

IMPLICATIONS FOR DESIGN

In terms of the design of gesture-based interactions, there are four points to develop from the ideas presented in this paper.

First, the idea that a gesture conveys meaning implies a discrete action which can be associated with a specific label. In much the same way, activity recognition often (but not always) concentrates on the detection of discrete tasks. In both cases, the aim would be to reflect instances of events in which the actor performs a defined gesture (or action) in a defined manner. The definition could be imposed by design (in which a learnt gesture set is employed) or statistical sampling (in which repeated versions of an action become associated with a label). What both approaches miss are the dynamics of performance and, to some extent, the importance of boundaries between events. So, I suggest that there is a need to think about gesture-based interaction not simply in terms of detecting specific actions but in terms of detecting variability and stability in action dynamics and in terms of event boundaries.

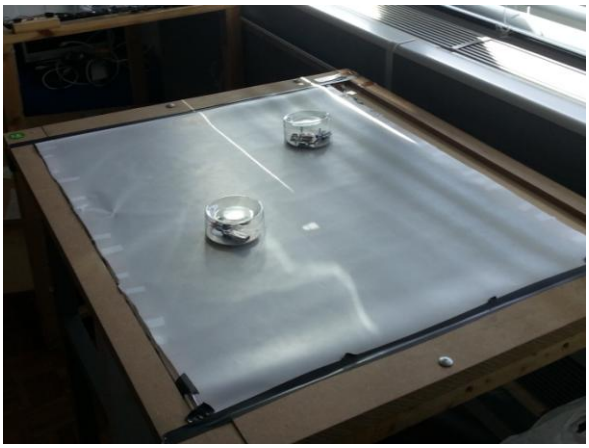


Figure 4. Instrumented objects on a multi-touch surface

Figure 4 shows two objects on a multi-touch surface. When the objects are placed on the surface, their location is registered by the computer controlling the surface and this location is transmitted to the object. When the objects are near other objects, proximity sensors turn on a light on the object's housing. The intention is to replicate the simple 'robots' developed by Brooks [11]. However, rather than having these objects behave autonomously, their movement is managed by a human player whose role is to interpret the signals from the objects and move them accordingly. By changing the rules that govern the turning on and off of lights on the objects, and by changing the communication patterns between objects, it is possible to create variations in play that create different challenges for the human. Further, the motion sensors in the objects can be defined by rules relating to the manner in which the person moves the objects, creating further variations in play. This provides a simple but effective platform to explore the interactions between people and objects in order to explore how gesture can be captured, displayed and interpreted in network of 'smart objects'.

Second, a focus on discrete action / gesture capture implies a focus on providing discrete feedback in response to the action. For example, a recognized gesture leads to a defined response by the computer or a recognized action can be defined as 'correct'. Related to the notion of dynamics, there is a need to consider how 'feedback' can be provided in a broader sense to the person, perhaps to encourage reflection on performance or to support dialogue directed at creative decision-making [12].

Third, a problem with providing the results of the recognition processes to the person can be related to earlier work on training using Knowledge of Results (KR). That work demonstrated that performance would improve when people had feedback, *but* that as soon as the feedback was removed, performance dropped down. This implies that KR changes the nature of the task. We see similar effects in speech and handwriting recognition, with people changing their 'natural' behavior to fit the demands of the technology they are using, and one would expect to see similar effects in the use of feedback to 'nudge' behavioral change.

REFERENCES

1. Cadoz, C. 1994. Le geste canal de communication homme-machine. La communication 'instrumentale', *Sciences Informatiques, numéro spécial: Interface homme-machine*. 13(1): 31-61.
2. Malafouris, L. (2013) *How Things Shape the Mind: a theory of material engagement*, Cambridge, MA: MIT Press.
3. Michotte, A. (1963) *The perception of causality* (trans. T. R. Miles & E. Miles). New York: Basic Books.
4. Scholl, B. J., & Tremoulet, P. D. (2000) Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*, 299-309.
5. Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*, New York: LEA
6. Sennett, R. (2008) *The Craftsman*, London: Penguin.
7. Lederman, S. J., & Klatzky, R. L. (1993) Extracting object properties through haptic exploration. *Acta psychologica*, *84*(1), 29-40.
8. Klatzky, R., Lederman, S.J., Hamilton, C., Grindley, M. and Swendsen, R.H. (2003) Feeling textures through a probe: effects of probe and surface geometry and exploratory factors, *Perception and Psychophysics, 65,* 613-631.
9. Kirsh, D. and Maglio, P. (1994) On distinguishing epistemic from pragmatic action, *Cognitive Science, 18,* 513-549.
10. Loader, P. (2012) The epistemic/pragmatic dichotomy, *Philosophical Explorations, 15,* 219-232.
11. Brooks, R. (1991) Intelligence without representation, *Artificial Intelligence, 47,* 139-159.
12. Pereira, A. and Tschimmel, K. (2012) The design of narrative jewelry as a perception-in-action process, *2nd International Conference on Design Creativity*.

A Cognitive Perspective on Gestures, Manipulations, and Space in Future Multi-Device Interaction

Hans-Christian Jetter
Intel ICRI Cities, University College London
Gower Street, London, WC1E 6BT, UK
h.jetter@ucl.ac.uk

ABSTRACT
In this position paper, I introduce my view of gestures, manipulations, and spatial cognition and argue why they will play a key role in future multi-device interaction. I conclude that gestural input will greatly improve how we interact with future interactive systems, provided that we fully acknowledge the benefits of manipulations vs. gestures, do not force users to interact in artificial gestural sign languages, and design for users' spatial abilities.

Author Keywords
gestures; manipulations; gesture sets; space; spatial memory; multi-device; cross-device; ad hoc.

ACM Classification Keywords
H.5.2. User Interfaces: Input devices and strategies.

INTRODUCTION
During the last decade, the rapid advances in sensor and display technology, CPUs, GPUs, and wireless networks have enabled a new generation of novel computing devices with a great variety of form factors and interaction styles, e.g., smart phones, smart TVs, mobile tablets, mobile projectors, tabletop computers, large multi-touch and pen-enabled whiteboards and tables, wearable computing such as smart watches or augmented reality glasses. Through these new devices, we have advanced from the *"personal computer era"* with *"one computer per user"* to the "*mobility era"* with *"several computers per user"* and we expect to advance to the "*ubiquity era"* with *"thousands of computers per user"* in 2020 and beyond [8].

In this coming era, the core challenge of HCI will be to understand, design, and implement user interfaces for a "natural" and efficient interaction with not only a single screen, device, or gadget. Instead we will have to focus on a seamless cross-device interaction with always changing *ad hoc* communities of several or many co-located devices. Current HCI research already reflects the critical role that a seamless interaction with multiple displays and devices will play, e.g., the use of second or third screens while viewing TV [2], cross-device interactions and collaboration using multiple co-located tablets or phones [9, 21, 22, 24, 27] and how to track their spatial configuration [10, 20, 23, 24], *proxemic interactions* based on spatial relations such as distance, orientation, and movement of devices [7], cross-device interaction in multi-tablet environments for active reading [1], or also my own work on collaborative sensemaking in multi-device environments with large screens and tabletops [12, 18]. The future will bring us an even greater wealth of different interactive devices that we will use in concert in countless different contexts within our workplaces, homes, and public spaces of our future cities.

Our challenge will be to design a seamless and efficient gestural interaction with these always changing *ad hoc* communities of interactive devices while providing the necessary flexibility, scalability, and robustness for unanticipated uses [15, 16]. I believe the key to addressing this challenge is a better understanding of and a cognitive perspective on the roles of gestures, manipulations, and the physical space in which they are performed. Therefore, in the following, I introduce my view of how these different concepts are related and I illustrate how it affects our current thinking about gestures, manipulations, and space.

THE STARTING POINT: GESTURES
There are many reasons for using gestures as user input in HCI. The most obvious ones are those based on context-specific or domain-specific requirements: For example, for a screen showing medical images in a sterile operating theatre or for a non-touch public display, gestural UIs with waving or pointing are a natural choice, because touch input has to be avoided or is impossible. Another example is the games domain for which Microsoft's Kinect demonstrates how gestural and full-body interaction of multiple co-located players can entirely change the players' experience and introduce a novel source of fun and motivation into gaming. However, another sort of argument for gestures is far less convincing and, in my opinion, does not hold after closer scrutiny: Technologists frequently claim that gestural input makes computing more "natural" by enabling

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single-spaced in TimesNewRoman 8 point font. Please do not change or modify the size of this text box.

communication with a computer the same way we also communicate with one another. I call this the *"gestures make computers more human"* hypothesis.

The flaw in this hypothesis is that the goal of letting a computer interpret natural human gesturing touches on unsolved grand challenges of computer science such as passing the Turing test or achieving strong AI. It is even more difficult than making a computer *understand* natural language. Developmental psychologist Michael Tomasello whose research is in the areas of cognitive development and comparative psychology between human and non-human primates argues that, compared with conventional human languages, gestures are very weak communicative devices, as they carry much less information "in" the communicative signal itself [32]. He illustrates this with the example of pointing at a bicycle in front of a library: When a person draws the attention of another person to the bike this way, this can mean anything ranging from "this is a nice bike" or "the library is still open" to "your ex-boyfriend is here", depending on the persons' context and shared experience. This resonates with linguistic anthropologist Charles Goodwin, for whom pointing is a *"situated interactive activity"* that is *"constituted as a meaningful act through mutual contextualization"* [6]. Consequentially for making a computer understand natural human gestures, we would have to make it understand the users' context and share experiences with the user. This is, at best, a very distant vision.

GESTURE SETS

A simple solution to this problem is to define gesture sets that constrain which gestures users can use and a system must recognize. Gesture sets assign a meaning (or function) to each gesture regardless of the users and their context. By this, gestures become unambiguous but are also reduced to mere symbols within a context-free sign language that is far less expressive than natural human gesturing and whose gestures have to be learned first. Thus, with regard to the *"making computers more human"* hypothesis, even the best gesture sets are only as close to natural human gesturing as chatbots are to understanding the meaning of natural language input, passing the Turing test, or strong AI.

We can, however, accept gesture sets as something not necessarily "natural" but as an artificial language that, if properly designed, most users will be able to adopt. Applications such as Matulic & Norrie's impressive tabletop system for document editing with pen and touch gestures demonstrates the great potential of application-specific gesture sets [25]. In the best case, the interaction with such gesture sets achieves a yet unequalled feeling of flow, control, and directness. In a mediocre case, users only save a few seconds (assuming that a gesture is faster than selecting a menu item or recalling and entering keyboard shortcuts), users are not interrupted (assuming that there is no need to switch between mouse/touchpad/keyboard anymore), and UI designers can save screen estate by leaving out menus or other administrative controls. However, in the worst case, discovering functions becomes guesswork and learning and remembering gestures turns out to be as difficult as using the keyboard shortcuts or command line languages, in particular if they are not used frequently. Often it is simply not possible to design a clear, unmistakable, and easy-to-remember mapping between commands and gestures for all application scenarios. This also shows in the relatively small agreement rates when multi-touch and/or pen gestures for tabletops or multi-display environments are elicited from users [26, 27, 30, 34]. In my eyes, it is therefore highly unlikely that a single self-explanatory, easy-to-learn standard gesture set is possible at all, at least unless future UIs are redesigned to be mainly controlled by *manipulations* not gestures.

FROM GESTURES TO MANIPULATIONS

Manipulations are a class of gestural input that is apparently easier to learn and to agree on. For example, in [26], there was a *"a clear trend towards higher agreement scores on actions that could be performed through direct manipulation and lower agreement scores on actions that were symbolic in nature"*. This resonates with George & Blake [5] who differentiate between two classes of gestural input: *gestures* and *manipulations*. For them, gestures are *"symbolic interactions"* for *"discrete, indirect, intelligent interaction"* while manipulations are *"literal interactions"* for *"continuous, direct, environmental interaction"*. Examples for *gestures* are symbolic stroke gestures such as "✓", "✗" for accepting or rejecting an item, or Apple's three-finger-tap to do a lookup on the word under your cursor. *Manipulations*, however, are continuous actions in space, e.g., dragging or flicking of an object across the screen, pinch-to-zoom, or two-finger-rotate. In [13], I also argue for this dichotomy of gestures vs. manipulations and that using too many gestures could result in a pseudo-natural UI which is close to a command line interface. My argument was that successful gestural input (e.g. the very popular pinch-to-zoom) is actually mimicking how physical objects could be manipulated in the real-world and that this can be explained with Hutchins et al.'s classic cognitive account of *direct manipulation* with two major metaphors for the nature of human-computer interaction: the *conversation metaphor* vs. the *model-world metaphor* [11].

The *conversation metaphor* implies that the user interface of a computer system is a language medium and that interacting with a computer means that users can converse with the system and tell it what to do, ideally in a natural way. However, in many cases, a conversation about what should happen is inefficient compared to direct action or direct manipulation to make it happen. Just imagine how cumbersome it would be to drive a car from the backseat by having to tell the driver about every necessary action. This becomes even more cumbersome, if we first have to learn the vocabulary and language of the driver. The opposed *model-world metaphor* for the nature human-computer

interaction is not based on the idea of a conversation [11]: *"In a system built on the model-world metaphor, the interface is itself a world where the user can act, and which changes state in response to user actions. The world of interest is explicitly represented and there is no intermediary between user and world. Appropriate use of the model-world metaphor can create the sensation in the user of acting upon the objects of the task domain themselves".* Using this metaphor, Hutchins et al. explain how *direct manipulation* user interfaces (e.g. the GUI) replaced the command line by making better use of the users' cognitive resources and their perceptual, spatial, and motor skills and relying on physical action with immediate visual feedback instead of conversation. *Direct manipulation* reduces the cognitive distance (the *gulfs of execution* and *evaluation*) between the forms of user input and system output. Users get the feeling of directness from the commitment of fewer cognitive resources.

SPACES OF BLENDED INTERACTION

It is important to notice that *direct manipulation* and the *model-world metaphor* are by no means restricted to mouse-operated GUIs with real-world metaphors like the "desktop". Already in the 1980s, *direct manipulation* was realized with different input devices such as pens or game paddles and in many "non-real-world" application or game UIs [31]. In the coming era of ubiquity, the principle of *direct manipulation* will extend beyond the boundaries of a few desktop or mobile devices into our entire environment that increasingly will be augmented with touch, gesture, and motion detection, tangible user interface elements, flexible displays, ubiquitous projectors, and many other sorts of interactive or "smart" objects connected by the so-called "Internet of Things". The world itself becomes one large model-world user interface and it will enable gestures and manipulations not only across devices but also across the physical and digital realms, e.g., picking up the content of a physical note by touching it with a finger and then pasting its content on a tablet or smart phone with another touch.

In [17], I provide a novel and more accurate description of the nature of human-computer interaction in such spaces called *Blended Interaction* that is based on recent findings from embodied cognition and cognitive linguistics. It uses Fauconnier and Turner's conceptual blends and conceptual integration [4] to explain how users always rely on familiar and real-world concepts whenever they learn to use new digital technologies. Designers should consider using and blending the vast amount of concepts that we as humans share due to the similarities of our bodies, our early upbringing, and our sensorimotor experiences of the world before resorting to elaborate conscious analogies such as the desktop metaphor. Similar to Dourish's embodied interaction [3], *Blended Interaction* draws on the way the everyday world works or, perhaps more accurately, the ways we experience the everyday world, instead of drawing on seemingly familiar artifacts.

SPACE AND ACTION

A good starting point for *Blended Interaction* is our shared experience and awareness of space and our skills to act and navigate in it. However, according to embodied cognition, spatial cognition happens not only in our head but is inseparable from our actions, gestures, manipulations, or movements in space.

For example, Wesp et al. [33] found that, unlike the commonly held belief that the sole role of gestures is to communicate meaning, gestures also serve a cognitive function and help speakers to maintain spatial images in short-term memory. This raises the question if HCI can use gestural input to help users better memorize locations or states of objects. We therefore conducted a study of gestural vs. mouse input for panning a map-like UI. We found that the accuracy of memorized object locations in the map was significantly better when the map was panned with touch instead of mouse [14]. We assume that the proprioceptive feedback during touch navigation with a 1:1 control-display ratio supported the encoding of locations in spatial memory. This effect cannot be observed without a 1:1 ratio, e.g., when using a mouse or when users use zooming in addition to panning. In a further study, we also investigated body movements for peephole map navigation [29]. We found that users are able to physically navigate a large map (292x82cm) with just a small tablet-sized peephole (23.5x13.2cm) with almost the same efficiency as when seeing the entire map. We attribute this to the proprioceptive feedback of peephole navigation and the absolute mapping between physical and map space. However, we found no significant difference in spatial memory performance [28]. This demonstrates both the potential but also the difficulty of understanding the interplay between physical action, space, and cognition.

CONCLUSION

In [19], David Kirsh describes the great potential of understanding the complex interplay between cognition, space, and physical action for interaction design. I fully agree with his vision of *"cognitively informed designers"*. I believe that gestural input will greatly improve how we interact with future interactive systems, provided that a new generation of *"cognitively informed designers"* fully acknowledges the benefits of manipulations vs. gestures, does not force users to interact with systems using artificial sign languages, and designs for users' spatial abilities.

REFERENCES

1. Chen, N., Guimbretiere, F. and Sellen, A. Designing a multi-slate reading environment to support active reading activities. *ToCHI 19*, 3 (2012), 1-35.
2. Courtois, C., Schuurman, D. and Marez, L. D. Triple screen viewing practices: diversification or compartmentalization? *In Proc. EuroITV*, ACM (2011).
3. Dourish, P. *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press, 2004.

4. Fauconnier, G. and Turner, M. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books, 2002.
5. George, R. and Blake, J. Object, Containes, Gestures, and Manipulations: Universal Foundational Metaphors of Natural User Interfaces. *In Natural User Interfaces (a CHI 2010 Workshop)*, Atlanta, USA, 2010.
6. Goodwin, C. Pointing as Situated Practice. *In Kita, S. (ed.) Pointing: Where Language, Culture, and Cognition Meet*, Lawrence Erlbaum (2003), 217-243.
7. Greenberg, S., Marquardt, N. and Ballendat, T., Diaz-Marino, R., Wang, M. Proxemic interactions: the new ubicomp? *interactions 18*, 1 (2011), 42-50.
8. Harper, R., Rodden, T., Rogers, Y. and Sellen, A. *Being human: Human-computer interaction in the year 2020*. Microsoft Research Ltd, Cambridge, England, 2008.
9. Hinckley, K. Synchronous gestures for multiple persons and computers. In Proc. UIST '03. ACM (2003).
10. Huang, D.-Y., Lin, C.-P., et al., MagMobile: enhancing social interactions with rapid view-stitching games of mobile devices. *Proc MUM '12*, ACM (2012).
11. Hutchins, E. L., Hollan, J. D. and Norman, D. A. Direct manipulation interfaces. *Human–Computer Interaction 1*, (1985), 311-338.
12. Jetter, H.-C. Design and Implementation of Post-WIMP Interactive Spaces with the ZOIL Paradigm. PhD Thesis, University of Konstanz, 2013.
13. Jetter, H.-C., Gerken, J. and Reiterer, H. Natural User Interfaces: Why We Need Better Model-Worlds, Not Better Gestures. *In Natural User Interfaces (a CHI 2010 Workshop)*, Atlanta, USA, 2010.
14. Jetter, H.-C., Leifert, S., Gerken, J., Schubert, S. and Reiterer, H. Does (Multi-)Touch Aid Users' Spatial Memory and Navigation in 'Panning' and in 'Zooming & Panning' UIs? *In Proc. AVI '12*. ACM (2012), 83-90.
15. Jetter, H.-C., and Rädle, R. Visual and Functional Adaptation in Ad-hoc Communities of Devices. *In Visual Adaptation of Interfaces (Workshop ITS '13)*, St. Andrews, Scotland, 2013.
16. Jetter, H.-C. and Reiterer, H. Self-Organizing User Interfaces: Envisioning the Future of Ubicomp UIs. *In Blended Interaction (a CHI 2013 Workshop,)* Paris, France, 2013).
17. Jetter, H.-C., Reiterer, H. and Geyer, F. Blended Interaction: understanding natural human–computer interaction in post-WIMP interactive spaces. *Pers. and Ubiq. Comp.*, DOI 10.1007/s00779-013-0725-42013).
18. Jetter, H.-C., Zöllner, M., Gerken, J. and Reiterer, H. Design and Implementation of Post-WIMP Distributed User Interfaces with ZOIL. *International Journal of Human-Computer Interaction, 28*, 11 (2012), 737-747.
19. Kirsh, D. Embodied cognition and the magical future of interaction design. *ToCHI, 20*, 1 (2013), 1-30.
20. Li, M. and Kobbelt, L. Dynamic tiling display: building an interactive display surface using multiple mobile devices. *In Proc MUM '12*. ACM (2012).
21. Lucero, A., Holopainen, J. and Jokela, T. Pass-them-around: collaborative use of mobile phones for photo sharing. *In Proc CHI '11*, ACM (2011), 1787-1796.
22. Lucero, A., Jones, M., Jokela, T. and Robinson, S. Mobile collocated interactions: taking an offline break together. *interactions, 20*, 2 (2013), 26-32.
23. Marquardt, N., Diaz-Marino, R., Boring, S. and Greenberg, S. The proximity toolkit: prototyping proxemic interactions in ubiquitous computing ecologies. *In Proc UIST '11*. ACM (2011), 315-326.
24. Marquardt, N., Hinckley, K. and Greenberg, S. Cross-device interaction via micro-mobility and f-formations. *In Proc UIST '12*. ACM (2012), 13-22.
25. Matulic, F. and Norrie, M. C. Pen and touch gestural environment for document editing on interactive tabletops. *In Proc ITS '13*. ACM (2013), 41-50.
26. Mauney, D., Howarth, J., Wirtanen, A. and Capra, M. Cultural similarities and differences in user-defined gestures for touchscreen user interfaces. *In CHI EA '10*. ACM (2010), 4015-4020.
27. Ohta, T. and Tanaka, J. Pinch: An Interface That Relates Applications on Multiple Touch-Screen by 'Pinching' Gesture. *In Proc ACE '12*. Springer (2012), 320-335.
28. Rädle, R., Jetter, H.-C., Butscher, S. and Reiterer, H. The Effect of Egocentric Body Movements on Users' Navigation Performance and Spatial Memory in ZUIs. *In Proc ITS '13*. ACM (2013), 23-32.
29. Rädle, R., Jetter, H.-C., Müller, J. and Reiterer, H. Bigger is not always better: Display Size, Performance, and Task Load during Peephole Map Navigation. *to appear in Proc CHI '14*. ACM (2014).
30. Seyed, T., Burns, C., Sousa, M. C., Maurer, F. and Tang, A. Eliciting usable gestures for multi-display environments. In Proc ITS '12. ACM (2012), 41-50.
31. Shneiderman, B. The future of interactive systems and the emergence of direct manipulation. *Behaviour & Information Technology, 1* (1982), 237-256.
32. Tomasello, M. *Origins of Human Communication*. MIT Press, Cambridge, MA, USA, 2008.
33. Wesp, R., Hesse, J., Keutmann, D. and Wheaton, K. Gestures Maintain Spatial Imagery. The *American Journal of Psychology, 114*, 4 (2001), 591-600.
34. Wobbrock, J. O., Morris, M. R. and Wilson, A. D. User-defined gestures for surface computing. *In Proc. CHI' 09*. ACM (2009), 1083-1092.

Design of a Portable Gesture-Controlled Information Display

Sebastian Loehmann, Doris Hausen, Benjamin Bisinger, Leonhard Mertl
University of Munich (LMU), HCI Group, Amalienstrasse 17, 80333 Muenchen, Germany
sebastian.loehmann@ifi.lmu.de, doris.hausen@ifi.lmu.de,
bisinger@cip.ifi.lmu.de, mertll@cip.ifi.lmu.de

ABSTRACT
Stopping by at somebody's office can be frustrating if the required person is absent. To offer visitors additional information in this case, we built a gesture controlled public display. We applied a user-centered design approach and as first step evaluated basic parameters, such as desired tracking area and preferred gestures. We incorporated these results in a standalone working prototype and achieved natural, intuitive gestures with a recognition rate above 80%.

Author Keywords
Freehand gestures, public display, user-centered design

INTRODUCTION
Students frequently come to our research lab with question about open topics for their bachelor thesis and visitors come by to chat about our research projects. However, oftentimes we are in a meeting or not in our office at all. Therefore, we propose to install a small public screen at the office door, displaying relevant information about staff members. For interactions with the display, we opted for freehand gestures over touch because this (1) keeps the display clean and without smudge, avoiding the need to constantly clean it; (2) considers hygiene issues and (3) allows us to explore this design space. In this paper, we give first insights into our design process, the definition of a gesture set and a preliminary prototype (see Figure 1). We follow a user-centered approach, involving users early in the development.

RELATED WORK
Gestural interaction with large screens has been introduced by Bolt in the late 1970s to support voice control of user interfaces, concluding "a gain in naturalness and economy of expression" [2].

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single-spaced in TimesNewRoman 8 point font. Please do not change or modify the size of this text box.

Figure 1. Prototype of a gesture-controlled information display which can be placed next to office doors.

Two decades later, CHARADE [1] allowed for the manipulation of presentation slides by performing only gestures, underlining the support of direct manipulation of user interface elements. In recent years, several studies on freehand interaction with large displays, such as distant pointing and clicking [8], have been conducted. Wachs et al. [9] provide an overview of pros and cons, tracking technologies and application areas for freehand gestures.

Moreover, several projects focused on smaller information displays on office doors. They offer visitors location and calendar details of the owner, the possibility to leave messages and retrieve private information after authenticating themselves [6]. Cheverest et al. [3] installed small information screens at their office doors and were able to send short messages, which were then visible on their display.

PRE-STUDY: GESTURE SET AND DESIGN SPACE
Following a user-centered approach, we involved users very early in the design process. In a pre-study, before starting the implementation of the system, we observed and interviewed 16 students and staff members while interacting with a screen using freehand gestures.

Setup
In order to avoid possible distractions, we used a picture frame as dummy (see Figure 2), representing the tablet computer we used for later prototyping. First, we asked participants to take the frame and place it on the wall at a height that seemed comfortable for them. Then, we invited

them to navigate in a virtual picture gallery by asking to "go to the next picture" or "scroll through the selection of pictures". We requested participants to only use touchless hand and arm movements that they consider appropriate for the current task.

During the study, we recorded all interactions, including the placement of the display dummy, on video from two perspectives. Through post-study video analysis, we were able to measure (1) the distances of the person and the hand to the display, (2) the interaction space used by each participant for each gesture, (3) the actual gestures they used as well as (4) the time they needed to perform each gesture. In order to achieve 'natural' results, we did not inform participants about our intention to take these measurements.

Figure 2. Dummy frame used for the pre-study to evaluate how participants want to interact with such a display.

Results

16 students and researchers from our lab, seven of them female, participated in our study. They were between 22 and 30 years old, with an average of 25 years. Six participants had at least some experience with freehand gestures before the study, mostly due to using the Microsoft Kinect for gaming. Four of them were PhD students from our lab, who dealt with gestures in their research on a theoretical and practical level at some point of their research. All participants owned a smartphone or tablet and used touch gestures daily.

As Figure 3 shows, the average distance between display and upper body was 51 cm. Within this range, 27 cm were actually used to perform gestures, with a minimal distance of 5 and a maximum distance of 32 cm to the display. 87% of all gestures were performed with a minimal distance of 15 cm. While executing gestures, the hand exceeded all four sides of the display by 10 cm on average. The duration of performing a gesture was 247 ms on average (minimum: 150 ms; maximum: 400 ms). Participants placed the display in an average height of 88% relative to their own size.

For both, vertical and horizontal, 94% of the performed gestures were swiping gestures in the corresponding direction. Qualitative results from the interviews confirmed this trend and revealed the relation to swiping gestures used for touch devices.

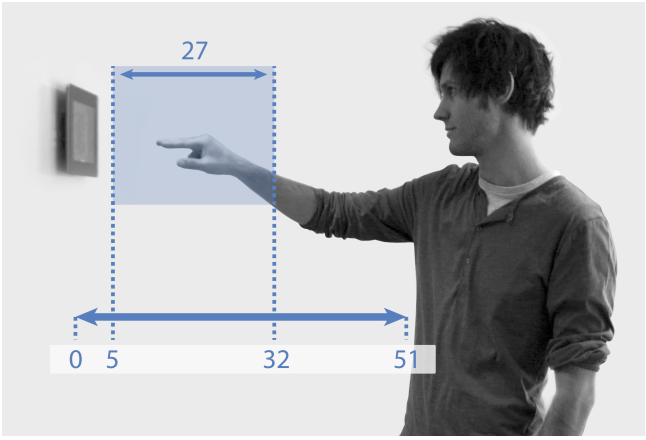


Figure 3. Preferred interaction space as found in the pre-study (in centimeters).

PROTOTYPE

Taking results from the pre-study into account, we built a first prototype. Our goal was a portable stand-alone system without a connection to a computer.

Hardware

The main component of the prototype is a 7" tablet PC running Android. To track hand movements in front of the display, we use Sharp infrared distance sensors, three on each side [5]. We chose to use six sensors to cover the vertical range of performed gestures during the pre-study: some participants performed the gestures right in front of the display, others just on the lower edge (probably trying to not occlude the display contents). To cover the horizontal area in which gestures were performed during the pre-study, we installed the sensors at an angle of about 45° towards the display's center (see Figure 4).

To compensate the jitter of the sensor values, we use capacitors and resistors in terms of hardware as well as a filter in our software.

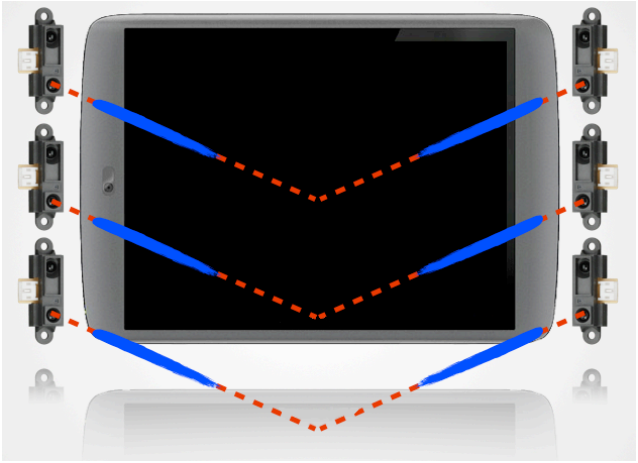


Figure 4. Arrangement of six infrared distance sensors around the display. Sensors are installed at an angle of about 45°.

To read and process the data produced by the sensors, we added an Arduino Mega ADK microcontroller board, which is connected to the tablet via USB. Any Android application can now access the data communicated by the Arduino via the Android Open Accessory protocol.

Housing

Using acrylic glass and a laser cutter, we built a chassis for the hardware. The front covers the display, which can therefore not be touched. A tinted foil, which does not block infrared light, hides the sensors. The Arduino and all wires are attached to the back wall of the housing and are thus not visible either. The only visible chord powers the Arduino and the tablet. Consequently, the prototype is portable and can be placed in arbitrary locations. Prototype can easily be fixed to any wall using power strips.

Software

The sensors can recognize objects that are between 4 and 30 cm away, receiving values between 0 and 650. By using a formula (a non-linear function), the values can be converted into the actual distance.

The Arduino reads values from all six sensors at a frame rate of 25 Hz. For each frame, a vector of six values (one depth sample per sensor) is sent to the tablet. An Android application implements the Dynamic Time Warping [7] algorithm to classify gestures executed in different speeds, as seen in the pre-study. To determine if one of four possible swiping gestures was performed, the DTW compares the data to recorded values of 64 sample gestures, 16 for each gesture, calculating the distance based nearest neighbor. Finally, the application updates the user interface according to the identified gesture.

STUDY: EVALUATING THE IMPLEMENTATION

In order to evaluate the prototype and to collect data for further development, we conducted a user study with 20 students and staff members. Please note that five of the staff members have already participated in the preliminary study. While one could argue that this bias can potentially influence the results, we think that there is no bias because (1) we did not inform participants about the results of the study or the following implementation of the system and (2) in the first study we used a nonfunctional display dummy in contrast to the fully functional interface in this study. Additionally, due to the user-centered approach, it was our purpose to include staff members into both studies as they will be future users of the system.

Setup

We implemented two applications: One showing a basic interface representing three staff members (see Figure 5) and information on their research topics.

Horizontal gestures navigate the staff and vertical gestures navigate the staff's information. We used this app for training purposes and to collect qualitative feedback on the look-and-feel of the prototype. We told participants that the interaction with the display works contactless, but not which gestures they would be able to use. They had now five minutes time to explore the app content.

The second application was implemented to measure the tracking accuracy of the prototype. We displayed arrows and asked participants to perform the swipe gesture in the corresponding direction. Each participant performed each swipe 20 times, for a total of 80 gestures per user, in a randomized order. After each gesture, we gave visual feedback by moving the content out of sight in the direction of the recognized swipe. After the study, we analyzed recorded videos to calculate tracking accuracies and a confusion matrix.

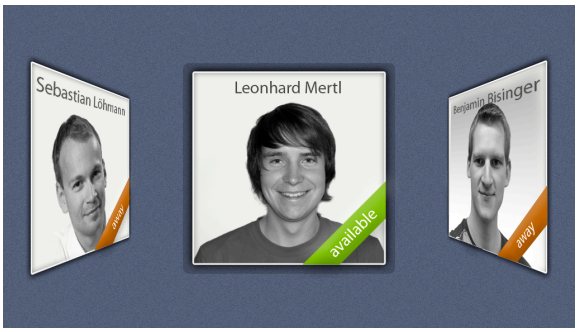


Figure 5. User interface for the first task of the user study. Horizontal swiping: Navigation through staff members. Vertical Swiping: Information about their research projects.

Results

20 students and staff members, seven of them female participated in our study. They were between 20 and 32 years old, with an average age of 25. Two of them were left-handed. Nine participants were familiar with the Microsoft Kinect for gaming.

The analysis of 1600 recorded gestures (20 participants x 4 gestures x 20 trials per gesture) lead to an average recognition rate slightly above 80%. Two participants, using gestural interfaces on a regular basis achieved higher rates of 88% and 90%. Recognition rates for each gesture were 81% for a swipe to the left, 85% to the right, 72% up and 84% down (see Figure 6).

Discussion & Implications

Our design process proved to be valuable, as participants – without instructions – immediately started to interact with the prototype using swiping gestures. We further believe that recognition rates were positively influenced by first exploring a suitable interaction space in front of the display. Looking closer at the average recognition rate for each individual trial from first to last performed gesture, the results improve from 75% to 85%. Thus, shortcomings of the gesture tracking can partly be compensated by the training effect.

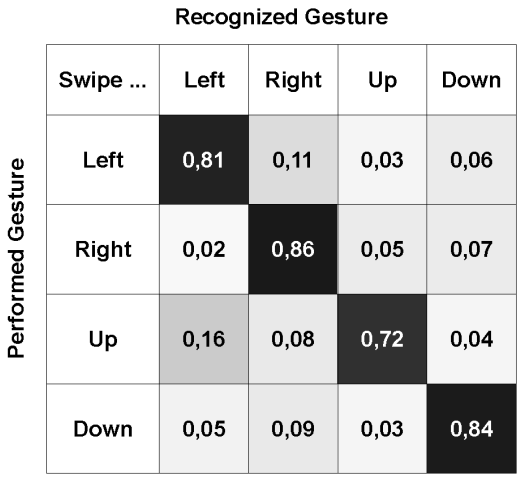


Figure 6. Confusion Matrix showing recognition rates for all four swiping gestures - left, right, up and down.

The relatively low recognition rate of 72% for the up gesture is due to the forearm that is still in the tracked area when the movement is in fact already completed, causing segmentation problems. Gestures that were performed rather fast resulted in false recognition due to the maximum sampling rate of the sensors. Generally, observations made during the exploration of the design space can be in conflict with constraints caused by the tracking technology, indicating a need to find trade-offs between both when prototyping gestural interfaces.

Another interesting observation was the influence of the distance between participants' hands and the display. Similar to the trade-off mentioned above, some users positioned their body unexpectedly far away from the prototype, causing the tracking to fail. Interestingly, they seemed to notice this issue without further clues and started to approach the display until they noticed a reaction of the system. We like to conduct further studies in order to find out how users explore the functionality of such a system. We also conclude that affordances need to be implemented, showing how and which gestures can be performed.

FUTURE WORK

Next steps include the improvement of the hardware. Due to the way we implemented the tracking algorithm, we can reorder the sensors by placing one above and one below the display to receive higher recognition rates, without the need to change the software. Taking new tracking technology into account, we are trying to replace the distance sensors with the Leap Motion Controller. This sensor is suitable for gesture tracking, but the current incompatibility between Leap and Android contradicts our design goal of a standalone device. As another alternative, we investigated the potential of the camera integrated into the tablet. When testing different libraries for gesture tracking using this approach, we noticed that (1) the tracking is unreliable as soon as other body parts (like the user's head) is visible and moving and (2) this kind of tracking becomes rather complex when extending the gesture set beyond swiping and thus causes delays due to the limited computing power of the tablet.

We will add further gestures that came up during the pre-study and are interested in how this influences recognition accuracy with different tracking technologies. Additionally, we will work on a more mature UI and add a wireless connection to the application to be able to add content on the fly. Considering the UI, a major issue will be the 'findability' of the possible gestures. In a new version of the interface, we try small icons (e.g. arrows) on the edges of the screen to indicate, where and how additional content can be obtained. Another challenge is to show first time users that the display can be controlled via gestures instead of touch. One approach is to utilize the distance of the hand to the display: when the hand gets to close, we progressively dim the display until the contents become invisible as soon as the display is touched.

We finally plan a long-term study by installing the prototype at our office door for several months to explore how users interact with the interface and its contents.

REFERENCES

1. Baudel, T., and Beaudouin-Lafon, M. Charade: remote control of objects using free-hand gestures. Comm. of the ACM 36, 7 (1993), 28-35.
2. Bolt, R. A. "Put-that-there": Voice and gesture at the graphics interface. ACM SIGGRAPH Comput. Graph. 14, 3 (1980), 262-270.
3. Cheverst, K., Dix, A., Fitton, D., Friday, A., and Rouncefield, M. Exploring the utility of remote messaging and situated office door displays. MobileHCI, (2003), 336-341.
4. Fitton, D., Cheverst, K., Kray, C., Dix, A., Rouncefield, M., and Saslis-Lagoudakis, G. Rapid prototyping and user-centered design of interactive display-based systems. Pervasive Computing 4, 4 (2005), 58-66.
5. Kratz, S., and Rohs, M. HoverFlow: expanding the design space of around-device interaction. MobileHCI, (2009), 8 pages.
6. Nguyen, D. H., Tullio, J., Drewes, T. and Mynatt, E. D. Dynamic Door Displays. Georgia Institute of Technology GVU Technical Report GIT-GVU-00-30, (2000).
7. Salvador, S., and Chan, P. Toward accurate dynamic time warping in linear time and space. Intelligent Data Analysis 11, 5 (2007), 561-580.
8. Vogel, D., and Balakrishnan, R. Distant freehand pointing and clicking on very large, high resolution displays. UIST, (2005), 33-42.
9. Wachs, J. P., Kölsch, M., Stern, H., & Edan, Y. (2011). Vision-based hand-gesture applications. Comm. of the ACM 54, 2 (2005), 60-71.

Gesture Design and Feasibility in Emergency Response Environments

Francisco Marinho Rodrigues, Teddy Seyed, Apoorve Chokshi, Frank Maurer
University of Calgary, Department of Computer Science
2500 University Drive NW, Calgary, Alberta, T2N 1N4
{fm.rodrigues,teddy.seyed,apoorve.chokshi,frank.maurer}@ucalgary.ca

ABSTRACT
Emergency response planning is a complicated process, with different forms of communication and information being exchanged amongst emergency response personnel during the course of an emergency. This information and communication exchange, led to the development of our multi-surface emergency response-planning prototype, to address communication and information sharing during emergency response situations. In this paper, we briefly describe the emergency response planning domain and how it informed the design of our prototype. We also discuss our preliminary findings on gesture design and feasibility of multi-surface interactions in the emergency response domain.

Author Keywords
Multi-surface environments; gestures; cross-device interaction; mobile devices; emergency response environments;

ACM Classification Keywords
H.5. 2 [Information interfaces and presentation]: User Interfaces— Input Devices and Strategies; Interaction Styles.

INTRODUCTION
When dealing with either real life emergencies, simulations or the planning of emergencies, Emergency Operation Centres (EOC) use numerous information sources. These information sources can come from sources that use their own unique protocols for information exchange. For example, during an emergency, a chain-of-command communication system is used by emergency response personnel to exchange information (e.g. firefighters, police, HAZMAT, EMS, armed forces), journalists (print, television and radio) utilize fact-checked information prior to broadcast, and citizens can live-tweet, post updates, or send emails.

Figure 1. The ePlan Multi-Surface Emergency Response Environment.

These different information sources have a significant impact in the workflow of emergency response personnel in the EOC. For example, traffic cameras can be live-streamed into the EOC, while tweets can arrive via text. The different information sources can appear on different screens in an EOC, which is typically driven by keyboard and mouse-based interactions with several large screens. EOC personnel continually need to examine up-to-date information on these screens and quickly determine its importance to an ongoing emergency, and in many situations, information is also shared amongst the EOC personnel, who are a group consisting of different backgrounds (e.g. firefighter, police).

Multi-surface environments, environments that utilize numerous different interactive devices (e.g. tablets, wall displays, and digital tabletops) provide an environment amenable for emergency response planning. Using concepts such as private and public spaces, and interactions that are possible with multi-surface environments (e.g. flicking, pouring), tasks such as information separation and segmentation for information triage in emergency planning is possible. In this paper, we present our ePlan multi-surface prototype, where iPads are used as personal workspaces from which EOC personnel can privately communicate with colleagues, a digital tabletop can be utilized as a collaboration and cooperation area, and a large high-resolution wall display aggregates information from traffic cameras, incident cameras, news feeds from local and national sources, along with live Twitter feed. We also discuss the challenges presented when implementing gestures and multi-surface interactions for EOC personnel.

Workshop on Gesture-based Interaction Design: Communication and Cognition 2014

RELATED WORK

Emergency response planning is comprised by many important tasks, from detecting and monitoring the emergency to the deployment of resources and communication management [1][6]. Even though this domain has been explored from the perspective of several different technologies, common rules on interactions to improve collaboration are scarce as the UI is heavily impacted by the domain and system's purpose, as highlighted by Bortolaso et al. Just co-locating people around a device does not mean that the collaboration will be improved since the tradeoff between simplicity and functionality must be evaluated multiple times during system's development.

uEmergency is a forest fire simulation system running on a very large-scale interactive tabletop [2]. This tabletop's dimensions (381x203cm) allow several users to collaborate using the system concurrently while considering personal space (local and private workspace) and a global space (shared among all users and synchronized through a button). Users can interact with the system using a digital pen or touch gestures. It's possible to translate and resize the map using gestures with one and two fingers, respectively; to perform annotations dragging and dropping markers from a menu into the map; and changing the simulation's time point through a slider available on each personal workspace. Since all users are sharing a very large-scale tabletop, collaboration is improved through visual cues from each user's actions.

Besides digital pen and touch gestures, physical tokens are also used in planning disaster systems on tabletops [3]. They act as input, changing simulation parameters according to their physical position above the tabletop, and provide feedback through images projected on them. The manipulation of physical tokens to interact with emergency systems has reduced the learning curve of these systems.

The research space of multi-surface environments is very well explored and significant research has been done in exploring the different ways in which interactions can take place [5][7][8].

The collaborative nature of the activities related to emergency response planning and the presence of multiple and heterogeneous devices in a room provide an opportunity for study and experimentation of different types of gesture-based interactions in the emergency response domain, described in this work by ePlan Multi-Surface.

EMERGENCY RESPONSE PLANNING

Working in collaboration with an emergency response simulation software company, C4i Consultants Inc.[1] (C4i), located in Calgary, Alberta, Canada, we designed the ePlan Multisurface Emergency Response Environment. It's a proof-of-concept environment designed based on our partner's specification for training purposes and, as shown in Figure 1, it consists of a large high-resolution wall display, digital tabletop and multiple iPads. The system was built using C4i's desktop software ePlan to drive the emergency simulation, as well as MSE-API [4], which provides communication between devices and multi-surface interactions. After discussions with C4i, researching multi-surface environments and based on their experience on emergency response, we decided to select a subset from the gesture set available in prior work [5] and implemented in MSE-API. To highlight the role of gestures and interactions in the prototype, we will describe the typical usage scenario in the context of an emergency response-planning scenario used by our industry partner:

Step 1: Emergency Alert Issued.

In the first stage, emergency response planners, in particular, the emergency response operation controller receives different information (text, email, and phone) from various sources (fire, emergency medical services, and police) about an emergency. The EOC then determines the type of emergency that is occurring and issues a state of emergency to a city or municipality, if necessary. During this time, the EOC is continually analyzing and receiving information on both their personal iPads, as well analyzing the situation on the large wall display that highlights information such as live camera traffic cameras and Twitter feeds. The EOC is often interrupted or simultaneously performing tasks due to the evolving nature of an emergency.

Step 2: Response Representatives Assemble.

After the alert has been issued, the response representatives assemble in the ePlan multi-surface emergency response environment. These representatives include the fire department, emergency medical services (EMS), law enforcement agencies, hazard materials unit (depending on the severity of the situation), among others. Each representative maintains their own personal iPad containing relevant information that is either shared at their discretion or used in their own assessment for allocating resources for the emergency. To share their information, a representative is able to use multi-surface interactions such as flick to send to the wall display, allowing all representatives to see updated information, the tabletop to assist in collaborative emergency response planning with other representatives (pouring from the iPad to the tabletop also performs the same function) or to other iPads, to facilitate communication between different representatives. These interactions are typically done in parallel with planning or analysis tasks in the environment, and notifications are used that visually prompt representatives of new data. The multi-surface interactions also serve as a visual prompt for representatives of new or updated emergency information.

[1] http://www.c4ic.com/

Figure 2. User performing "flick" gesture on iPad's screen to send content from tablet to tablet (left); tablet to tabletop (center); and tablet to wall display (right).

Step 3: Emergency Response Planning.

During the emergency response planning session(s), which last until the end of an emergency situation, numerous types of interactions occur. This session is the most critical component of emergency response planning, as significant coordination and planning is done. In ePlan multi-surface, emergency response, personnel are continually collaborating and consuming new information rapidly using iPads, while also simultaneously trying to keep track and manage the emergency through the wall display and digital tabletop. At the end of an emergency response situation, a report is typically generated that summarizes the emergency and the contributions of the emergency response personnel.

INTERACTIONS AND GESTURES IN EPLAN PROTOTYPE

Touch-based gestures on ePlan Multi-Surface are used for content transfer to support EOC personnel during emergency planning scenarios. These gestures are based on a device's type (iPad, digital tabletop, and wall display), its physical position and relative position to other devices, different gestures and interactions might be performed. These include the following

1. Pulling content from another device

Considering the case an EMS specialist wants to analyze part of the global situation, presented on wall display, in his iPad. He orients his iPad towards the wall display and perform a pull gesture using one finger from the top to the center of the screen. The content will then be copied from the wall display into his mobile device.

2. Pouring content from tablet into tabletop

During emergency planning activities, users collaborate among each other by sharing analysis (annotations) made privately in their devices. On ePlan Multi-Surface it's possible to make content made on one's device globally available: a user share his annotations by placing his iPad above the tabletop and "pouring" its content into tabletop's screen, as shown in Figure 2. The content will then be displayed on wall display and available for everyone's analysis.

3. Sending content through flick gestures

A common situation in emergency planning is one specialist informing others or discussing about a given aspect of the emergency – for example, firefighters informing an evacuation plan to the police. Since ePlan users can interact with their own devices in a local fashion, it's possible to share information with specific users through flick gestures on tablet's screen. It's possible to perform tablet-tablet, tablet-tabletop and tablet-wall display content transfer through such gestures, as shown in Figure 3. MSEAPI identifies who should receive the content according to sender's position and orientation.

Figure 3. User "pouring" content over tabletop

GESTURE DESIGN AND FEASIBILITY

Building upon prior work of multi-surface interactions and gestures [5], the goal of building ePlan multi-surface was to examine multi-surface gestures in the context of emergency response planning. In the oil and gas domain we observed that multi-surface gestures and interactions still require a significant amount of work to be adopted, and adoption barriers, such as learnability, still need to be overcome.

In our preliminary work presented here, design sessions, feedback and informal discussion has suggested similar findings to [1]. We believe this reflects not just the nature of how we've considered multi-surface interactions and gestures, but the literature for the space as well. One important aspect of gesture design, especially in the context of multi-surface environments, is that the focus has primarily been on using gestures to transfer content. There has been a noticeable shift away from the GUI based techniques as the technology of devices has evolved. With more information being available from devices (e.g. gyroscope, accelerometer) to essentially mimic or create "natural" interactions (e.g. pick and drop, pouring, etc), this seems like a logical research track to follow. However, in the case of emergency response planning, this shift seems to be at the expense of the user, as appears to be both more efficient and feasible for emergency response personnel to not use such gestures in the traditional sense of replacement, but merely augmenting their tasks.

Another interesting implication of gesture design and interactions from a multi-surface perspective, is their grounding. For the most part, we have introduced gestures that were designed or inspired by prior work, and have seen that in practical, real-world settings, the results are less than ideal. This seems to suggest, that gesture and interaction design should first begin in grounded domains before being examined and tested more generically, as is much of the case in the gesture design literature.

For us, in the context of multi-surface environments and emergency response planning, this opens a number of interesting questions, that we'd like to explore in future work, which are the following:

- What are alternative ways to designing gestures in a multi-surface context, especially when they are so new?
- How can we leverage or augment traditional techniques (e.g. 2D interfaces) in the design of multi-surface gestures and interactions?
- Are multi-surface interactions and gestures truly feasible or faster for domains such as emergency response planning?

Also, as future work, we intend to evaluate the environment prototype in a training situation with our partner, since to run a test during a real emergency would carry great risk for the ones involved.

REFERENCES

1. Bortolaso, C., Oskamp, Matthew, Graham, N., and Brown, D. OrMiS: a tabletop interface for simulation-based training. In *Proc*. *ITS 2013*, ACM Press (2013), 145-154.
2. Yongqiang Qin, Jie Liu, Chenjun Wu, and Yuanchun Shi. 2012. uEmergency: a collaborative system for emergency management on very large tabletop. In Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces (ITS '12). ACM, New York, NY, USA, 399-402.
3. Kazue Kobayashi, Atsunobu Narita, Mitsunori Hirano, Ichiro Kase, Shinetsu Tsuchida, Takaharu Omi, Tatsuhito Kakizaki, and Takuma Hosokawa. 2006. Collaborative simulation interface for planning disaster measures. In CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06). ACM, New York, NY, USA, 977-982.
4. MSE-API. https://github.com/ase-lab/MSEAPI
5. Seyed, T., Burns, C., Costa Sousa, M., Maurer, F., and Tang, A. Eliciting usable gestures for multi-display environments. In *Proc*. *ITS 2012*, ACM Press (2012), 41-50.
6. Hofstra, H., Scholten, H., Zlatanova, S., and Alessandra, S. Multi-user tangible interfaces for effective decision-making in disaster management. *Remote Sensing and GIS Technologies for Monitoring and Prediction of Disasters*, Springer (2008), 243-266.
7. Dachselt, R., and Buchholz, R. Natural throw and tilt interaction between mobile phones and distant displays. *CHI '09 Extended Abstracts on Human Factors in Computing Systems (CHI'99),* Boston, MA, USA (2009), pp. 3253-3258.
8. Rekimoto, J. Pick-and-drop: a direct manipulation technique for multiple computer environments. *Proceedings of the 10th annual ACM symposium on User Interface Software and Technology (UIST'97),* Banff, Alberta, Canada (1997), pp. 31-39.

Embodying Diagramming through Pen + Touch Gestures

Andrew M. Webb
Interface Ecology Lab
Dept. of Computer Science and Engineering
Texas A&M University
andrew@ecologylab.net

Andruid Kerne
Interface Ecology Lab
Dept. of Computer Science and Engineering
Texas A&M University
andruid@ecologylab.net

ABSTRACT

Ideation, the process of generating new ideas, is central to design tasks in which the goal is to find novel solutions around a set of requirements. Designers create diagrams as external representations of ideas. Ideas and the design process from which they emerge are embodied by both the gestures used to create diagrams and the diagrams themselves. Interaction design needs to leverage embodiment to support creative cognitive processes. We hypothesize that expressive embodied gestures for transforming diagrams will stimulate design ideation. We introduce new bimanual gestures for creating diagrams. We posit implications for the design of gestural interaction to support design ideation.

INTRODUCTION

Ideation, the font of innovation, means the process of generating new ideas. Ideation is central to design. *Design* is a purposeful and creative process in which means to an end is laid down, e.g. a solution is derived. Design processes are supported by embodied representations, including gestures, tangibles, and diagrams, which have been found to help people think [15, 12, 24, 16]. An embodied representation externalizes mental models and ideas to supplement cognition through movement of the body. For example, one can externalize models of the physical world to help in navigating a set of directions through hand gestures, rotating the hand and pointing fingers left and right, up and down [12]. Leveraging embodiment in interaction design is key to mitigating the cognitive and neuromuscular load inherent in design environments with large visual spaces and complex command sets.

Building upon Dourish's definition of embodied interaction [5], we define *embodied gestures* as not simply gestures that involve the body (as this includes all gestures), but that the "approach to design and analysis of [gestures] takes embodiment to be central to, even constitutive of, the whole phenomenon." Embodied gestures take into account the physical body, operations performed, and associated cognitive processes. We do not consider a gesture embodied if it is designed strictly for efficiency without regard for kinaesthetic qualities of movement and how the qualities relate to operations performed and help in cognitive processes. The proliferation of sensory interaction modalities, such as multi-touch surface, pen, and computer vision, enables cost-effective development of new forms of embodied interaction.

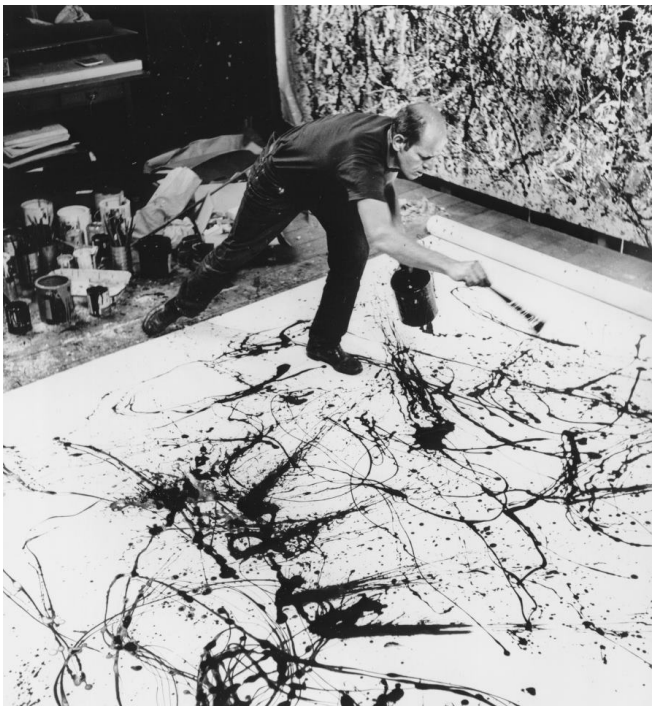


Figure 1: Photograph by Hans Namuth of Jackson Pollock engaging in gestural painting [23].

A *diagram* is a design thinking tool that enables and stimulates imagination, facilitating conceptualization. Diagrams mediate exploration of relationships between concepts, using ambiguous visual representations to foster varied, flexible interpretations. Architects see a diagram as an "engine of novelty" [17], documenting aspects of integral design thinking processes [21]. The philosopher and cultural theorist, Deleuze, identifies the diagram as an abstract machine, "defined by its informal functions and matter and in terms of form makes no distinction between content and expression, a discursive formation and a non-discursive formation" [4].

We hypothesize that expressive embodied gestures for creating diagrams will stimulate design ideation. Gestural interaction (e.g. Figure 1) supports exploratory ideation processes, such as sketching ideas and using one's hands (sometimes

Submitted for review.

with tools) to transform representations in different ways. We are developing a new embodied diagramming environment to investigate impact on creative cognitive processes. We derive new bimanual pen + touch gestures (combining pen and multi-touch input) for diagramming. We are evaluating our new environment and gestures in architecture education.

This paper begins with a discussion of related work. Next, we introduce a new diagramming medium and propose novel pen + touch gestures stimulating design ideation. We describe a context for evaluation. We conclude with this research's potential implications for designing gestural interaction.

RELATED WORK

We ground design of our new medium and pen + touch gestures in embodied cognition. We build upon prior sketch design environments and bimanual interaction techniques.

Embodied Interaction to Promote Cognition

Cognitive models, including those associated with creativity, are embodied [6]. Embodied interactions promote cognition through physical movement. For example, when readers manipulate objects that correspond to characters and actions in a text, it greatly enhances comprehension and memory, as measured by both recall and inference tests [7]. Tversky et al. demonstrate that gestures help people not just to communicate meaning, but further, to remember and to understand complex ideas [27]. Of particular value are iconic gestures, whose shape maps directly to what they mean, metaphoric gestures, that use spatial representation to convey relationships (such as distance), and embodied gestures, which encode knowledge motorically, as images and diagrams encode pictorially. These gestures augment memory, and the representation of meaning. Together, these species of gestures provide us with cognitive guidance for how to develop gestural interfaces that will help designers manipulate diagrams and create ideas as they interact.

Sketch Design Environments

Sketches act as interactive imagery [8] in which strokes are drawn, re-drawn, drawn over, and erased, transforming ideas. Ideas are externalized, manipulated, reflected upon, and reinterpreted [25]. Designers quickly sketch out ideas while simultaneously manipulating diagrams in various ways, supporting creative discovery as externalized ideas are combined and restructured [28]. Sketches can be rigorous in visually describing details, but also ambiguous using abstract forms and implicit visual features.

SILK is a user interface (UI) design tool that enables rapid prototyping interface designs through sketching [19]. The primary goal of SILK is not to promote creative UI design, but to facilitate quickly designing UIs and testing basic interactions. The explicit shapes and symbols needed to form recognizable sketches of UI components oppose our goal to facilitate ambiguity to promote design ideation. Electronic Cocktail Napkin is a pen-based collaborative design environment that supports abstraction, ambiguity, and imprecision in sketching [9]. Instead of focusing on sketch recognition, the present research addresses how gestural interaction can support creative processes.

Bimanual Interaction

We seek to embody diagram interactions through bimanual pen + touch gestures. Guiard developed one of the first models of bimanual interaction, the *kinematic chain* [10]. In a kinematic chain, the non-dominant hand (NDH) acts to define a reference frame for the actions of the dominant hand (DH). When drawing on paper, this is equivalent to the NDH positioning and rotating the paper in conjunction with the DH making marks with a pencil. We will design bimanual interaction techniques using the kinematic chain model.

The benefits of bimanual interaction are well documented. Toolglass widgets are translucent interface tools that are positioned with the NDH, and interacted with using the DH [2]. A kinematic chain is formed, where the NDH hand defines which elements are affected by the widget through positioning, and the DH selects which operations to perform and provides fine grain control over how operation parameters are manipulated. Hinckley et al. recommend that the pen by itself always makes marks, but when combined with touches creates new forms of interaction [11]. Brandl et al. investigated benefits of bimanual interactions across different input modalities [3]. They compared time and errors to complete path following tasks using touch with both hands, pen with both hands, and pen with DH and touch with NDH. Findings indicated that pen and touch was quicker, more accurate, and more preferred.

EMBODIED DIAGRAMMING ENVIRONMENT

We are developing a new diagramming environment to investigate impact of embodied gestures on creative design processes. We introduce a new diagramming medium and propose pen + touch gestures for transforming diagrams. We are deploying our environment in architecture education to investigate impact on design ideation. We are engaging in iterative design using formative evaluations in architecture education to improve our diagramming environment and gestures.

Information Composition + Sketching

We introduce a new diagramming medium that integrates information composition with sketching in an infinite zoomable space. *Information composition*, a diagramming medium for representing a personal information collection as a connected whole, supports reflection when performing information-based ideation tasks [29]. Designers engage in *information-based ideation* tasks, using information as support for generating new ideas [13], such as investigating how properties of different materials will impact a design. Composition authoring is a process of gathering clippings from information resources. These clippings function indexically, enabling access back to the information resources. As a diagram, an information composition expresses relationships between gathered ideas through implicit visual features, such as spatial positioning, size, color, and translucence. As holistic sensory media, compositions are designed to engage thinking about, authoring, annotating, and reflecting on collections as records, sources, and media of ideation.

In our new diagramming medium, designers gather clippings while sketching out ideas and relationships amidst the col-

Figure 2: Early stage information composition + sketching diagram created by an architecture student in our field study after the first session. Students were asked to create a diagram on multiple scales investigating how contrast and juxtaposition of scales express relationships. Each visual clipping serves as an index for accessing information resources on the architect, Alberto Campo Baeza, and his work. The student has begun sketching over image clippings to emphasize important ideas and explore relationships in geometry.

lected information (see Figure 2). Complex interaction issues arise as designers intermix actions involving sketching, gathering clippings, and visual transformations.

Pen + Touch Gestures

We seek to reduce cognitive and neuromuscular load associated with tools supporting complex interaction by providing gestural interactions that are natural and intuitive and more directly connect designers to diagrams. The goal is not simply to make diagramming easier, but to aid designers in forming abstractions and investigating ambiguous representations through embodied experiences.

While diagrams vividly convey spatial relationships and ideas, much of the thinking and mental models of the author are encoded in embodied creative processes of transformation. We define *diagram transformation* as any operation that changes a diagram to encode meaning, e.g.: adding and removing elements, affine transforms, color-space transforms, cropping, and distortion. Just as a designer uses her hands to transform physical diagrams (e.g. rotation, transparent overlays, folding or bending material), embodied gestures are needed for transforming our new diagramming medium.

The ability to quickly sketch an idea is important as suggested by others [1, 9, 26, 22]. As recommended by Hinckley et al [11], we propose that the pen, when used by itself, always makes marks. The exception is when pen input is combined with touches, that act as modifiers, invoking commands. This enables designers to fluidly switch between sketching ideas and manipulating or transforming diagram elements.

Designers form kinematic chains when interacting with physical media, such as orienting paper while sketching or rotating a model to find an advantageous angle for adding or removing parts. We propose using gestures with kinematic chains where the NDH gestures to select the transformation and element(s) affected and the DH (with a pen) precisely performs the transformation. Transforming a diagram may be exploratory and require reverting changes to investigate alternate ideas. In these kinematic chain transformation gestures, the NDH can also perform a gesture to undo or redo a series of transformations.

In an infinite, zoomable information space, designers will have difficulty keeping track of ideas and relationships between ideas due to limits of working memory [20]. We hypothesize that embodied gestures for manipulating scale, rotation, position, and zoom will support spatial cognition helping designers remember where ideas are located, as well as, think about relationships between ideas represented visually through spatial distance. A single touch gesture positions diagrammatic elements. A two point pinch gesture, involving either two fingers or a one finger and the pen, performs simultaneous scale, rotate, and translate.

Embodied gestures should not be limited to finger touches and pen points. Designers use their whole hands, arms, and body when creating diagrams. We propose whole hand ges-

tures to transform multiple elements or regions of a diagram. Whole hand gestures are coarser than finger touches, allowing for broad transformations, such as aligning elements along a line represented by the side of one hand with fingers extended straight or sweeping elements into groups or piles.

Evaluation: Architecture Education and Experiments

As a context for evaluating effects of embodied gestures on design ideation, we are engaging in an ethnographic investigation and field study in architecture education. Students in the graduate course, Visual Thinking: Theories and Methods of Diagramming, used a preliminary version of our diagramming environment on a course assignment. We collected the diagrams they created, and video recorded their gestural interactions while working in the environment. We are in the process of analyzing this data. We will derive codes for recorded gestural interaction, looking specifically at sequences of gestures performed, numbers of hands used, and qualities of movement (e.g. speed, direction, effort [18]). Additionally, we are recruiting architecture Ph.D. students to use the diagramming environment for a longer period as part of developing their dissertations. We seek to gain better understandings of architecture design processes and how embodied diagramming environments can support these processes.

We will conduct controlled laboratory experiments to evaluate impact on ideation of embodied gestures. Participants will create diagrams to answer two open-ended design problems. The independent variables will be (1) whether or not the environment used supports embodied gestures with pen and touch or non-embodied gestures with mouse and keyboard; and (2) the design tasks being performed. The dependent variables are the diagrams that participants create and a set of measures for analyzing those diagrams. We will use information-based ideation metrics [14] to measure the novelty (uniqueness of ideas), variety (diversity of ideas), fluency (number of ideas), and quality of ideas within the diagrams. We hypothesize that diagramming with embodied gestures will help participants externalize, think about, and reflect upon ideas while working within large information collections, leading to more novel and varied designs. We will validate our hypothesis through comparison of information-based ideation measures between independent variables.

CONCLUSION

HCI researchers have the opportunity to transform design processes with the development of new embodied pen + touch gestural interaction. Emerging sensing technologies make this transformative research possible. We need gestures that help offload cognitive processes to external forms addressing limits of human attention. Digital environments enable creation and exploration of large information spaces that can be dynamically transformed. Gestural interaction needs to be expressive to support diverse transformations. Developing gestural interaction techniques that use kinematic chains will help designers think about how elements are transformed and the relationships between elements and transformations.

REFERENCES

1. Arnheim, R. *Visual Thinking*. University of California Press, 1969.
2. Bier, E. A., Stone, M. C., Pier, K., Buxton, W., and DeRose, T. D. Toolglass and magic lenses: the see-through interface. In *Proc. SIGGRAPH* (1993).
3. Brandl, P., Forlines, C., Wigdor, D., Haller, M., and Shen, C. Combining and measuring the benefits of bimanual pen and direct-touch interaction on horizontal interfaces. In *Proc. AVI* (2008), 154–161.
4. Deleuze, G. *Foucault*. Univ. of Minnesota, 1988.
5. Dourish, P. *Where the Action is: The Foundations of Embodied Interaction*. Bradford Books, 2004.
6. Glenberg, A. Why Mental Models Must Be Embodied. *Advances in Psychology 128* (1999), 77–90.
7. Glenberg, A. M., Brown, M., and Levin, J. R. Enhancing comprehension in small reading groups using a manipulation strategy. *Contemporary Educational Psychology 32*, 3 (2007), 389 – 399.
8. Goldschmidt, G. The Dialectics of Sketching. *Creativity Research Journal 4*, 2 (1991), 123–143.
9. Gross, M. D., and Do, E. Y.-L. Ambiguous intentions: a paper-like interface for creative design. In *Proc. UIST* (1996), 183–192.
10. Guiard, Y. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of motor behavior 19* (1987).
11. Hinckley, K., Yatani, K., Pahud, M., Coddington, N., Rodenhouse, J., Wilson, A., Benko, H., and Buxton, B. Pen + touch = new tools. In *Proc. UIST* (2010).
12. Jamalian, A., Giardino, V., and Tversky, B. Gestures for thinking. In *Proc. of the Cognitive Science Society Meetings* (2013).
13. Kerne, A., Webb, A. M., Smith, S. M., Linder, R., Moeller, J., Lupfer, N., Qu, Y., and Damaraju, S. Evaluating information-based ideation with creativity measures of curation products. *ACM Transactions on CHI in minor revisions* (2013).
14. Kerne, A., Webb, A. M., Smith, S. M., Linder, R., Moeller, J., Lupfer, N., Qu, Y., and Damaraju, S. Evaluating information-based ideation with creativity measures of curation products. *will appear in Transactions on CHI* (2014).
15. Kim, M., and Maher, M. Comparison of designers using a tangible user interface & graphical user interface and impact on spatial cognition. In *Proc. Human Behaviour in Design* (2005).
16. Kirsh, D. Thinking with external representations. *AI & Society 25*, 4 (2010), 441–454.
17. Kwinter, S. The genealogy of models: The hammer and the song. *Diagram Works, ANY 23* (1998), 57–62.
18. Laban, R., and Lawrence, F. *Effort*. MacDonald and Evans, 1947.

19. Landay, J. A., and Myers, B. A. Interactive sketching for the early stages of user interface design. In *Proc. CHI* (1995), 43–50.

20. Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review 63*, 2 (March 1956), 81–97.

21. Peponis, J., Lycourioti, I., and Mari, I. Spatial models, design reasons and the construction of spatial meaning. *Philosophica – Diagrams and the anthropology of space 70* (2002), 59–90.

22. Plimmer, B., and Apperley, M. Computer-aided sketching to capture preliminary design. *Aust. Comput. Sci. Commun. 24*, 4 (Jan. 2002), 9–12.

23. Rose, B., Ed. *Pollock: Painting*. Agrinde Publications Ltd., 1980.

24. Suwa, M., and Tversky, B. What do architects and students perceive in their design sketches? a protocol analysis. *Design Studies 18*, 4 (1997), 385 – 403.

25. Suwa, M., Tversky, B., Gero, J., and Purcell, T. Seeing into sketches: Regrouping parts encourages new interpretations. In *Visual and spatial reasoning in design* (2001), 207–219.

26. Trinder, M. The computer's role in sketch design: A transparent sketching medium. In *Computers in Building*. Springer, 1999, 227–244.

27. Tversky, B., Heiser, J., Lee, P. U., and Daniel, M. P. *Explanations in gesture, diagram, and word*. Oxford University Press, Oxford, 2009, 119–131.

28. Verstijnen, I., van Leeuwen, C., Goldschmidt, G., Hamel, R., and Hennessey, J. Sketching and creative discovery. *Design Studies 19*, 4 (1998), 519 – 546.

29. Webb, A. M., Linder, R., Kerne, A., Lupfer, N., Qu, Y., Poffenberger, B., and Revia, C. Promoting reflection and interpretation in education: Curating rich bookmarks as information composition. In *Proc. Creativity and cognition* (2013).

Tangible Meets Gestural: Gesture Based Interaction with Active Tokens

Ali Mazalek[1], Orit Shaer[2], Brygg Ullmer[3,4], Miriam K. Konkel[5]

Synaesthetic Media Lab[1]
Digital Media & GVU Center
Georgia Institute of Technology
Atlanta, GA, USA

Dept. of Computer Science[2]
Wellesley College
Wellesley, MA, USA

School of EECS[3], Center for Computation and Technology[4], and Dept. of Biological Sciences[5]
Louisiana State University
Baton Rouge, LA, USA

mazalek@gatech.edu, oshaer@wellesley.edu, {ullmer, konkel}@lsu.edu

ABSTRACT
Emerging multi-touch and tangible interaction techniques have a potential for enhancing learning and discovery but have limitations when manipulating large data sets. Our goal is to define novel interaction techniques for multi-touch and tangible interfaces, which support the exploration of and learning from large data sets. In this paper we discuss conceptual, cognitive, and technical dimensions of gestural interaction with active tangible tokens for manipulating large data sets.

Author Keywords
Tangible interaction; active tokens; gesture-based interaction; big data.

ACM Classification Keywords
H.5.2 [Information Interfaces and Presentation]: User Interfaces---*input devices and strategies, interaction styles*.

INTRODUCTION
To date, little research has been devoted to investigating tangible and multi-touch interaction in data-intensive domains such as genomics, environmental studies, and social networks. Here, learning and discovery rely on manipulating large data sets using sophisticated computational methods [25]. Tangible and tabletop interactions provide form factors that foster collaboration through visibility of actions, multiple access points, and egalitarian input [10, 17, 26], and support distributed cognition [22, 26]. In the context of data exploration, the ability to support collaborative work and enhance reasoning could lead to new discoveries.

Designing tangible and multi-touch systems that support learning and discovery in *data-intensive* areas requires going beyond the application of existing interaction techniques. While direct touch is currently a standard input method for interactive surfaces, in data-intense applications visual representations are typically small, finger size and occlusion may interfere with direct interaction of small targets through touch [9, 32]. Similarly, WIMP-style control elements provided by various multi-touch toolkits, such as scrollbars, sliders, check boxes, and text fields, may often pose the same challenges for effective and accurate touch interaction, or take expensive screen real-estate [9]. For some data manipulation tasks that require high precision, touch-based graphical representations such as knobs and sliders are less effective than their physical counterparts [8].

Several researchers have considered novel multi-touch gesture-based interaction techniques for data driven applications; e.g. [9, 12, 32]. However, while providing advantage over touch interaction with WIMP style controls, multi-touch gestures often suffer from low discoverability and lack of persistence [9]. Considering these limitations of multi-touch interaction for Big Data exploration, we suggest that tangible systems with clear feedback and strong constraints provide an alternative approach for exploring Big Data.

Such systems can utilize both soft (graphical) and hard (physical) tokens and constraints to guide users in querying and interpreting large data sets, enabling users to collaboratively engage in problem solving. Technological advances and mass-market offerings such as Sifteo Cubes [1] also open possibilities for the use of *active tokens* [36].

Active tokens are programmable physical objects with integrated display, sensing, or actuation technologies (e.g., [1, 16, 28, 35, 36]). Thus, they can be reconfigured over time, allowing users to dynamically modify their associations with datasets or controls. The use of active

tokens expands the design space of token and constraints interaction [24, 27]. Combining interactive multi-touch surfaces with active tokens could facilitate the presentation and manipulation of Big Data while preserving benefits of tangible interaction such as support for two-handed interaction, co-located collaboration, and strong affordances. We focus on a sub-class of active tokens that can be manipulated using gestures independently from global constraints. Gestural interaction with active tokens blurs the boundaries between tangible and gestural interaction, and fits within the area defined as "tangible gesture interaction" [31].

In this paper, we briefly consider the conceptual dimensions for gestural interaction with tangible active tokens, discuss cognitive foundations for gesture-based interaction with active tokens for discovery and learning, and examine recent technical configurations that are relevant to this area.

CONCEPTUAL DIMENSIONS

Tangible Tokens and Constraints (TAC) systems [24, 27] engage the physical expression of digital syntax through configurations of tokens and constraints. For example, token and constraint relations such as presence, position, sequence, proximity, connection, and adjacency are utilized to encode information as well as to communicate to users what kinds of interactions an interface can (and cannot) support. The manipulation of a token in respect to its constraints results in modifying both physical and digital states of the system. Gestural interaction with active tokens expands the design space of TAC interaction, blurring boundaries between tangible and gestural interaction.

Several prior systems have explored gesture-based interaction with active tokens. For example, the Tangible Video Editor [36] employed active tokens to represent video clips. SynFlo [35] utilized active tokens to simulate a biology experiment and evoked gestures such as pouring and shaking. However, aside from the parameter bars of Tangible Query System [28], the Big Data context has yet to be engaged.

In [30] we investigate user-generated gestures for exploring large data sets. Our findings highlight three characteristics of gestural interaction with tokens: space, flow, and cardinality: *Space* describes where an interaction takes place: typically on-surface (integral), on-bezel (proximal), and in-air (distal). The dimension of *flow* is adopted from [34] and may be regarded as having both discrete and continuous dimensions. *Cardinality* indicates the number of hands and tokens involved in a gesture, with atomic, compound, and parallel subelements. These characteristics are elaborated in [30]. However, gesture sets are yet to be evaluated within task and data-driven scenario.

COGNITIVE FOUNDATIONS

The centralist (brain-centric) view of cognition has in recent decades been shifting to what Killeen and Glenberg [13] call an "Exocentric Paradigm." This posits that cognition is a process that involves the brain, the body, and the environment. This paradigm is supported by a wide array of empirical evidence, which falls broadly under terms like "embodied cognition," "situated cognition," and "distributed cognition" [11, 15, 33]. From the perspective of our active token and Big Data discussion, we are especially interested in how evolving notions of cognition can further our understanding of how people's physical actions and interactions with their environment support scientific reasoning; and how this understanding can inform the design of physical and computational tools for discovery and learning.

External representations and scientific reasoning

From early childhood, our interaction with physical objects appears to be closely connected with our learning and thinking processes. For example, researchers have shown that touching physical objects can help young children learn how to count by helping them keep track of their activities, and by allowing them to connect each physical object with a number [2]. Studies with children have also shown a co-development of language and gesture [6] and the origins of gestures appear to be connected to physical actions.

In thinking about complex problems, scientists employ external artifacts (e.g., models, diagrams, instruments) to support their reasoning [20, 21, 23]. A prominent example is the double helix model of DNA built by Watson and Crick, which enabled the two scientists to quickly form and test out hypotheses by manipulating the model's physical structure. Physical models can thus provide an entry point for the cognitive apparatus in the form of both conceptual and material manipulation [5].

Computational systems can also embody knowledge. Typically, visualizations are used to make computational models accessible to human cognitive capabilities. Some visualizations can be interactively explored and filtered in order to find patterns that might enhance understanding. However, the interaction with most visualizations is not closely connected to the underlying model of the studied phenomenon or system. That is, the interactions users have with most interactive visualizations (e.g., using button clicks, menu selections, etc.) are very unlike Watson and Crick's manipulation of the physical DNA model. In the latter case, the actions made with the physical model were tightly coupled with the scientists' emerging conceptual model, which helped to leverage the connection in the brain between motor, perceptual, cognitive processes in the development of insights [7]. We believe that systems that employ active tokens have the potential to leverage gestural interaction/manipulation in order to create a similar connection between the computational model/data and the user's conceptual model in areas of scientific problem solving.

Tokens and gestures for thinking and learning

Martin and Schwartz [18] have investigated how physical actions impact thinking and learning. They provide four ways in which this happens: induction, off-loading, repurposing, and physically distributed learning. Although our focus is not limited to children, we use these categories as a framework for considering how gestural interactions with active tokens might support thinking and learning.

Induction is when people do not have stable ideas, but they are acting in a stable environment that offers clear feedback and strong constraints that can guide interpretation [18]. In this case, physical actions can enable them to query the environment and test their hypotheses. From an interaction perspective, well-designed feedback and constraints could thus allow TAC systems to support testing of hypotheses and problem solving. For example, graphical (soft) or physical (hard) constraints and the shape of tangibles can suggest ways in which tangibles can be placed on an interactive surface or combined together.

Off-loading is when both people's ideas and the environment in which they operate are stable [18]. In this case, people rely on the environment to reduce cognitive load of a task -- often called distributed cognition [11]. From an interaction perspective, physical tokens can support distributed cognition as users spread and group them in different ways [3, 4, 22]. Although it is also possible to spread and group digital artifacts, e.g. via multi-touch interaction, Antle and Wang's comparison of TUI and multi-touch interaction in a puzzle-solving task [4] revealed that the TUI condition supported more efficient and effective motor-cognitive strategies.

Repurposing is when people have stable ideas about the given problem but their environment is adaptable and can be changed to achieve their goals [18]. This relates to Kirsch and Maglio's distinction between pragmatic and epistemic actions [15]. Pragmatic actions bring people closer to their goal; epistemic actions mostly support people's ability to think about the problem. Although tokens have physical form factors and constraints that suggest ways to manipulate them, the characteristics of gestural interaction with tokens described above (space, flow, cardinality) point to ways in which TAC systems might leave room for individual customization. For example, tokens placed on-surface may have certain defined behaviors, while on-bezel or in-air interaction with the same tokens might allow users to redefine their functions, allowing each person to develop their own strategies for problem-solving.

Physically distributed learning is when people's ideas and the environment are both adaptable [18]. Here, people may interact with their environment without knowing exactly what steps they need to take or even the final state. By studying how children learn fractions with different materials, Martin and Schwartz [18] found that the emergence of new interpretations through physical adaptations of the environment is a benefit of physical action for learning abstract ideas. This suggests that system designers need not always provide tightly structured environments, but should allow people to create their own structures for problem solving. The combination of gestural interaction with active tokens can provide ways to make TAC interaction more adaptable and open-ended.

TECHNICAL CONFIGURATIONS

Here we provide a brief overview of some recent technical advancements relevant to TAC systems.

Tables, tablets, and smartphones

Interactive tables and tablets have been available in varying forms for several decades. While interactive tables have not reached the mass market, the commercial release first of Microsoft PixelSense [19], and more recently of lower-cost capacitive tables, are laying the hardware foundations for broader dissemination. Even more impactful is the pervasive consumer adoption of smartphone and tablet technologies. Many of these devices are sensor-rich and some including RFID/NFC technologies. Tablets and smartphones provide near-ready platforms for the mediation of diverse 1D and 2D constraints. Mass commercialization of inch-scale devices such as Sifteos [1] offer compelling platforms for active tokens.

Embedded computing

The last decade has witnessed explosive growth and adoption of the Arduino and Raspberry Pi processors, which offer a compelling mix of mass-market economics, mass community investment, and high-level software environments. Viewed from a TAC perspective, in synergy with mass-market tablets, such embedded tools can complement sensing and mediation capacities such as sensing on the central active surface, and on the bezels. Bezel integrations can both extend the interaction real estate of individual devices [29]; and help stitch together tiled arrays of devices.

CONCLUSION

In this paper we considered conceptual, cognitive, and technical dimensions for gestural interaction with tangible active tokens. Gestural interaction with active tokens expands the design space of tangible Token and Constraints system and offers new possibilities for learning from and understanding of large data sets.

REFERENCES

1. (n.d.). Retrieved from Sifteo: https://www.sifteo.com/
2. Alibali, M.W., & DiRusso, A.A. (1999). The function of gesture in learning to count: more than keeping track. Cognitive Development 14(1), 37–56.
3. Antle, A. N., & Wise, A. F. (2013). Getting down to details: Using theories of cognition and learning to inform tangible user interface design. Interacting with Computers, 25(1).

4. Antle, A.N., Wang, S. (2013).Comparing Motor-Cognitive Strategies for Spatial Problem-Solving with Tangible and Multi-touch Interfaces, In Proc. of TEI '13, ACM.
5. Baird, D. (2004). Thing Knowledge: A Philosophy of Scientific Instruments. Berkeley: University of California.
6. Bates, E., & Dick, F.(2002). Language, gesture, and the developing brain. Developmental Psychobiology 40, 293–310.
7. Chandrasekharan, S. (2009). Building to discover: A common coding model, Cognitive Science, 33: 1059-1086.
8. Crider, M. et. al. (2007) A mixing board interface for graphics and visualization applications. In Proceedings of Graphics Interface 2007, pages 87–94. ACM.
9. Drucker, S., Fisher, D., Sadana, R., Herron, J., & Schraefel, M. C. (2013). TouchVix: A Case Study Comparing Two Interfaces for Data Analytics on Tablets. In Proc. of CHI, ACM.
10. Hornecker, E., Marshall, P., Dalton, N. S., & Rogers, Y. (2008). Collaboration and interference: awareness with mice or touch input. In Proc. of CSCW, ACM.
11. Hutchins, E. (1995). Cognition in the Wild. Cambridge, MA: MIT Press.
12. Isenberg, P., Isenberg, T., Hesselmann, T., Lee, B., Von Zadow, U., & Tang, A. (2013). Data visualization on interactive surfaces: A research agenda. Computer Graphics and Applications, IEEE, 33(2), 16-24.
13. Killeen, P.R., Glenberg, A.M. (2010). Resituating cognition, Comparative Cognition & Behavior Reviews, 4, 66-85.
14. Kirsh, D., Maglio, P. (1994) "On distinguishing epistemic from pragmatic action," Cognitive Science, 18(4), 513-549.
15. Kirshner, D. & Whitson, J. A., Eds. (1997). Situated cognition: Social, semiotic, and psychological perspectives. Mahwah, NJ: Lawrence Erlbaum.
16. Klum, S., Isenberg, P., Langner, R., Fekete, J. D., & Dachselt, R. (2012). Stackables: combining tangibles for faceted browsing. In Proc. of AVI. ACM.
17. Marshall, P., Hornecker, E., Morris, R., Sheep Dalton, N., & Rogers, Y. (2008). When the fingers do the talking: A study of group participation with varying constraints to a tabletop interface. In Proc. of ITS, IEEE.
18. Martin, T., Schwartz, D.L. (2005). Physically Distributed Learning: Adapting and Reinterpreting Physical Environments in the Development of Fraction Concepts, Cognitive Science, 29, 587-625.
19. Microsoft. (n.d.). Microsoft PixelSense. Retrieved from Microsoft PixelSense: http://www.microsoft.com/en-us/pixelsense/default.aspx
20. Nersessian, N.J. (2002). The cognitive basis of model-based reasoning in science. The cognitive basis of science. P. Carruthers, S. Stich and M. Siegal, Eds, Cambridge University Press: 133-153.
21. Nersessian, N.J. (2008). Creating scientific concepts. Cambridge, MA, MIT Press.
22. Patten, J., & Ishii, H. (2000). A comparison of spatial organization strategies in graphical and tangible user interfaces. In Proc. of DARE, ACM.
23. Rheinberger, H.-J. (1997). Toward a history of epistemic things: Synthesizing proteins in the test tube. Stanford, CA, Stanford University Press.
24. Shaer, O., Leland, N., Calvillo-Gamez, E. H., & Jacob, R. J. (2004). The TAC paradigm: specifying tangible user interfaces. Personal and Ubiquitous Computing, 8(5).
25. Shaer, O., Mazalek, A., Ullmer, B., & Konkel, M. (2013). From Big Data to Insights: Opportunities and Challenges for TEI in Genomics. In Proc. of TEI. ACM.
26. Shaer, O., Strait, M., Valdes, C., Feng, T., Lintz, M., & Wang, H. (2011). Enhancing genomic learning through tabletop interaction. In Proc. of CHI, ACM.
27. Ullmer, B., Ishii, H., & Jacob, R. J. (2005). Token+ constraint systems for tangible interaction with digital information. ACM TOCHI, 12(1), 81-118.
28. Ullmer, B., Ishii, H., & Jacob, R. J. (2003). Tangible query interfaces: Physically constrained tokens for manipulating database queries. In Proc. of Interact.
29. Ullmer, B., Ardaud, G., Dell, C., and et al. Employing and extending mass-market platforms as core tangibles. In Proc. of TEI'12 (Works in Progress), 2012.
30. Valdes, C., Eastman, D. Grote, C., Thatte, S., Shaer, O., Mazalek, A., Ullmer, B., Konkel, M.K. (2014) Exploring the Design Space of Gestural Interaction with Active Tokens through User-Defined Gestures. In Proc. CHI 2014.
31. van den Hoven, E., Mazalek, A. (2011). Grasping gestures: Gesturing with physical artifacts. AI EDAM 25(3).
32. Voida, S., Tobiasz, M., Stromer, J., Isenberg, P., & Carpendale, S. (2009). Getting practical with interactive tabletop displays: designing for dense data, fat fingers, diverse interactions, and face-to-face collaboration. In Proc. of the ITS, ACM.
33. Wilson, M. (2002). Six views of embodied cognition, Psychonomic Bulletin & Review 9(4): 625-636.
34. Wobbrock, J. O., Morris, M. R., & Wilson, A. D. (2009). User-defined gestures for surface computing. In Proc. of CHI, ACM.
35. Xu, W., Chang, K., Francisco, N., Valdes, C., Kincaid, R., & Shaer, O. (2013). From wet lab bench to tangible virtual experiment: SynFlo. In Proc. of TEI. ACM.
36. Zigelbaum, J., Horn, M. S., Shaer, O., & Jacob, R. J. (2007). The tangible video editor: collaborative video editing with active tokens. In Proc. of TEI, ACM.

Animation Killed the Video Star

Voicu Popescu
Purdue University
popescu@purdue.edu

Nicoletta Adamo-Villani
Purdue University
nadamovi@purdue.edu

Meng-Lin Wu
Purdue University
wu223@purdue.edu

Suren D. Rajasekaran
Purdue University
srajase@purdue.edu

Martha W. Alibali
University of Wisconsin
mwalibali@wisc.edu

Mitchell Nathan
University of Wisconsin
mnathan@wisc.edu

Susan Wagner Cook
University of Iowa
susan-cook@uiowa.edu

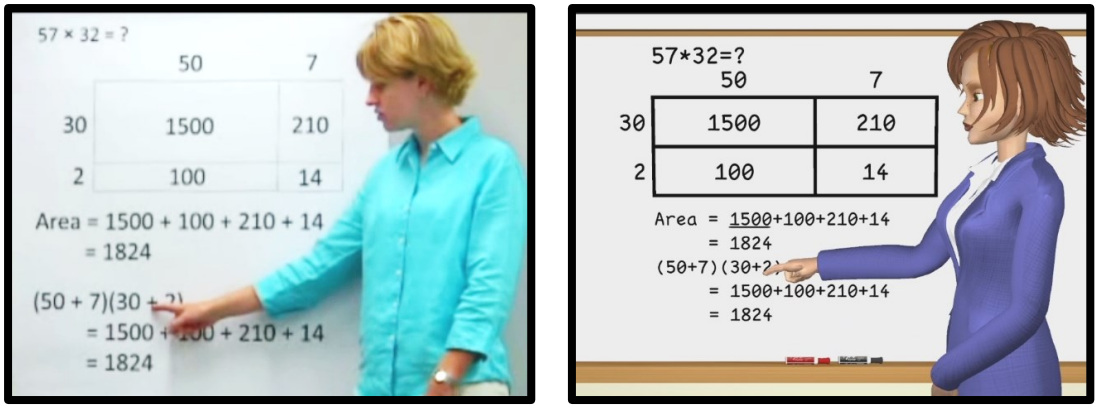


Figure 1: Visual stimulus used in instructor gesture research: video (left), animation (right).

ABSTRACT
In this position paper, we describe a novel approach for creating visual stimuli for research on gesture in instruction. The approach is based on a system of computer animation instructor avatars whose gesture is controlled with a script. Compared to video recording instructor actors, the approach has the advantage of *efficiency*—once the script is written, it is executed automatically by the avatar, without the delay of script memorization and of multiple takes, and the advantage of *precision*—gesture is controlled with high fidelity as required for each of many conditions, while all other experiment parameters (e.g. voice tone, secondary motion) are kept constant over all conditions, avoiding confounds. We have begun implementing the approach, and we will test it in the context of connecting mathematical ideas in introductory algebra.

Author Keywords
Instructor gesture, instructor avatar, computer animation, connecting mathematical ideas, introductory algebra.

Copyright 2014 ACM

Permission to make digital or hard copies of part or all 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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org or Publications Dept., ACM, Inc., fax +1 (212) 869-0481.

ACM Classification Keywords
H.5.1 Multimedia Information Systems—Animations
H.5.m. Information interfaces and presentation (e.g., HCI)—Miscellaneous.

INTRODUCTION
Computer animation has the potential to become a powerful platform for research on gesture in instruction, overcoming many of the challenges associated with the traditional approach of creating lesson stimuli by video recording instructors. We have begun implementing a system of instructor avatars that we plan to test in introductory algebra (Figure 1).

Importance of gesture in education
Communication is an integral element of nearly all forms of instruction, including tutoring, peer collaboration, and classroom instruction. Instructors use a range of modalities to communicate, including language, drawing, writing, and non-verbal behaviors such as gestures.

Many recent, naturalistic studies have highlighted the importance of teachers' gestures, particularly in the domains of mathematics and science. For example, researchers have argued that teachers use gestures as a "semiotic resource" for developing and refining ideas [e.g., 2, 7], as a means to link ideas and representations [1], and as a means of fostering joint attention or shared understanding [8], particularly in cases where students are having difficulty with the material.

However, empirical data that address whether teachers' gestures are actually beneficial for students' learning are relatively scarce. A handful of studies have compared lessons with gestures to lesson without gestures, and shown that lessons with gestures lead to greater student learning [e.g., 3, 4, 9] Although these studies represent a valuable first step, they do not address the range of variation in teachers' gestures, or whether some types of gestures are more effective than others. What is needed are empirically validated recommendations about what type of gestures are most effective. However, one challenge in research on gesture in instruction is creating appropriate lesson stimuli.

Challenges of video

A common approach is to create video recordings of lessons that vary in whether and how the teacher uses gestures; lessons are scripted so that, ideally, only gesture varies between lessons. This allows for strong inferences to be drawn about whether the gestures contribute to students' learning. In some cases, the same audio track is used across conditions in a study, and the teacher-actor "lip-syncs" the speech while producing the scripted gestures. This approach has the advantage of realism—video depicts the instructors exactly as students see them in classrooms.

However, this methodological approach also has many challenges. It is often difficult for teacher-actors to memorize all the needed scripts for each condition in a study. Even more problematic, teacher-actors have to keep all aspects of behavior other than gesture the same between conditions, including eye gaze, posture, and facial expression. Because it is effortful to simultaneously follow the script and manage other aspects of behavior, the resulting videos often look unnatural, and they often require many "takes".

The challenges do not end there. It is also often challenging for teacher-actors to manage auxiliary visuals, such as writing on the board, while also producing scripted gestures. For this reason, many experiments give up on developing visuals incrementally, in real time, therefore losing the benefits of focusing student attention and of controlling complexity by conveying the solution progressively.

ANIMATION: POTENTIAL, CHALLENGES, SOLUTIONS

Potential of animation

Advances in technology have enabled developing and deploying computer animation applications on consumer level computing technology such as personal computers, laptops, tablets, and smart phones. A computer animation character is well suited to serve as an instructor proxy in studies of gesture: a humanoid character can make any gesture a human instructor can make, and an animation character has perfect memory and infinite energy.

Once the script is complete, the instructor avatar can render the stimulus in one "take", with no errors. Slight modifications of the script produce the stimuli for any number of additional conditions, with perfect control over secondary parameters. The avatar does not have to worry about remembering what to say and do, and when, and the resulting lessons can be, paradoxically, less contrived and more natural than in the case of human teacher-actors.

Instructor avatars are malleable, so they simplify generating stimuli for studies that take into account instructor characteristics such as gender, personality, and age. Finally, if desired, animation can be bigger than life. The whiteboard can magically animate concepts, the avatar can have perfect drawing skills, and the avatar can have cartoon-character-like qualities, such as a stylized appearance and an extroverted personality reflected through speech and action (e.g., backflips to celebrate correct answer).

Challenges of animation

The first question is whether what is learned about gesture in the context of instructor avatars does transfer to human instructors. Preliminary analyses of data collected in our first studies indicate that, yes, benefits of gesture previously measured with human instructors are replicated when instructor avatars are used. Moreover, effective instructor avatars are valuable in and of themselves, enabling the creation of effective digital learning materials.

However, before instructor avatars can become an effective platform for gesture research, several challenges have to be addressed. First, the animation needs to be of sufficient quality. We define animation quality in this context both at a low level, which includes rendering nuanced gestures with precision, clarity, and life-like quality, and at a high level, which includes conforming to avatar behavior and speech rules for effective social and teaching interactions [12, 13]. Second, for the approach to scale, one has to be able to create stimuli without the prerequisites of artistic and programming talent. Gesture researchers need to be able to create their own experiment stimuli without assistance from digital artists and programmers.

Solutions to animation challenges

A solution to these challenges is a system of computer animation instructor avatars that are controllable through a script. The script is a text file that contains directions for the avatar, much the same way scripts are used by researchers to direct human teacher-actors when generating video stimuli. No programming expertise is needed; instead, the approach relies on scripting, an interface already familiar to researchers. Compared to the natural language and loose rules used when scripting a human actor, the scripting of the avatar has to be done in a language that can be interpreted automatically, i.e., an English-like language that is complete, concise, and unambiguous.

The need for artistic talent is removed by using a database that already contains all the digital art assets needed for generating the stimuli. These assets include the computer animation characters for the avatars, pre-recorded joint

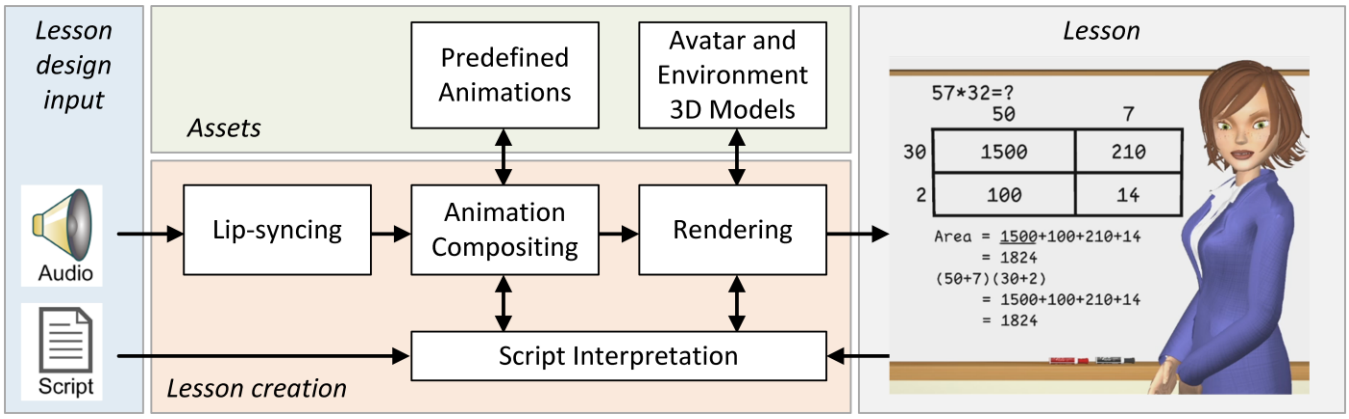


Figure 2: Overview of system of computer animation instructor avatars for generating lesson stimuli for gesture research.

angle values for complex animations, and auxiliary visuals (e.g., a whiteboard capable of displaying any polynomial multiplication). Simple animations, such as pointing at a specific location on the board, are computed automatically on-the-fly, as needed, using animation algorithms.

For animated avatars to be effective and believable instructors, their animation should be of life-like quality. To achieve this, the animation must adhere to fundamental principles. The 12 principles of animation [5] are a set of procedures taught as fundamental rules of the "language of movement" at the Walt Disney Studio in the late 1930s. Five of these principles are essential in our context.

Anticipation. Anticipation is preparation for action. Actions are preceded by smaller actions in the opposite direction (e.g., to jump with joy, an avatar has to bend its knees first). Without anticipation, gestures appear abrupt, stiff, and unnatural.

Follow-through and overlapping action. Follow-through is the termination of an action. "Actions very rarely come to a sudden and abrupt stop, they are generally carried past their termination point" [6]. For example, when the avatar lands after jumping, it has to bend its knees again before assuming the standing pose. Moreover, movements of different parts of a living creature do not occur at the same time; motions start, move, and stop at different points and at different rates. Lack of adherence to this principle yields mechanical motion with an undesirable "stop-start" quality.

Secondary Action. Any movement of a living creature is composed of a primary action and multiple secondary actions. For example, when the avatar makes a beat gesture with its right hand, the avatar might sketch a similar gesture of reduced amplitude with its left hand. Representing different types of interconnected motions is fundamental to the creation of "a believable whole within any movement" [6]. Failure to adhere to this principle results in disjointed, puppet-on-a-string-like characters; the motion does not appear to come from within but rather to be applied from an external source.

Slow-out and slow-in. Body parts do not usually move at constant speed; they show a certain degree of acceleration and deceleration. Actions should progressively accelerate out of a key pose and decelerate into a resting pose. Failure to adhere to this principle confuses the viewer with sheer defiance of the laws of physics.

Arcs. Living creatures cannot perform linear movements. The same way the laws of physics do not tolerate abrupt changes in velocity magnitude, they do not tolerate abrupt changes of velocity direction. Failure to adhere to this principle results in mechanical, robotic motion.

A SYSTEM OF INSTRUCTOR AVATARS

We are developing a system of animation avatars with the overall architecture given in Figure 2.

The lesson designer provides two types of input: instructor audio and lesson script. The designer records the instructor audio using voice talent that matches the instructor avatar. Synthesizing the audio directly from text would result in a robotic voice. The audio file is used by a lip-syncing module to derive the facial animation of the instructor avatar needed to utter the recorded words.

The lesson script is a text file that controls the avatar and the auxiliary visuals. The scripting language supports a small number of commands, each with subcommands and arguments. For example the command SAY *audioA* 1.2 10.5 makes the avatar say the instructor audio recording stored in file *audioA*, from seconds 1.2 to 10.5. The command GESTURE DB *idkGesture* makes the avatar gesture a pre-defined "I don't know gesture" by raising its shoulders, bending its arms at 90º, and turning its hands such that the palms point upwards. The command GESTURE POINT LEFT INDEX *targetA* makes the avatar point with its left index to *targetA,* which was previously defined to correspond, for example, to a point on the graph. The scripting language also allows precisely synchronizing gesture and speech by defining time points relative to the beginning of an audio file, which can then be used as starting times for gestures.

The script is interpreted automatically. The necessary animations are either retrieved from the assets database (for GESTURE DB commands) or from the lip-syncing module, or they are computed on the fly through inverse kinematics animation algorithms (e.g., for the GESTURE POINT command). The various animation elements are composited and applied to the avatar computer animation character.

We have developed an avatar named Julie (Figures 1 and 2). Julie is a partially segmented 3D animation character comprised of eight polygonal meshes with a total polygon count of 171,907. Julie's body is rigged with a skeletal deformation system that includes 69 joints. Her face is rigged with a combination of joint deformers and blendshape deformers. 29 joints are used for opening and closing the mouth and for eye/eyebrow deformations, 20 blendshape targets are used for facial deformations, and eight targets provide the visemes required for animation of dialogue (four for the consonant sounds and four for the vowel sounds). The polygon meshes are skinned to the skeleton with a dual quaternion smooth bind with a maximum number of influences of five.

The avatar is rendered in the environment to obtain the lesson stimuli. Initially, we will focus on non-interactive lessons. Interactive lessons will be supported by collecting input from learners to which the avatar will react, according to the script. Once a script is written for a first condition, the scripts for the other conditions are generated by modifying the first script. This procedure is efficient and keeps all parameters constant except for the gesture conditions that are the target of study.

Whenever possible, a gesture is implemented algorithmically through inverse kinematics. This includes all deictic gestures which abound in instructors' non-verbal communication repertoires. The advantage of algorithmic animation is efficiency—one inverse kinematic algorithm animating the arm allows pointing anywhere on the whiteboard by simply changing the target parameter. The challenges of algorithmic animation are lower quality, as rigorous conformance to animation principles cannot be achieved in target independent fashion, and lack of support for complex gestures, which have to be animated manually by digital artists.

The reliance on the database of predefined animations brings scripting language simplicity. The language does not have to support the definition of new avatar poses and the linking of poses to form new gestures. The tradeoff is that users cannot define new gestures, which can be a potential bottleneck preventing scalability to other concepts, subject matters, and student age groups. Whenever a new gesture is needed, that gesture has to be defined by a digital artist using an animation software system (e.g. Maya) and added to the database.

One option is to extend the scripting language to support a precise and inherently complex description of the behavior of the avatar, as was done for example in the SmartBody system which relies on SAIBA framework's Behavior Markup Language standard [14].

However, we believe that the important advantages of scripting language simplicity and similarity to the scripts already used by gesture researchers when creating video stimuli do not have to be sacrificed. We anticipate that the database of gestures shared online will quickly grow to cover all gestures needed. For example we have already defined a complex vocabulary of charisma gestures covering 18 parameter combinations: inward, vertical, or outward; one hand, two hands unsynchronized, or two hands parallel; small or large amplitude. Moreover, gestures can be retargeted to any animation character with the same skeletal structure.

As prior research indicates, the precise timing between gesture and utterances is essential [10, 11]. The scripting language does allow defining time steps with fine granularity (i.e., milliseconds). However, finding the precise time step presently requires a trial and error approach. Proceeding with a binary search requires ten tries to find a time stamp with millisecond accuracy within a one second time interval. We will investigate direct synchronization using the textual representation of speech. We will leverage the availability of the text, which simplifies the audio to text mapping.

INSTRUCTOR GESTURE RESEARCH

We will use experimental lessons created with the instructor avatars to address a several key questions about how teachers' gestures contribute to student learning. Our initial focus will be on instruction in algebra, which typically involves instructors drawing links across multiple related representations. For example, in a lesson about polynomial multiplication, an instructor might seek to connect a procedure for multiplying polynomials that is expressed in symbolic form with a procedure that is depicted using an area model (see Figure 1). We will eventually expand our focus to other types of mathematical and statistical content (e.g., geometry, confidence intervals). Where possible, we will compare our findings to those obtained in research with human teachers.

As a starting point, we will investigate whether students learn more about links between ideas when teachers gesture to corresponding ideas sequentially or when they do so simultaneously (i.e., pointing to two corresponding ideas with the two hands). We will also investigate whether students learn more about links between ideas when teachers use gesture to make element-by-element mappings or more global mappings. We will also ask whether deictic or representational gestures are more communicatively effective. The avatar system will allow us to design focused tests of these and many other research questions.

LONG TERM GOALS

Our long term goal is to provide a powerful system for creating stimuli for research on gesture and beyond. Although video *did* kill the radio star (as claimed by the Buggles in their hit song in 1979), we do not foresee that animation will ever completely replace video recordings of human instructors. Video stimuli are by definition photorealistic, and the talent of a gifted instructor will continue to elude lossless translation into scripted animation for the foreseeable future.

The instructor avatar approach opens the door for research on learning personalization, where the avatar offers the fine-grained instructor control needed to adapt to a variety of learner characteristics. Our initial focus is on introductory algebra with middle and high school learners, but the approach can be extended to other gestures, topics, disciplines, and learner age groups. The system of instructor avatars can also become a powerful platform for creating interactive digital learning activities, and for creating materials for pre- and in-service teacher professional development. Beyond standard education, the system of instructor avatars can also support multimodal instruction for special education. Finally, our system can be repurposed for developing non-verbal communication skills for other forms of public speaking, beyond instruction.

ACKNOWLEDGMENTS

We thank Jian Cui, Howard Friedman, and Katherine Duggan who have helped develop the initial version of the computer animation instructor avatar system on which the present system is based. We are particularly grateful to Jian Cui for his help as we extended the initial system. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A130016, and by the National Science Foundation, through Grant 1217215. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences, of the U.S. Department of Education, or of the National Science Foundation.

REFERENCES

1. Alibali, M. W., Nathan, M. J., Wolfgram, M. S., Church, R. B., Johnson, C. V., Jacobs, S. A., & Knuth, E. J. (2014). How teachers link ideas in mathematics instruction using speech and gesture: A corpus analysis. *Cognition & Instruction, 32*(1), 65-100.

2. Arzarello, F., Paola, D., Robutti, O., & Sabena, C. (2009). Gestures as semiotic resources in the mathematics classroom. *Educational Studies in Math., 70*(2), 97-109.

3. Church, R. B., Ayman-Nolley, S., & Mahootian, S. (2004). The role of gesture in bilingual education: Does gesture enhance learning? *International J. of Bilingual Education & Bilingualism*, 7, 303–319.

4. Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child Dev., 84*(6), 1863-1871.

5. Johnston, O., & Thomas, F. The illusion of life: Disney animation. *Disney Editions;* Rev Sub edition, Oct 5, 1995.

6. Lasseter, J. Principles of traditional animation applied to 3D computer animation. *Computer Graphics* (1985), 21(4): 35-44.

7. Rasmussen, C., Stephan, M., & Allen, K. (2004). Classroom mathematical practices and gesturing. *J. of Mathematical Behavior, 23*, 301-324.

8. Reynolds, F. J., & Reeve, R. A. (2001). Gesture in collaborative mathematics problem solving. *J. of Mathematical Behavior, 20*(4), 447-460.

9. Valenzeno, L., Alibali, M. W., & Klatzky, R. L. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemp. Educ. Psych., 28*, 187–204.

10. Striegnitz, K., Tepper, P., Lovett, A. & Cassell, J. (2008). Knowledge representation for generating locating gestures in route directions. In K.R. Coventry, T. Tenbrink & J. Bateman (Eds.), *Spatial Language and Dialogue (Explorations in Language and Space)*. Oxford: Oxford University Press.

11. Nass, C., Isbister, K. & Lee, E.-J. (2000). Truth is beauty: Researching embodied conversational agents. In *Embodied conversational agents* (pp.,374-402). Cambridge, MA: MIT Press.

12. Finkelstein, S., Ogan, A., Walker, E., Muller, R., & Cassell, J. (2012). Rudeness and rapport: Insults and learning gains in peer tutoring. In *Proceedings of Intelligent Tutoring Systems*,.Berlin: Springer..

13. Ogan, A., Finkelstein, S., Mayfield, E., Matsuda, N., & Cassell, J. (2012). Oh dear Stacy! Social interaction, elaboration, and learning with teachable agents. In *Proceedings of CHI 2012.* Austin, TX.

14. Thiebaux, M., et al., (2008). SmartBody: behavior realization for embodied conversational agents. In *Proceedings of 7th international conference on autonomous agents and multiagent systems* (pp. 151-158). Estoril, Portugal.

Help systems for gestural interfaces and their effect on collaboration and communication

Davy Vanacken, Anastasiia Beznosyk, Karin Coninx
Hasselt University - tUL - iMinds, Expertise Centre for Digital Media
Wetenschapspark 2, 3590 Diepenbeek, Belgium
{firstname.lastname}@uhasselt.be

ABSTRACT

Due to limited discoverability and standardization, people often have difficulties with learning and remembering all the gestures of a gestural interface. To overcome these difficulties, help systems are integrated to enable users to quickly explore the available gestures. These systems typically target single-user applications and research is mainly focused on studying the efficiency of various kinds of help. In this paper, we study textual and animated help in a collaborative tabletop game. We not only consider the efficiency of the help systems, but we also examine how they impact collaboration and communication. Our study shows that animated help has a positive effect on the level of collaboration, as users work together to explore and learn the game.

INTRODUCTION

Interactive tabletops have enabled new types of co-located collaborative applications. Tabletops invite users to gather around the table and typically allow for playful interaction through a gestural interface. As a result, tabletops are popular in the area of digital entertainment, offering an experience that is similar to the typical board games, where several people sit around a table, engaged with both the game and each other. However, the use of a gestural interface also introduces new challenges.

An important advantage of a GUI is that commands do not have to be memorized. Instead, the possibilities of the GUI can be discovered by exploring the various widgets. In addition, as a result of the standardization of GUIs, a user can usually depend on prior knowledge of other applications. Gestural interfaces, on the other hand, typically provide little to no means of discoverability and there are few common conventions, resulting in a lack of consistency across applications and a low memorability of the gestures [5]. This is especially true for cooperative gestures, which are usually more complex, as users need to synchronize and coordinate their actions.

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. Copyrights for components of this work owned by others than the authors must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission from the authors.
Gesture-based Interaction Design: Communication and Cognition,
http://hci.uncc.edu/~mmaher9/CHI-gesture-interaction/
CHI 2014 Workshop, April 26 2014, Toronto, ON, Canada
Copyright is held by the authors.

To overcome these challenges, a variety of help systems have been proposed (a more comprehensive overview of literature is provided in [6]): videos of available gestures [8], animated virtual hands that demonstrate gestures within the context of use [7], dynamic visualizations of the current state of recognition and possible completion paths [1, 2], the use of physical metaphors [4], etc. These systems are typically evaluated in a single-user context, with a focus on learning efficiency.

Having multiple users interact simultaneously in a co-located setting introduces some additional factors that are interesting to investigate. In a multi-user setting, users can not only learn gestures by using the help system, but they can also observe each other and explain gestures to each other. In addition, the type of help system that is provided might have an influence on collaboration and communication. This paper presents the results of a study on two commonly used types of help systems in a collaborative tabletop game: textual and animated help. With this study, we briefly look into learning efficiency, but we mainly aim to shed some light on how help systems can contribute to the collaborative experience.

GESTURE-BASED PUZZLE GAME

We conducted a study with a puzzle game on a multi-touch tabletop (Figure 1). Each puzzle piece is represented as a cube, with a picture on one of its sides. However, the game is designed in such a way that it encourages and sometimes requires two users to collaborate tightly, so not all pieces behave in the same way. To enforce actions to be executed by a particular user and to log data about individual actions, each user is assigned an avatar (the purple and orange 'disks').

We briefly explain the interactions in the game. To select a puzzle piece, move your avatar close to it and tap the avatar twice. To deselect, double tap the avatar again. To move a piece, press the avatar with one finger to lift the piece up from the floor and at the same time drag the piece around with another finger. To rotate a piece, first lift it up. To rotate in 2D, put two fingers on the piece and move one of them. To rotate in 3D, put two fingers on the piece and then spin the piece around by moving over it with a third finger. To enlarge a small piece, move a special purpose piece close to it and while both pieces are selected, simultaneously press the two confirmation buttons at the top of the screen. Heavy pieces move and rotate slower than light pieces. However, a heavy piece can be selected and lifted by two avatars simultaneously to increase the speed of those actions. The game carries on

Figure 1. Two persons playing the puzzle game. One player interacts with a puzzle piece while her partner reads the help.

until all small pieces are enlarged, and the pieces are (more or less) in the correct orientation and position.

Two different help systems were implemented to explain the gestures: textual help with an annotated image and animated help with annotations (Figure 2). Each user has an individual panel that displays the help on his or her side of the screen. In both cases, users are able to interact with the game while consulting the help, so they can try gestures while reading text or watching an animation. Both the textual and animated help are individualized to some degree, as the image and animations always show the avatar of the corresponding user.

The animated help visualizes the gestures by animating virtual fingers on top of a scaled down copy of the relevant parts of the user interface (i.e. the avatars and puzzle pieces). To further clarify the animations, we annotate them with a minimal amount of text. The annotations point out important aspects or nuances of the interaction, such as the type of puzzle piece that is involved. We used pilot studies to optimize the speed and length of the animated sequences, to make sure that people are able to read the annotations comfortably and interpret all the actions. This resulted in animations that take around 20 to 25 seconds.

When the game starts, the help panel automatically opens. In contrast to the textual help that offers the explanations of all the possible gestures at once, animations are shown one by one. At the start, an animation first shows how to select a puzzle piece. Once the user successfully selects one of the pieces, the game automatically demonstrates how to move it and after some time how to rotate it, and in case of a small piece, how to enlarge it. If a single user selects and manipulates a heavy puzzle piece, the game displays an animation as a reminder after a while, to indicate that this type of piece can be manipulated faster with the help of another user.

Help can also be accessed at any time during the game by pressing the question mark button in the upper corner of the screen. In case of textual help, the button simply opens or closes the text. In case of animated help, it opens a context-sensitive menu that only lists actions that are currently available to the user. If the user has nothing selected, only 'select' will be available in the menu. Once the user has selected a puzzle piece, all the actions that are available on that type of piece will be listed in the menu, as illustrated in Figure 2.

Experimental Design

We recruited 28 volunteers (3 female and 25 male, ranging in age from 22 to 45), who were randomly divided in 14 pairs. We used a between-subjects design: 7 pairs had to solve the puzzle with the textual help, and 7 pairs with the animations. The condition was randomly assigned to the pairs.

The participants were asked to read a brief introduction beforehand, but we did not explain the help system or gestures. During the experiment, we encouraged the think-aloud protocol and we filmed each session. Two observers also took notes, but participants were not allowed to ask the observers any questions. The participants had a printed image of the finished puzzle at their disposal. It took approximately 30 minutes for each pair to complete the puzzle.

The following dependent variables were logged during the puzzle task or extracted from the recorded videos: (1) task completion time; (2) number of times help was accessed; (3) amount of time players acted individually; (4) amount of time players acted collaboratively; (5) amount of time players watched animations together; (6) amount of time players read text together; (7) amount of time players communicated; (8) when players communicated for the first time; (9) when players collaborated for the first time.

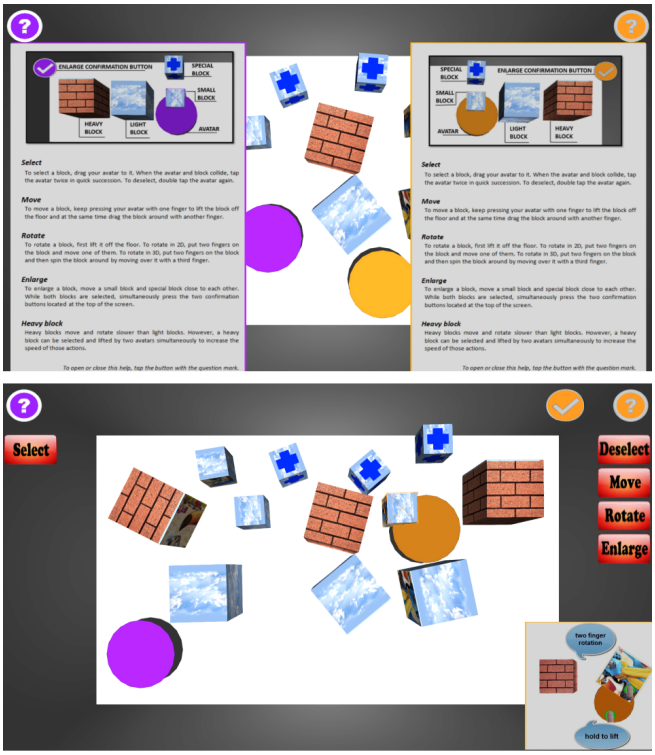


Figure 2. Two help systems that explain the gestures: textual help with an annotated image (top) and animated help with annotations (bottom).

Subjective data was collected through a paper-based post-experiment questionnaire, using a visual analogue scale: participants were asked to give a rating by placing a mark at the appropriate position on a continuous ten-centimeter line, representing the point between 'not at all' and 'very much' that they felt represented their perception best.

RESULTS AND DISCUSSION

First, we analyze the use of the textual help and animations. Afterwards, we analyze how the different types of help contribute to the level of collaboration and communication. The findings are based on performance logs, analysis of the video recordings, the observers' notes, and the subjective data from the questionnaires (more details on the analysis of the data are provided in [3]).

The Help Systems

Although the primary focus of our study was on how a help system can influence the behavior of users when simultaneously interacting in a co-located setting, the questionnaire included a few questions on the efficiency of the help system. It allows us to verify that both help systems were effective in terms of accessibility and understandability. The differences between textual help and animations that are reported in this section are statistically insignificant based on a paired samples t-test, which shows that both systems served their purpose more or less equally.

All pairs successfully completed the puzzle, and they were able to use the help system without any explanation. With the textual help, some participants first read the complete text and then tried the various gestures, while others only read the first part and tried that particular gesture before reading the next part. In some cases, one participant read part of the text out loud, while the other participant performed the gestures. All participants had to reopen the textual help at least a handful of times at a later stage.

Although one pair never used the menu and learned all gestures through the automatically triggered animations, other participants mostly ignored them, except for the one that was shown when the game started. Once participants successfully selected a piece, the game automatically triggered the animation on how to move, but that animation was typically ignored as they first tried deselecting and selecting other pieces. Afterwards, they explicitly initiated the next animation through the menu. This behavior makes it difficult to automatically show help at appropriate times, although simply waiting for the participant to be inactive for a while before triggering a new animation may somewhat lessen the problem.

When asked if the help explained the gesture clearly, animations ($\bar{x}$=5.6, s=1.9) were rated higher than text ($\bar{x}$=4.1, s=2.4). The animations were annotated to highlight important aspects, thereby combining the advantage of text and animations to a certain degree. However, we took a minimalistic approach with the annotations, to avoid having to read too much and animations becoming lengthier as a result. Several participants made errors because they overlooked something important that was not included in the annotations. To select a puzzle piece, for example, participants had to double tap on the avatar, but occasionally they tried to double tap on the piece instead. We wrongfully presumed that the location would be sufficiently clear from the animation itself. This issue could have been avoided by extending the annotation. The textual help was complemented with an image of all the essential components, but a graphical representation (e.g. a small storyboard) of each gesture may improve its clarity.

Figure 3. Two participants watch the same animation together.

The cooperative gestures were more complex, requiring participants to collaborate tightly. Results from the questionnaire indicate that the textual help ($\bar{x}$=5.9, s=2.0) performed marginally worse than animations ($\bar{x}$=6.3, s=1.7) in making clear how to perform collaborative tasks. However, in our observations we clearly noticed that quite a few participants struggled with enlarging a small puzzle piece in case of textual help. It took them a while to figure out that they needed to collaborate, as some first tried, for instance, to select two pieces with only one avatar, which is not possible.

Animations ($\bar{x}$=6.5, s=2.1) allowed participants to discover gestures slightly quicker than textual help ($\bar{x}$=5.3, s=2.5), thanks to the visual nature that makes interpretation quicker. We regularly noticed participants trying a gesture while reading the text out loud, and actually counting the number of fingers they put down on the tabletop. A recurrently observed issue was participants accidentally performing a gesture without truly knowing how they did it or what happened. Additionally, some participants executed gestures in an uncomfortable or uncontrolled manner. Our help only explains gestures, but does not offer feedback during or after their execution, except for green dots to confirm recognized touches and the actual effect on the puzzle piece when an action is performed correctly. An approach like Gesture Play [4] provides both feedback and feedforward during execution, and clearly indicates whether the gesture is being performed appropriately. This kind of feedback improves a user's ability to successfully execute an action the way it was intended to be.

Participants also rated different ways of learning gestures. In both conditions, they learned the most from consulting the help (text: $\bar{x}$=7.2, s=2.4; animations: $\bar{x}$=7.6, s=2.0), from talking to their partner (text: $\bar{x}$=7.1, s=2.2; animations: $\bar{x}$=6.9, s=1.0), and finally from watching their partner's actions (text: $\bar{x}$=5.9, s=2.7;

Table 1. Comparison of collaboration and communication for textual and animated help.

Parameter	Text		Animation		Statistics
	$\bar{x}$	s	$\bar{x}$	s	
Collaboration					
Individual activity, %	31.4	13.8	16.9	7.5	$t(12)=2.44$, $p=0.031$
Collaborative activity, %	29.0	9.2	41.1	9.6	$t(12)=-2.39$, $p=0.034$
Using help together, %	1.8	1.3	5.0	2.3	$t(12)=-3.28$, $p=0.007$
Time of first collaboration, %	37.7	16.5	19.1	15.5	$t(12)=2.17$, $p=0.05$
Communication					
Amount of communication, %	32.6	13.3	37.3	16.1	$t(12)=-0.59$, $p>0.001$
Time of first communication, %	5.7	2.7	1.2	0.8	$t(12)=4.15$, $p=0.001$

animations: $\bar{x}=5.7$, $s=2.3$). We expected participants to learn more by watching each other than by talking, but our observations clarify these findings. First of all, explaining something verbally allows you to continue your current task, a behavior that we observed repeatedly. In addition, participants regularly explained an interaction when seeing the other person doing it wrong. They first tried to correct the mistakes of the other participant verbally, and only if that did not succeed, they demonstrated the action. These results highlight that social learning is an important aspect in a multi-user environment.

An interesting conclusion in the context of gaming is that participants reported that textual help ($\bar{x}=5.1$, $s=3.2$) took them out of the game experience, which was less of an issue with animated help ($\bar{x}=3.3$, $s=1.8$).

Level of Collaboration and Communication

The video analysis was performed by the two observers according to a list of guidelines. We considered activities to be completely individual if the participants really focus on two separate tasks and do not communicate in any way, as shown for example in Figure 1. Collaboration, on the other hand, includes two participants either working closely together to accomplish a task (e.g. enlarge a small puzzle piece, manipulate a heavy piece together), communicating with each other (e.g. discussing a strategy, explaining gestures), or consulting the help together (e.g. watching an animation together, as shown in Figure 3). Table 1 summarizes the results regarding individual and collaborative performance, as well as communication.

We calculated percentages by dividing the individual and collaborative time by the total amount of time it took to complete the task. The results clearly show a statistically significant increase ($t_{12}=2.44$, $p=0.031$) of completely individual work in case of textual help ($\bar{x}=31.4\%$, $s=13.8$) compared to animated help ($\bar{x}=16.9\%$, $s=7.5$), and a statistically significant decrease ($t_{12}=-2.39$, $p=0.034$) in collaboration ($\bar{x}=29.0\%$, $s=9.2$) compared to animated help ($\bar{x}=41.1\%$, $s=9.6$). As a result of the higher level of collaboration, participants also reported to be more aware of their partner's actions during the game in the animation condition ($\bar{x}=7.7$, $s=1.7$) compared to the text condition ($\bar{x}=6.2$, $s=2.3$).

One cause of these differences is the moment participants started their first collaboration, as it took them significantly longer ($t_{12}=2.21$, $p=0.047$) with textual help ($\bar{x}=511s$, $s=316$) than with animations ($\bar{x}=214s$, $s=164$) (Table 1 includes these results as percentages in stead of absolute values). Logically, this moment also influenced the subjective experience of the participants, as we found a significant negative correlation ($r=-0.55$, $p=0.043$) that implies that the later the collaboration starts, the lower the perceived level of collaboration is. The fact that it took longer for the first collaboration to happen is partially attributable to some pairs reading the whole text before starting to interact. Nonetheless, it does not completely explain the difference in amount of individual activities and collaboration, as the other pairs also had to spend some time watching other animations once they figured out the basic interactions.

There are a few other factors to consider when analyzing the level of collaboration. First of all, if we look specifically at how much time participants spent consulting help together (thus reading the same text or watching an animation together), we see a statistically significant difference ($t_{12}=-3.28$, $p=0.007$) between textual ($\bar{x}=1.8\%$, $s=1.3$) and animated help ($\bar{x}=5.0\%$, $s=2.3$), which accounts for part of the aforementioned difference in amount of individual and collaborative activities. A plausible reason is that animations are easier to watch from different angles, while text is more difficult to read from an angle. Secondly, if a participant is manipulating a heavy puzzle piece, the game displays an animation after a while to remind the participant that the piece can be manipulated faster with two. The animated help thus invites participants to collaborate more.

We also asked participants to rate their subjective experiences with regard to collaboration, but in contrast to the objective results, no significant differences were found between both conditions. The results are summarized in Figure 4. One interesting finding is that participants stated that the help actually contributed to the level of collaboration, in case of textual help ($\bar{x}=6.2$, $s=2.4$) as well as animations ($\bar{x}=6.6$, $s=2.4$). When a certain task could be performed both collaboratively and individually (e.g. moving heavy pieces), participants reported that it was mostly accomplished together, with textual help ($\bar{x}=7.2$, $s=2.5$) and animations ($\bar{x}=7.4$, $s=2.0$).

Participants rated their amount of individual contribution a little higher in case of textual help ($\bar{x}=5.5$, $s=1.2$) compared to animations ($\bar{x}=4.3$, $s=1.6$), but overall they found that everyone contributed equally, both with textual help ($\bar{x}=7.8$, $s=1.5$) as well as animations ($\bar{x}=7.0$, $s=1.9$). Surprisingly, participants

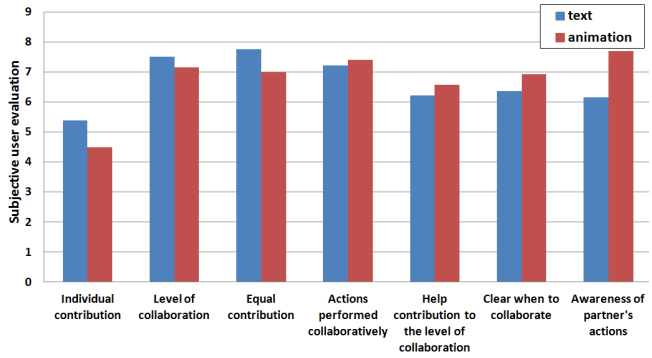


Figure 4. Means of the questionnaire results regarding the collaboration between two players.

experienced a very similar level of collaboration in the text ($\bar{x}$=7.5, s=1.3) and animation ($\bar{x}$=7.2, s=2.1) condition. The participants' subjective perception thus differs from our objective analysis, but this can be partially attributed to our broader definition of 'collaboration', as we also included actions such as watching animations together.

Finally, we analyzed the amount of communication. Participants communicated a lot with each other about various (not always task related) topics. We only found a slightly higher amount of communication in case of animated help ($\bar{x}$=37.3%, s=16.1) compared to text ($\bar{x}$=32.6%, s=13.3), but communication did start sooner in the animation condition ($\bar{x}$=13.6s, s=9.1) compared to the text condition ($\bar{x}$=67.3s, s=25.3), which is in line with the collaboration that also started sooner.

CONCLUSIONS

We presented the results of a study on two commonly used types of help, namely text and animations, within a collaborative tabletop game. Two users had to solve a puzzle by using a variety of gestures, some of which required the users to synchronize and coordinate their actions. We not only analyzed the subjective experiences of the users regarding the two help systems, but we also analyzed the effect of each help system on collaboration and communication.

The study shows that animated help allowed users to quickly discover the available interaction possibilities, with less of a negative impact on the game experience. Further studies are needed, however, to investigate if the nature of the gestures influences these results. Animations might be more suitable for very complex gestures, for instance, but not for simple gestures. Although subjectively both types of help contributed almost equally to the perceived level of collaboration, animated help had a positive effect on the actual level of collaboration, as users worked together to explore and learn the game. The high level of collaboration also reinforced awareness of each other's actions. In addition to the help, users learned a lot by talking and watching each other, which highlights the importance of social learning in a multi-user setting.

Our findings indicate that the choice of help system can make a difference, for instance to invite strangers to work together in a casual game on a public display. However, care has to be taken with generalizing these results, as more in-depth studies are needed to account for possible confounding factors with respect to interpreting the collaboration results. Furthermore, the study was conducted in a lab environment. In a real walk-up-and-use setting, users will be less committed to completing their task, and factors such as social embarrassment come into play.

We investigated two specific implementations of textual and animated help in a two-player game. With different group sizes, people develop different strategies, such as a group splitting in smaller subgroups, which in turn can influence how users deal with the help and how they support one another. Furthermore, each user had a separate help panel. Having one shared help panel in the center of the screen will result in different behaviors, and has the added benefit of being more flexible regarding the number of users, as the current approach is optimized for two users.

ACKNOWLEDGMENTS

We thank the participants of our study for their valuable time and effort.

REFERENCES

1. Anderson, F., and Bischof, W. F. Learning and performance with gesture guides. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI, ACM (2013), 1109–1118.
2. Bau, O., and Mackay, W. E. OctoPocus: a dynamic guide for learning gesture-based command sets. In *Proceedings of the 21st ACM symposium on User interface software and technology*, UIST, ACM (2008), 37–46.
3. Beznosyk, A. *An experimental perspective on factors influencing collaborative user experience in virtual environments and games*. PhD thesis, Hasselt University, 2012.
4. Bragdon, A., Uguray, A., Wigdor, D., Anagnostopoulos, S., Zeleznik, R., and Feman, R. Gesture Play: motivating online gesture learning with fun, positive reinforcement and physical metaphors. In *Proceedings of the ACM international conference on Interactive Tabletops and Surfaces*, ITS, ACM (2010), 39–48.
5. Norman, D. A. Gestural interfaces: a step backwards in usability. *Interactions 17*, 5 (2010), 46–49.
6. Vanacken, D. *Touch-based interaction and collaboration in walk-up-and-use and multi-user environments*. PhD thesis, Hasselt University, 2012.
7. Vanacken, D., Demeure, A., Luyten, K., and Coninx, K. Ghosts in the interface: meta-user interface visualizations as guides for multi-touch interaction. In *Proceedings of the 3rd IEEE international workshop on Horizontal Interactive Human Computer Systems*, TABLETOP, IEEE Computer Society (2008), 81–84.
8. Vogel, D., and Balakrishnan, R. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th ACM symposium on User interface software and technology*, UIST, ACM (2004), 137–146.

Towards Biomechanicly-Inspired Index of Expert Drawing Gestures Complexity

Myroslav Bachynskyi
Max Planck Institute for Informatics and
Saarland University

ABSTRACT

In this paper we describe our approach for measuring an in-air drawing gesture complexity based on biomechanical data. Our method is based on optical motion capture based biomechanical simulation and muscular synergy extraction. We assume that complexity of trajectory formation by our brain can be calculated in inverse way: we observe a movement trajectory and a posture, then we interpret the movement in terms of muscular forces needed to produce it and corresponding muscle activations, further we find similar muscle co-activation patterns which correspond to synergies of muscles which work together in response to single activation signal sent by our brain. Our claim is that number of distinct synergies necessary to produce the gesture reflects the complexity of the movement. Thus trajectory which needs more synergies is more complex and is harder to remember and repeat. The main goal of this paper is to introduce objective measure of complexity for in-air gestures which can be used for assessment of gestural interfaces.

Author Keywords

Biomechanical simulation; muscle activations; synergies; gestural interfaces, gesture complexity.

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces: Evaluation/methodology

INTRODUCTION

Recent development of accelerometers, computer vision sensors, and body tracking technology allows wider implementation of gestural interfaces for computer control. These interfaces are advantageous for multiple scenarios where many degrees of freedom need to be controlled, distant application control or sterile environment are required. Examples cover multiple cases from games and exergames, public displays, 3D modeling, industrial applications or medicine. Different use cases pose specific and critical requirements to the gestures, which makes gestural design a hard problem. Designers of gestural interfaces need a huge amount of knowledge external to what they know and use with traditional desktop-based interfaces.

Goal of gestural interfaces is to provide for users natural and easy interaction, nevertheless many of them are designed without proper consideration of physical and cognitive ergonomics, as well as human performance. The reason for this is the fact, that gestural design space is much larger than design space for desktop-based interaction, as result it is very hard to cover all important factors and consider all possible alternatives. Empirical analysis of such huge design space is expensive and time consuming. An alternative is usage of conceptual models of human performance, however they are underdeveloped for gestures. In desktop-based interfaces most widely used model is Fitts' law, which describes movement time and difficulty of aimed movement. In contrast design of gestures lacks similar working model which would describe difficulty or complexity of gesture.

Of course there were some earlier works dealing with gestures: adaptations of Fitts' law, applications of 2/3 power law or finding invariance in movements, for example in end-effector velocity profile [3]. But they analyzed only trajectory of end-effector, which is too narrow for in-air gestures, where posture of whole arm plays significant role and differs depending on orientation or location of a movement. Furthermore, they haven't defined any measure of gesture complexity. Gesture complexity measure should describe how difficult it is to learn, produce and repeat the gesture.

In this paper we want to improve theoretical knowledge about gestures by considering them from physiological viewpoint. We can take into account not only an end-effector data as in earlier studies, but also the data about posture of a limb and corresponding indices from inside human body. We record body posture during gesture using optical motion capture. Then we apply biomechanical simulation to extract data about the movement from inside of our body, including activations of all muscles recruited during the gesture. Based on these muscle activations we identify all muscle synergies recruited during the movement.

Our hypothesis is that *complexity of gesture is correlated with number of distinct synergies recruited during the movement produced by experienced person.* If more synergies are recruited, than gesture is more complex and harder to remember and reproduce.

If gesture complexity measure is defined, it could be applied by designers to create better interfaces, with each gesture

Submitted to the workshop "Gesture-based Interaction Design: Communication and Cognition"

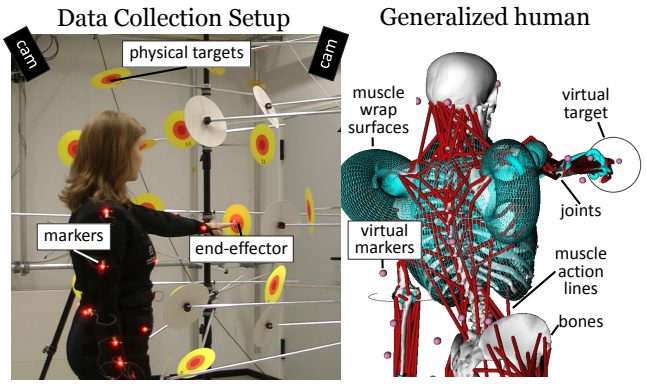


Figure 1. ***Left:*** **subject in motion capture suit with markers and target setup.** ***Right:*** **Inverse Kinematics with biomechanical model scaled to proportions of subject.**

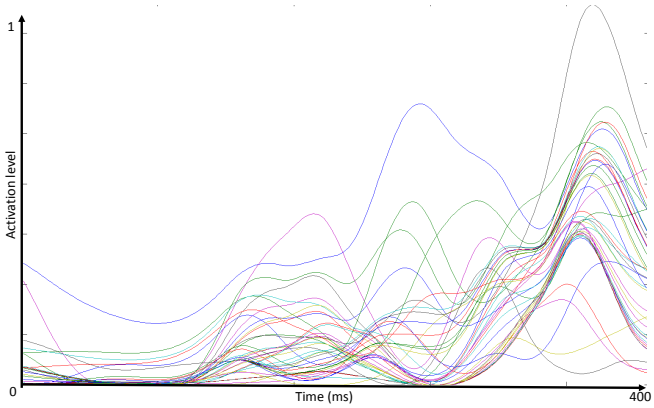


Figure 2. Activations of the upper extremity muscles plotted over time for single aimed movement. Each line corresponds to one muscle.

fine-tuned to the computer action it induces. As example, too complex gestures can be avoided in most cases, or they could be intentionally introduced for some critical actions like "permanent delete". This will improve usability and ergonomics of gestural interface and make learning curve less steep.

In this paper we describe our approach and possible complexity measurement steps. We start from description of optical motion capture and biomechanical simulation, further we write about muscle synergies and their extraction, based on them we describe our hypothesis with additional details and example of synergies for movements in 3D space, we conclude with discussion of limitations of our assumption and future work.

There are still issues with our approach concerning validation of gesture complexity measure. We also want to try it on real complex gesture and demonstrate practical HCI application of the complexity measure.

BIOMECHANICAL SIMULATION: FROM MOVEMENT OBSERVATION TO ACTIVATIONS OF ALL MUSCLES

MoCap-based biomecchanical simulation is a method for estimation of processes inside human body based on external observations. It is combination of two techniques: motion capture and biomechanical simulation of musculoskeletal models.

Today the human motion can be recorded in multiple ways: using set of goniometers attached to limbs, optical recording of locations of active or passive markers attached to human body, estimating the posture of human body based on recordings of multiple cameras and statistical models of the human, using accelerometer sensors and gyroscopes, etc. The most accurate and fairly non-intrusive method is marker-based optical motion capture, which is applied to collect data for further biomechanical simulation. Modern marker-based systems allow recordings accurate to the millimeter level, for example the system we use, PhaseSpace Impulse with 12 cameras and 43 active markers attached to the suit (Figure 1 *left*) records position of markers with 1/5 mm accuracy and with up to 480 fps speed.

Biomechanical simulation processes motion capture data and augments it with number of indices from inside the body. There are few software systems which can run biomechanical simulation: OpenSim, LifeModeler, AnyBody, SIMM. Some of them come together with musculoskeletal model of the human body. We use free opensource tool OpenSim [2] for running biomechanical simulations. The model we have selected is commercial SIMM Full-Body musculoskeletal model shown on Figure 1(*right*) [4]. The simulation pipeline in OpenSim consists from multiple steps which update the model and transform the data from one form to the other. All relevant steps are described below:

- *Scaling* updates generalized musculoskeletal model to the proportions of particular subject. All body segments can be scaled separately in all dimensions, model mass is scaled, although generalized mass distribution is preserved.
- *Marker adjustment* corrects placements of virtual markers to match recorded ones according to the value of trust specified for each marker.
- *Inverse Kinematics* calculates generalized coordinates (angles at joints) based on positions of markers in 3D space by minimizing sum of squared errors between corresponding virtual and real markers.
- *Inverse Dynamics* computes moments at joints based on outputs of inverse kinematics, mass distribution of the model and available external forces.
- *Static Optimization* resolves total moments at joints to the forces as well as activations of separate muscles. This step assumes that for a movement humans recruit their muscles in optimal way. This matches the recruitment of muscles of the person which has already learned particular movement.

Among the simulation outputs the most interesting data with respect to gesture complexity is muscle activations necessary to produce the trajectory. Their calculations were recently validated against EMG recordings for aimed movements in all directions covering whole reachable space of upper extremity [1]. Example of muscle activations for single aimed movement is demonstrated on Figure 2.

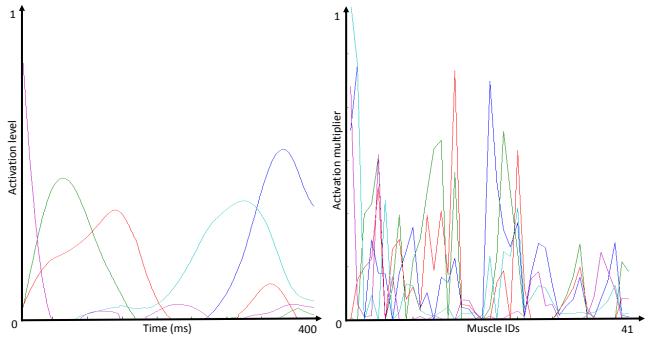


Figure 3. ***Left:*** **activations of 5 synergies,** ***Right:*** **multipliers for each muscle of 5 synergies.**

MUSCLE SYNERGIES AS DEFINING FACTOR OF MOVEMENT

Although muscle activations present huge amount of information about the movement, there is additional concept between them and our brain: muscle synergies. This theory proposed by Bernstein as solution to degrees-of-freedom problem (How our brain can control with high efficiency synchronously and precisely thousands of motor units and hundreds of muscles in whole our body?) According to the theory, for learned movement the motor units from different muscles are combined in groups which are simultaneously co-activated with the same signal patterns sent by our brain [6]. Within this theory control of movement is executed as follows:

- Motor cortex sends single signal through the neural system.
- Motor center in spinal cord receives signal from motor cortex, and forwards it to multiple motor units with particular multiplier for every motor unit.
- Each motor unit receives signal from motor center and activates all own muscle fibers according to the signal.
- Activated muscle fibers contract producing active forces, which together accelerate our limb towards target posture.

Within such scheme it is much simpler for our brain to efficiently control variety of learned movements which can be produced. However, if it is necessary to produce some completely new movement (never produced any similar movement before), during learning stage humans recruit synergies existing from previously learned movements or control separate muscles, then after some number of repetitions for separately controlled muscles new synergy is established.

For identification of muscle synergies from activations of all muscles multiple matrix factorization algorithms were adopted and tested [5]. We use Non-Negative Matrix Factorization as it produces plausible results and fulfills natural constraints of non-negativity of neural signals. On Figure 3 is shown an example of synergies and their activations extracted using above mentioned algorithm for whole arm aimed movement. The extracted 5 synergies reconstruct $>85\%$ of variability of the source muscle activations.

In order to produce complex movement the sequence of activation signals is sent to the set of synergies. Our hypothesis is that *number of synergies recruited within the gesture is correlated with the complexity of the movement.* Thus movements which require more distinct synergies have higher complexity, and are harder to learn and repeat.

GESTURE COMPLEXITY

Our hypothesis describes gesture complexity as measure correlated with number of distinct synergies activated within the movement.

We assume that complexity can not be dependent on separate muscles: according to degrees of freedom problem stated by Bernstein, which is base for muscle synergies theory, for learned movements human brain does not control each muscle separately. From other viewpoint gesture can not be stored in our brain as trajectory of end-effector only, because it will need a lot of computations during motor planing phase, which would exclude high performance of movement.

To learn a gesture our brain should remember synergies activation sequence. When it is necessary to produce a gesture motor cortex retrieves stored pattern, slightly updates it to the environmental situation and applies it. We consider only distinct synergies, because if one synergy is applied repetitively, than its signal can be simplified by separating repeating component from repetition pattern, however this assumption is not very strong and it can turn out that all synergy recruitments correlate better with gesture complexity than distinct synergy recruitments.

CASE: SYNERGIES OF AIMED MOVEMENTS

We have tried our hypothesis on the data we have collected in pointing experiment in the 3D space reachable by the arm. We have uniformly distributed 25 targets within reachable space (Figure 1 *left*). In the experiment we have recorded aimed movements between each pair of targets. Figure 4(*left*) demonstrates visualization of recorded end-effector trajectories in half-sphere reachable by the arm. As discussed in the sections above we have recorded optical motion capture data and run all steps of biomechanical simulation for all recorded movements. Then we have extracted 10 synergies which reconstruct 85% of variability in muscle activation data of whole dataset. Figure 4(*center and right*) demonstrates segments of trajectories on which two example synergies (Synergy 1 or Synergy 2) were activated.

It can be clearly seen that on corresponding segments Synergy 1 mostly accelerates limb in lower and central part of space in the direction *Up*, and is applied in some cases for corrective movements close to the targets in upper part of space.

Synergy 2 mostly accelerates limb in the direction to *Right Middle* as well as participates in corrections close to the targets in central and right part of space.

Other identified synergies also have specific directionality and part of space where they are active. These properties support the validity of identified synergies. Additionally, activation patterns of aimed movements match prevoius knowledge. Short and middle length aimed movements recruit in

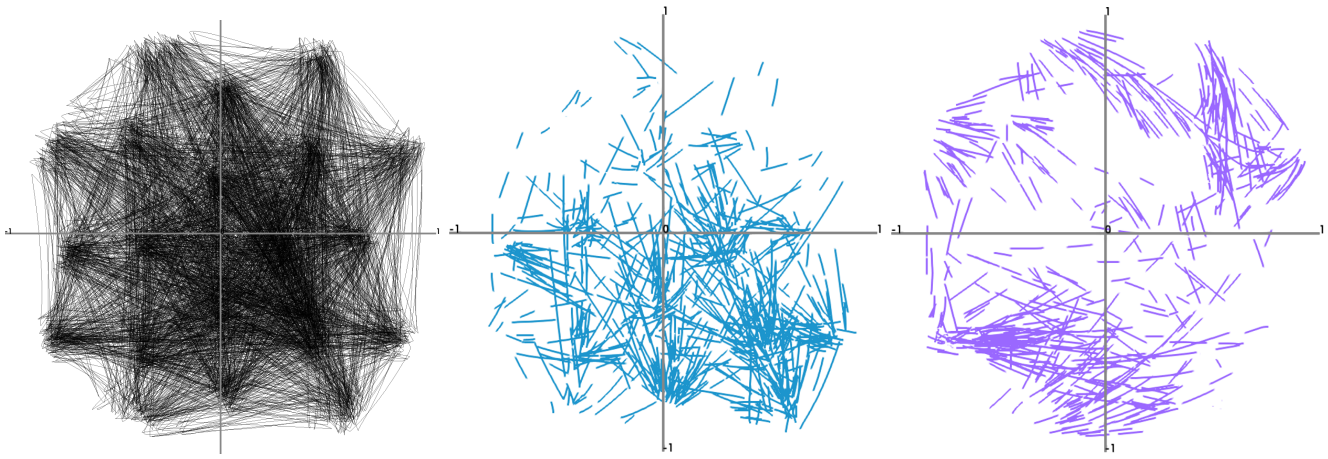


Figure 4. ***Left:*** **Complete trajectories of aimed movements in reachable 3D space.** ***Center and Right:*** **Segments of movements produced by the single synergy.** ***Center:*** **Synergy 1,** ***Right:*** **Synergy 2**

most cases two synergies: agonist synergy accelerates end-effector towards target, antagonist synergy decelerates end-effector at the target. In contrast long movements between opposite ends of the movement space recruit 3 and more synergies: besides agonist at the movement begin and antagonist at the movement end, there are additional synergies in the middle phase of movement. According to our hypothesis, this means that such movements have higher complexity, which is supported by the fact that their velocity profile deviates from bell-shaped.

DISCUSSION

In the paper we propose new objective measure of complexity of gestures based on biomechanical and neural processes inside human body. Such measure of complexity could be used to create more effective, easier to learn and use gestural interfaces. Gesture designers will be able to compare alternative gestures, search for optimal solution in iterative way, or even analytically search for optimal gestures if synergies from possible input space will be available as primitives.

Additionally proposed complexity measure could provide not only information about cognitive complexity, but also information about physical difficulty of movements. It is based on biomechanical data, which can be extensively used for physical ergonomics evaluation of interfaces. Many intermediate results as joint angles, joint moments, muscle forces and muscle activations can be considered by gesture designers to reduce physical ergonomics cost by avoiding extreme values of those measures and keeping integrated value of muscle forces and activations as low as possible.

There are still some challenges that have to be solved. At first we need to check validity of the gesture complexity predictions. Second is to validate between-subject reproducibility of corresponding synergies and complexity measure. This can be done by conducting experiments with variable complexity of gestures and checking learnability and reproducibility of them.

Than we need to create “recipe” for application of complexity measure and check its usability and designers gain-in knowledge on some experimental task. Then application of the complexity measure on real design task need to be checked and its influence on gestural design process.

There are also some technical difficulties with application of biomechanical simulation. First is necessity of precise marker placement, although in near future it can be solved by markerless motion capture. Second is still high computational cost of static optimization in OpenSim, however there is already published work claiming to have solved this problem.

REFERENCES

1. Bachynskyi, M., Oulasvirta, A., Palmas, G., and Weinkauf, T. Is motion-capture-based biomechanical simulation valid for hci studies? study and implications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press (2014).
2. Delp, S. L., Anderson, F. C., Arnold, A. S., Loan, P., Habib, A., et al. OpenSim: Open-source software to create and analyze dynamic simulations of movement. *IEEE Trans. Biomedical Engineering 54*, 11 (2007), 1940–1950.
3. Gibet, S., Kamp, J.-F., and Poirier, F. Gesture analysis: Invariant laws in movement. In *Gesture-based communication in human-computer interaction*. Springer, 2004, 1–9.
4. Holzbaur, K. R. S., Murray, W. M., and Delp, S. L. A model of the upper extremity for simulating musculoskeletal surgery and analyzing neuromuscular control. *Ann. of Biomed. Eng. 33* (2005), 829–840.
5. Tresch, M. C., Cheung, V. C. K., and d’Avella, A. Matrix factorization algorithms for the identification of muscle synergies: Evaluation on simulated and experimental data sets. *J. Neurophysiol. 95* (2006), 2199–2212.
6. Tresch, M. C., and Jarc, A. The case for and against muscle synergies. *Current Opinion in Neurobiology 19*, 6 (2009), 601 – 607. Motor systems Neurology of behaviour.

How Do Users Interact with an Error-prone In-air Gesture Recognizer?

Ahmed Sabbir Arif[1], Wolfgang Stuerzlinger[1], Euclides Jose de Mendonca Filho[2], Alec Gordynski[3]

[1]York University
Toronto, Ontario, Canada
{asarif, wolfgang}@cse.yorku.ca

[2]Federal University of Bahia
Salvador, Bahia, Brazil
euclidesmendonca.f@gmail.com

[3]Flowton Technologies
Toronto, Ontario, Canada
ag@flowton.ca

ABSTRACT
We present results of two pilot studies that investigated human error behaviours with an error prone in-air gesture recognizer. During the studies, users performed a small set of simple in-air gestures. In the first study, these gestures were abstract. The second study associated concrete tasks with each gesture. Interestingly, the error patterns observed in the two studies were substantially different.

Author Keywords
Error behaviours; in-air gestures; gestures.

ACM Classification Keywords
H.5.m User Interfaces: Miscellaneous.

INTRODUCTION
Invisible user interfaces are becoming increasingly popular. It refers to an interface that is either invisible or becomes invisible with successive learned interactions. Users interact with such interfaces mainly via gestures, potentially in the air [11]. Smart televisions, video game consoles, or the Leap Motion enable such interactions. Current gesture recognizers achieve up to 99% accuracy [1], provided that sufficient training data is available and reliable input technologies are used. However, in practice, gesture-based techniques are more error prone than traditional ones, likely due to gesture ambiguity [8] and the lack of appropriate feedback [5]. As the number of easily *performable* gestures is limited, most techniques utilize the same or similar gestures for multiple tasks [8]. This makes it difficult for users to associate gestures to concrete tasks and for the system to disambiguate them. The absence of direct visual feedback for in-air gestures poses additional problems [11].

A well-regarded theory in psychology error research, called the mismatch concept [3], holds both the user and system responsible for committing errors, but attributes errors to the mismatch in the interaction among these two. This implies that a deeper insight into how users interact with error prone systems is crucial for user-friendly interfaces and effective recognition techniques. For instance, if a large number of users experience a given mistake, it is prudent to globally change that specific system feature. Similarly, an adaptive system could calibrate itself to individual user behaviours. Yet, error behaviours for gesture interfaces have not been well studied. To inform in-air gesture research, we conducted two pilot studies to explore error behaviours.

Copyright is held by the owner/author(s).

CHI 2014, Apr 26 - May 01 2014, Toronto, ON, Canada
Workshop on Gesture-based Interaction Design: Communication and Cognition

INVESTIGATED GESTURES
Five simple in-air gestures, *push*, *left*, *right*, *up*, and *down*, were used in the studies. These gestures were performed by placing the dominant hand in the *rest* position (resembling a stop or wait gesture), moving the hand *forward*, *left*, *right*, *up*, or *down*, respectively, and then bringing it back to the rest position. Illustrated in Figure 1. The rest position was used as the origin and was updated during each *push* gesture.

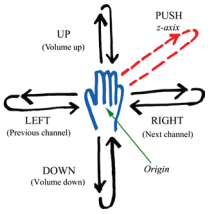


Figure 1. The five gestures used during the pilot studies.

Performance Metrics
The following metrics were calculated during the pilots.

Gesture Attempts (GA): The average number of attempts it took to perform a gesture. A flawless gesture recognizer will result in a GA of one, provided there was no human error.

Error Rate (ER): This is the average percentage of errors committed with the system. This is a compound of the *Human Error Rate* (ER_H) and the *System Error Rate* (ER_S). The first is the average committed by the users, with the second committed by the system. The experimenter manually recorded all user actions and system reactions. Incidents where users performed the correct gesture, but the system failed to (correctly) recognize it, were classified as system errors. If users performed the wrong gesture, this was seen as a human error. The experimenter also kept a watchful eye for incidents where both the human and the system made mistakes. However, such incidents did not occur during the pilots.

PILOT STUDY 1
This study investigated abstract gestures, i.e. gestures that were not associated with concrete tasks.

Apparatus and Participants
A custom application, developed with the OpenNI [10], was used during the pilot. It sensed motion using a Microsoft

Kinect and recognized gestures using the NiTE™ computer vision library [9].

Fourteen participants, aged from 20 to 35 years, average 23.6, voluntarily participated in this pilot. All were right-handed. Seven were male. Three frequently interacted via in-air gestures with Kinect, Wii, and/or PS3 controllers, nine occasionally, and the rest had no prior experience.

Figure 2. Pilot study 1 setup. The Kinect was placed on a table 20″ above the floor. The participant sat approximately 40″ away from the Kinect facing the device and inputted gestures as instructed by the experimenter (bottom right). Here, the participant is holding his hand in the initial *rest* position. The experimenter manually recorded the inputted gestures, failed attempts, and the types of errors.

Procedure and Design

Participants were instructed to place their hands in the initial *rest* position and to wait for the experimenter's instructions. The experimenter then guided them through the study by instructing them on the gestures to perform. The experimenter also verbally informed them of any mistakes made by them or by the system. In other words, the experimenter simulated auditory feedback for the system. Figure 2 illustrates the study setup. If an error was made, participants were requested to try the same gesture again until it was successfully recognized by the system. If no error was made, they were asked to input the next gesture. The experimenter manually recorded the type and numbers of attempted gestures, successful or unsuccessful attempts, and the types of errors committed. Upon completion of the study, participants were interviewed on the perceived easiness of the gestures, fatigue, and general comments. We uses a within-subjects design: 14 participants × 4 blocks × 4 trials × 6 gestures per trial (*push*, *left*, *right*, *up*, *down*, *push*) = 1,344 gestures in total. The four inner gestures were counterbalanced in each block via a balanced Latin square. The first and the last *push* gesture activated and deactivated the gesture recognition system, correspondingly. There was no practice, but the experimenter demonstrated how to perform the gestures before the study.

Results

An Anderson-Darling test revealed that the study data was normally distributed. A Mauchly's test confirmed that the data's covariance matrix was circular in form. Therefore, we used repeated-measure ANOVA for all analysis.

Gesture Attempts (GA)

An ANOVA failed to identify a significant effect of gesture type on GA ($F_{4,13}$ = 0.13, ns). There was also no significant effect of block ($F_{3,13}$ =1.62, p > .05) or gesture type × block ($F_{12,156}$ = 1.08, p > .05). On average, the overall GA during the four blocks were 1.08 (SE = 0.02), 1.09 (SE = 0.02), 1.05 (SE = 0.02), and 1.03 (SE = 0.01), respectively. Figure 3 illustrates the average GA for each gesture.

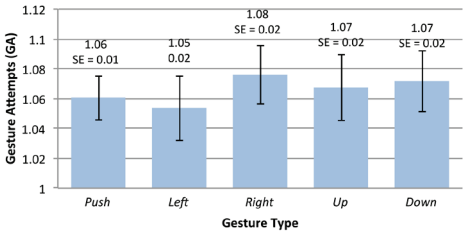


Figure 3. Average gesture attempts (GA) for the five gestures. The error bars represent ±1 standard error (SE).

Error Rate (ER)

An ANOVA failed to identify a significant effect of gesture type on ER ($F_{4,13}$ = 0.49, ns). There was also no significant effect of block ($F_{3,13}$ = 1.62, p > .05) or gesture type × block ($F_{12,156}$ =1.01, p > .05). On average, overall ER during the four blocks were 7.44 (SE = 2.2), 8.33 (SE = 2.38), 5.06 (SE = 1.53), and 3.27% (SE = 1.05), respectively. Figure 4 illustrates the average ER for each gesture.

Experimenter records revealed that 80.3% of all errors were committed by the system. The remaining 19.7% were human errors. A Chi-squared test found this to be statistically significant (χ^2 = 36.00, df = 1, p < .0001).

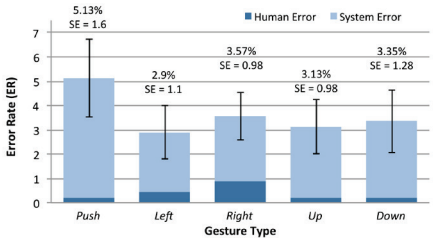


Figure 4. Average error rate (ER) for the five gestures. The error bars represent ±1 standard error (SE).

An ANOVA failed to identify a significant effect of gesture type on ER_H ($F_{4,13}$ = 1.07, p > .05). There was also no significant effect of block ($F_{3,13}$ = 1.42, p > .05) or gesture type × block ($F_{12,156}$ = 1.05, p >.05). Likewise, an ANOVA failed to identify a significant effect of gesture type on ER_S ($F_{4,13}$ = 0.62, ns). There was no significant effect of block ($F_{3,13}$ = 2.28, p > .05) or gesture type × block ($F_{12,156}$ = 0.81, ns).

Human Effort and Fatigue

No straightforward trend was observed regarding this. About half the users reported high post-study fatigue, while the rest were mostly neutral. A Chi-squared test did not find this to be statistically insignificant (χ^2 = 0.999, df = 2, ns).

Discussion

The system recognized 94% of all gestures correctly. As at least 97% accuracy rate is necessary for the users to find a gesture-based system useful [7], our setup served well as an *error prone system*. The results showed that gesture type has no significant effect on attempts per gesture or accuracy. This means none of the gestures were substantially more error prone than the others. One potential explanation is that the gestures were abstract and users did not have to

associate tasks with each one of them. There was also no significant effect of block. This is not unusual considering the length of the study. More interestingly, we identified the following behaviours by observation and user responses.

- About 57% users found the *down* gesture uncomfortable to perform. One potential reason is that they were performing this gesture in a seated position, which did not give them enough space to freely move their hands (far enough) downwards. Other gestures were rated mostly neutral.
- About half the users made relatively shorter gestures by the second block, i.e. they kept their hands closer to their body compared to the first block.
- Users often briefly got confused between *left* and *right*. That is, they started performing the opposite gesture, i.e. *left* instead of *right* or vice versa, but corrected themselves almost immediately. The experimenter did not record such incidents as errors.

PILOT STUDY 2

This pilot study investigated task-associated gestures, i.e. gestures that were associated with concrete tasks.

Apparatus and Participants

A custom application, developed with OpenNI [10], was used during this study. It sensed motion using a Microsoft Kinect and recognized gestures using the NiTE™ computer vision library [9]. The application was shown full-screen on a 21″ CRT monitor, to mimic a television screen. It displayed the current channel number in digits at the centre, and current volume level in a slider at the bottom of the screen. The monitor also served as a stand for the Kinect. The application logged all interactions with timestamps. The experimenter used a separate application to present random gestures in each session and to record user behaviours and errors. It was launched on a laptop computer to facilitate fast logging via keyboard shortcuts, see Figure 5.

Seven participants, aged from 21 to 43 years, average 29.0, voluntarily participated in this pilot. They all were right-handed and two were female. One frequently interacted via in-air gestures with a Nintendo Wii or Leap Motion and two occasionally, while the rest had no prior experiences. They all received a small compensation.

Procedure and Design

Participants were instructed to place their hands in the initial *rest* position and to make a *push* gesture when they were ready. This started a session and presented a random task on the top of the screen for them to perform. Participants performed four tasks: load the previous channel (*left)*, load the next channel (*right)*, raise the volume (*up)*, and lower the volume (*down)*, see Figure 1. Error correction was forced. That is, participants had to keep trying until the system performed the intended task. They were provided with visual feedback—they could see the channel changing and the volume bar moving upon successful recognition of the corresponding gestures. Upon completion of the study, they were asked to fill out a short questionnaire. A within-subjects design was used: 7 participants × 3 sessions × 64 gestures (*left*, *right*, *up*, *down*, each 16 times, *randomized*) = 1,344 gestures in total. There was no practice block, but the experimenter demonstrated how to perform the tasks prior to the pilot. There was a mandatory 5 min. break between sessions.

Figure 5. Pilot study 2 setup. Participants sat on the left chair facing the monitor and the Kinect. The experimenter sat on the right chair with a clear view of the setup and the participant. The chairs were approximately 18″ high. The Kinect was placed above the monitor in 4° angle and approximately 46″ above the floor. The distance between the participants and the Kinect was kept at approximately 48″. The experimenter used an extended keyboard, not visible here, to log user behaviours.

Results

An Anderson-Darling test revealed that the study data was normally distributed. A Mauchly's test confirmed that the data's covariance matrix was circular in form. Therefore, repeated-measure ANOVA was used for all analysis.

Gesture Attempts (GA)

An ANOVA identified a significant effect of gesture type on GA ($F_{3,6} = 5.7$, $p < 0.01$). A Tukey-Kramer test revealed that *left* took significantly more attempts than *up* and *down*. However, no significant effect of session ($F_{2,6} = 0.47$, ns) or gesture type × session ($F_{6,36} = 0.37$, ns) was identified. On average, the overall GA during the three sessions were 1.07 (SE = 0.02), 1.05 (SE = 0.01), and 1.1 (SE = 0.02), respectively. Figure 6 illustrates the average GA for each gesture.

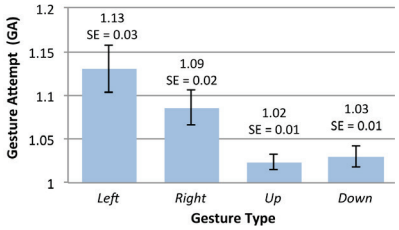


Figure 6. Average gesture attempts (GA) for the four gestures. The error bars represent ±1 standard error (SE).

Error Rate (ER)

An ANOVA identified a significant effect of gesture type on ER ($F_{3,6} = 5.87$, $p < .01$). A Tukey-Kramer test revealed that *left* suffered from significantly more errors than *up* or *down*. Yet, no significant effect of session ($F_{2,6} = 0.48$, ns) or gesture type × session ($F_{6,36} = 0.41$, ns) was found. The overall average ER during the three sessions were 6.70 (SE = 1.91), 4.69 (SE = 1.0), and 8.93% (SE = 4.30). Figure 7 illustrates the average ER for each gesture.

Experimenter records revealed that 91.2% of all errors were committed by the system and the remaining 8.8% by humans. A Chi-squared test found this to be statistically significant ($\chi^2 = 67.24$, $df = 1$, $p < .0001$).

An ANOVA found no significant effect of gesture type on ER_H ($F_{3,6} = 0.26$, ns). There was also no significant effect of session ($F_{2,6} = 0.18$, ns) or gesture type × session ($F_{6,36} = 1.25$, $p > .05$). Yet, an ANOVA identified a significant effect of gesture type on ER_S ($F_{3,6} = 6.48$, $p < .005$). A Tukey-Kramer test revealed that *left* had significantly more system errors than *up* and *down*. There was no significant effect of session ($F_{2,6} = 0.37$, ns) or gesture type × session ($F_{6,36} = 0.36$, ns).

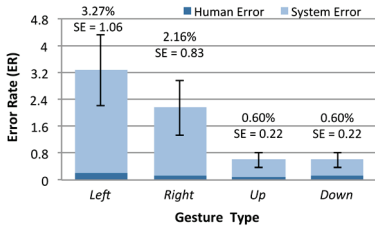


Figure 7. Average error rate (ER) for the four gestures. The error bars represent ±1 standard error (SE).

Human Effort

A Friedman test failed to identify significance with respect to mental ($\chi^2 = 2.467$, $df = 3$, ns), physical ($\chi^2 = 0.864$, $df = 3$, ns), or temporal demand ($\chi^2 = 1.429$, $df = 3$, ns), performance ($\chi^2 = 6.414$, $df = 3$, ns), effort ($X^2_{(3)} = 1.333$, $df = 3$, ns), or frustration ($\chi^2 = 4.761$, $df = 3$, ns) for the four gestures. Also, no straightforward trend was found regarding fatigue. About 43% users reported that they experienced high post-study fatigue, while the rest were mostly neutral.

DISCUSSION

The system recognized about 93% of all gestures, which is comparable to the first pilot. However, unlike the first pilot, significant effects of gesture type were observed for both gesture attempts and accuracy. This indicates the possibility that error patterns may be different for meaningless and meaningful gestures. One explanation is that recall-based tasks are more challenging. Remarkably, the *left* gesture was significantly more error prone and took more attempts than the other ones. The reason is that users often performed the *right* or the *down* gestures instead of *left*. While we do not have a definite reason for this behaviour, the fact that all our participants were right-handed may have contributed to this phenomenon. Similar to the first pilot, there was no significant effect of session. User behaviours towards the system were also comparable to the first pilot. But this time we collected more data to further analyse the behaviours.

- About 71% users believed that their performance got faster with time. The results do not support this. An ANOVA failed to identify a significant effect of session on task completion time ($F_{2,6} = 1.7$, $p > .05$). It is possible that users' gesture recall time decreased with practice. However, this is difficult to verify in a short study as the difference between the novice and the expert recall and preparation time is only ~600ms [6].
- Users easily got impatient with system errors and made repetitive attempts without allowing the system to react to the first re-attempt. This caused additional system errors (which we recorded correspondingly). About 60% errors took more than one attempt to fix. Compared to behaviours in a different error prone system [2], a Chi-squared test found this to be statistically significant ($\chi^2 = 4.0$, $df = 1$, $p < 0.05$). During the interviews, all users stated that their reaction was instinctive.
- About 43% users experienced high post-study fatigue. Observation revealed that these users continued making longer gestures (moved their hands further away from their body), while the others started making shorter gestures by the second session.
- Users often got confused between the *left* and the *right* gestures. About 82% of the total errors were committed while performing these two gestures.

CONCLUSION AND FUTURE WORK

We presented results of two pilot studies that investigated error behaviours with an error prone in-air gesture recognizer. In the first study users performed abstract gestures, while in the second they performed task-associated gestures. Results show that although error patterns during the two studies were substantially different, users reactions to the errors were similar.

The first pilot instructed users verbally to perform a task and provided them with auditory feedback, while the second pilot used a custom application and provided visual feedback. In the future, we will investigate the effect of these factors and of different user expertise on the study results.

REFERENCES

1. Anthony, L. and Wobbrock, J.O. $N-Protractor: A fast and accurate multistroke recognizer. *GI 2012*, 117-120.
2. Arif, A. S. and Stuerzlinger, W. Predicting the cost of error correction in character-based text entry technologies. *CHI 2010*, 5-14.
3. Brodbeck, F. C., Zapf, D., Prumper, J., Frese, and M. Error handling in office work with computers: A field study. *J. Occup. Organ. Psychol. 66*, 4 (1993), 303-317.
4. Forlines, C., Wigdor, D., Shen, C., and Balakrishnan, R. Direct-touch vs. mouse input for tabletop displays. *CHI 2007*, 647-656.
5. Gustafson, S., Bierwirth, D., and Baudisch, P. Imaginary interfaces: Spatial interaction with empty hands and without visual feedback. *UIST 2010*, 3-12.
6. Kieras, D. E. Using the keystroke-level model to estimate execution times. University of Michigan, MI, USA, 1993.
7. LaLomia, M. User acceptance of handwritten recognition accuracy. *CHI 1994*, 107-108.
8. Morris, M. R., Wobbrock, J. O. and Wilson, A. D. Understanding users' preferences for surface gestures. *GI 2010*, 261-268.
9. NiTE™ Middleware. http://www.openni.org/files/nite
10. OpenNI Framework. http://www.openni.org
11. Wigdor, D. and Wixon, D. *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. Morgan Kaufmann, Burlington, MA, USA, 2011.

Challenges in Gesture Recognition for Authentication Systems

Gradeigh Clark and Janne Lindqvist
Rutgers University
gradeigh.clark@rutgers.edu and janne@winlab.rutgers.edu

ABSTRACT
In this paper we describe current popular gesture recognition methods that can be applied to gestures on mobile devices for authentication. Three different methods are considered : geometric recognizers, Hidden Markov Models, and Dynamic Time Warping. A brief description of each method is given along with a series of design considerations that we believe should be kept in mind when trying to develop recognizers for authentication.

Author Keywords
Gesture recognition; authentication systems

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces

INTRODUCTION
The proliferation of sensor-loaded interactive mobile devices like tablets, smartphones, and touchscreen laptops have allowed the generation of a variety of different gesture types that can be interpreted by these devices to perform any number of tasks. Broadly speaking, these tasks can be as diverse as tilting a screen for visual effect in a game or for changing screen resolution via pinching motions. There is prior work on defining gesture sets to accomplish some of these tasks that explore the applicability of both user-generated gesture sets and pre-defined gesture sets [5, 3, 9]. There do exist gesture sets have succeeded as standards from a 2D perspective: pinching for zoom, swiping to move a screen in a given direction, etc. In current smartphone designs, phones react to 3D gesture motions such as picking up the phone and moving it towards the ear to answer a call.

Gestures represent large potential as an authentication schema for smartphones, tablets, or any other sensor-rich device available in the marketplace. Gesture based methods have advantages over current popular authentication methods for devices (e.g. text entry or personal identification numbers) given that these gestures can potentially be done faster, require less concentration and lower accuracy while also being customizable, easier to remember and benefit from increased security. The most ubiquitous form of gesture authentication is the familiar nine pin password seen on Android devices.

Herein, we present on a number of popular recognition schemes that can be applied to mobile devices. We will narrow our focus to recognition techniques for 2D gestures. 2D gestures are gestures that use the touch screen for drawing a gesture (e.g. tracing a circle on the screen) and we will presume no restriction on the number of fingers that can interact with the screen. We examine three types: geometric methods [2, 1, 8, 6, 13], Dynamic Time Warping, and Hidden Markov Models [4, 7, 12]. These three appear to be the more prevalent types of recognizers used for the given gesture type, although that is not to say they are the only ones available for us; there are also Bayesian Networks, Artificial Neural Networks, Support Vector Machines, et cetera.

DESIGN CONSIDERATIONS
When considering gestures for authentication, we need to think about how a system should be configured to manage inputs. Drawing inspiration from Wobbrock [13] and others, we can make a list of design considerations for recognition with regards to authentication when designing a recognizer:

1. Sample invariant. Gestures that are recorded across different devices with their own configurations on how touch events are sampled along with the natural variances in the speed and time with which a user performs a gesture can lead to inputs that, can be spatially correct but mismatched with regards to vector size. The recognizer should resample the input such that it obtains an accurate portrait of the input data and normalize the size at the same time.
2. Trainable. A good recognizer should allow for the design and learning of the system to handle new inputs. It should not run only off a predefined set of gestures from the designer; users need to be allowed to design their own sets of gestures otherwise it is not possible to leverage the full utility of the password space afforded to gestural inputs on a mobile device.
3. Computationally feasible. When designing for mobile platforms, it is important to reduce the overall CPU and time consumption of the algorithm performing recognition. The user experience for authentication will be degraded if there is a delay for every sign in attempt, especially given how many times per day an average person now checks their smartphone.
4. Configurable. A gesture recognizer should allow for many different options for both the user and the developer. This could be small things like control over the sampling rate or how many stored templates used.
5. Resistant. A recognizer should be capable of of rejecting false users. To achieve this end, it may be more important

to leverage more biometric features of a gesture when performing the recognition. This could be anything from the finger length (as used to effect by Sae-Bae et al. [9]) to the pressure of the finger. Combining this feature with configurability would make a much more potent recognizer.

6. Ease of use. Although recognizers can be described in papers and pseudocoded, it may not be enough to communicate its effectiveness to developers. A difficult, mathematically intensive and hard to understand recognizer can cause issue with adoption especially so if the recognizer happens to be quite good. Most developers that may want to use it, say college students or startups, would have trouble picking the recognizer up if it is beyond their comprehension. So it is important to make a recognizer that is also clean, neat, and API friendly while trying – as much as possible – to reduce algorithmic complexity and bring the knowledge base a little closer to Earth.

There are also other considerations that need to be introduced due to the nature of smartphones and tablets and how users interact with them. These devices can be held in at least 4 different natural combinations (vertically up, vertically down, horizontally left, horizontally right), which can effect how gestures are interpreted if templates or features were extracted during training rounds were in a different orientation layout. These considerations are as follows:

1. Location invariance: No matter where the correct gesture is drawn on the screen, it should be authenticated correctly.
2. Scale invariance: No matter what size the correct gesture is drawn to on the screen, it should be authenticated correctly.
3. Rotation invariance: No matter what angle the correct gesture is drawn at on the screen, it should be authenticated correctly.

We would argue that location and scale invariance are important when dealing with cross-platform authentication, because the screen dimension inherently limits what size the gesture can be drawn to and the area over which a gesture can be performed would cause wild variations in where it would be drawn depending on the user. Rotation invariance is useful for reducing computational complexity when dealing with individualized free-form gestures as we had.

How do all of these design choices affect the security of the authentication procedure, in terms of being able to authenticate correctly and not letting false authentication attempts through, remains an open question. Therefore, one obvious way to move forward is to relax and restrict the above choices, and try out the performance with participants.

Speaking heuristically, however, there are obvious tradeoffs. Location invariance sounds quite strict; this requires the user to draw a gesture the same place every time. But this actually is not uncommon; the Android nine pin configuration system already employs this. But the problem there is the fact that the password space is more limited and there are cues for recall, lowering overall security. A more ideal case for a secure gesture input to the device would be a multitouch one where the user draws without the trace being revealed on the screen and without constraint on path as in the Android layout. In this instance, location invariance can be a hindrance; with a free-form gesture like the one described, a user needs to remember exactly where to draw on the screen or otherwise face repeating their trials. Some users could prefer this, hence why configurability should be an option. We do not envision this being as big an issue on a smartphone as it would be on a tablet since the screen size creates more possibilities.

As far as scale invariance, this is an issue that should be addressed in some cases. Algorithms that are scale invariant, while they pass the correct gestures through in spite of their size, also allow other anomalies to happen. Rectangles get interpreted as squares; circular shapes are all condensed to one size. It is also questionable whether or not users can comfortably remember the correct size (at least, after some trials) to draw the gesture to each time. There is some idea that can support this; after enough practice, a user should be able to get most things to the same size every time – same idea behind having consistent handwriting.

Rotation invariance, similar to scale invariance, introduces anomalies as well. An arrow that is drawn facing right would be interpreted the same as an arrow drawn in any other orientation. So any type of directional cues are lost in translation with regards to rotation invariance. Out of the three listed, this should be the easiest one to relax from a user's perspective since it seems intuitive that an 8 should fail to authenticate against an infinity symbol.

GESTURE RECOGNITION TECHNIQUES

Geometric Methods

The current popular methods for this space are the $1 - Recognizer [13] and its appropriate derivatives [2, 1, 6]. Important notice should be given to Rubine [8], who laid the ground work for $1, and is included in this section even though Rubine [8] extracts more features as well for recognition. These methods are effectively a nearest neighbor, template matching approach.

1. These template matching algorithms will begin by resampling the raw gesture to N points, where N is some integer.
2. Afterwards, the resampled gesture is rotated until the angle that the first point in the sequence makes with the centroid of the gesture sequence is zero degrees.
3. The gesture is then translated to the center of the 2D plane it occupies.
4. The gesture then has its size normalized so the points are contained within a unit box.

Users need to enter a number of templates that can be used to train the recognizer, each of which go through the above process. Once there are training templates in the system, inputted gestures undergo the next steps:

1. The recognizer will then do a distance comparison (Euclidean, cosine, Mahalanobis, et cetera) between successive points and aggregate them to compute a score [2],

2. However, after performing the score computation, the gesture needs to be continuously rotated by some angular distance and have another score computed for each rotation - this is to find the best score when comparing to the template, whose orientation may not be the same as the inputted gesture even after normalization [6],[13].

Protractor [6] gets around the last step by solving a minimum angular distance problem and computing the score after that - much quicker than what is seen from Rubine [8] or Wobbrock [13].

Hidden Markov Models

Markov Models contain a discrete series of states and state transitions that correspond to measurable events[4, 7]. For gesture recognition, these events would be processes related with drawing the gesture. For a normal Markov Model, there is a state transition given a known, measurable event – for example, when using a vending machine; unless the appropriate amount of measurable change is present when selecting an item there will be no transition to a state where someone receives what they want from the machine. Hidden Markov Models (HMMs) are called "hidden" because there is no way to measure how the gesture is transitioning based on the input data, hence we must make an estimate as to which state to move to.

Instead, there are multiple possible transitions. There is a probability matrix that describes the chances of transitioning to one state from a given state. A gesture class would be classified with a sequence of states (say, move up or move down, etc.) and probabilities corresponding to a series of state transitions are calculated and compared against each other. Where the gestures are defined and specified by predetermined HMMs. For the system to learn a HMM for a gesture, it needs to take gesture data from the user and adjust the model probabilities to get the best idea for the state transitions. The model that is trained correctly can be used to evaluate and classify the incoming gesture[4, 7]. Downsides to this implementation is that HMMs require a large number of training examples [12] and authenticate slower than the previous two methods [13].

Dynamic Time Warping

Gestures typically record time data along with coordinate points. The gestures tend to be mismatched because users do not enter their gestures at the same rate on successive tries - a plot of the x gesture versus time for two attempts would show that they do not line up and thus they cannot be evaluated against each other; one figure would be a time shifted version of the other, assuming correct inputs. Dynamic Time Warping(DTW) scales the gestures in time such that they can be compared correctly – a matrix is constructed to align the time series path that is filled with these distances or "costs". The matrix is populated with the time difference between successive points in each series – if two gestures have time length of N and M respectively, the matrix is of size NxM consisting of distances between all points in one series with respect to the other. [11, 9].

DTW then works to create the lowest cost path or "optimal warp" between points. This is done by traversing the matrix and imposing boundary conditions; it must start at some corner origin (i,j) of the matrix and end up at another corner (n,m). The algorithm is restricted to not jump in the time index and it cannot go backwards. Assurances are built into the computations such that the traversal path does not stray far from the diagonal of the two sequences in the matrix. The average variance and standard deviation for a gesture against the template is then computed and measured against a threshold to determine authentication [11, 9].

CONCLUSIONS

We have outlined three popular methods that broadly define the gesture recognition space for 2D gestures that can be done on smartphones or tablets along with design considerations for what features an ideal recognizer strives for.

Geometric methods are the simplest and easiest ones to understand; distance based measures that are computed as scores. Computationally, these are also feasible; the resampling can use as little as six points [6] without severe degradation to recognition capability, allowing for fast scoring. It is difficult to get beyond anything more with these, however; they do not require biometric input data and they only work using one set of features – the spatial component of the gesture. At most, current geometric recognizers allow optionalities for rotation invariance but not location or scale. Geometric recognizers currently see some industry use through being implemented in the Android library's classes on gestures.

HMMs are the most mathematically complex and difficult to understand of the ones presented in this paper. Preparing and defining states for the different gesture types as well as algorithmically setting up probability matrices requires training data from the user. They tend to scale well aftet training since new inputs do not affect prior ones; the vocabulary of the model only increases. Tradeoffs, though, are that they require a lot of data to maintain and are not effective until there are already a larger number of training templates than the other two methods; they are computationally weighty. Whitehead et al. [12] concluded that, for their HMM gesture recognizer, a developer should have 250 training samples optimally and around 27 transition states for their 3D recognizer. However, they represent the most possibility of the three presented here to achieve the range of design considerations since the model can use many features to make choices about the state transitions. HMM see adopted use (along with SVM, mentioned earlier) for use with gesture shortcuts (e.g. pinching to zoom and swiping to pan the screen) – these are easier since the set is predefined.

DTW is a hungrier method as compared to geometric recognition. In general, DTW algorithmic complexity obeys a square law, and the template storage is on par with HMM. For mobile devices, this is not recommended on this stance alone. As far as design characteristics, DTW examines differences usually in space or in time but does not give consideration to other features. DTW similarly requires the use of stored templates for matching, as in the geometric method, and can

Criteria	Geometric	HMM	DTW
Sample Invariant?	Yes	Yes	Yes
Trainability	Low	Low	High
Computation	Low	Medium	High
Configurable	Low	Low	High
Resistance	?	?	?
Ease of Use	High	Medium	Low

Table 1. Summary Table. Note that resistance is left as a '?'; this is a highly subjective category that is difficult to properly measure. It will require much future work in a controlled environment to properly determine.

require more templates than that method to achieve similar results [6, 13].

There is not a clear winner yet for a ubiquitous recognizer that can be used strictly for authentication given the considerations outlined. A free, open design space for gesture recognition techniques has given rise to the the novel innovations by researchers outlined in this paper and as a result methods in this area are generated on an ad hoc basis. Ad hoc is used here to mean that researchers decide they want to begin the implementation of a new type of recognizer by utilizing using different features rather than authentication versus whats currently published and show how well it works in comparison. On the flip side of that, entrants to the field who want to implement gesture recognizers have a wide selection and have to do additional work to decide which method they need to use for their application.

It can be suggested that HMM is the most optimal type of recognizer to be used for more varieties of applications but is greedy and more difficult to program. This can additionally be argued with DTW since it requires complicated processing steps on data (e.g. matrix operations).

As far as authentication is concerned, it is best for an entrant to the field to attempt these geometric algorithms since they are neat, ordered, and simple to program. Given the level of customization present in free-form gestures (as far as the appearance of the gesture is concerned), geometric recognizers will work fine when dealing across templates. Without prior knowledge of the gesture itself, an attacker to the device would be stopped by the recognizer unless they were capable of generating a gesture close enough to the stored templates and if that is the case then it's more of a question on the security content of that stored gesture (Is it too simplistic, e.g. a circle?) than a judgment on the recognizer. One way to quantify the security of a generated gesture would be a recent one presentd by Sherman et. al [10]. They generate a security score by measuring the mutual information between a gesture and its template after the removal of "predictable" features.

Overall, a more fluent design approach would be to try to hit goals outlined in the design considerations and outlining how a recognizer succeeds and how it fails when moving ahead to design ones for authentication purposes.

REFERENCES

1. L. Anthony and J. O. Wobbrock. A lightweight multistroke recognizer for user interface prototypes. In *Proc. of GI '10*.
2. L. Anthony and J. O. Wobbrock. $n-protractor: a fast and accurate multistroke recognizer. In *Proc. of GI '12*.
3. C. Bo, L. Zhang, X.-Y. Li, Q. Huang, and Y. Wang. Silentsense: Silent user identification via touch and movement behavioral biometrics. In *Proc. of MobiCom '13*.
4. J. Fierrez, J. Ortega-Garcia, D. Ramos, and J. Gonzalez-Rodriguez. HMM-based on-line signature verification: Feature extraction and signature modeling. *Pattern Recogn. Lett.*, dec 2007.
5. S. A. Grandhi, G. Joue, and I. Mittelberg. Understanding naturalness and intuitiveness in gesture production: insights for touchless gestural interfaces. In *Proc. of CHI '11*.
6. Y. Li. Protractor: a fast and accurate gesture recognizer. In *Proc. of CHI '10*.
7. D. Muramatsu and T. Matsumoto. An HMM on-line signature verifier incorporating signature trajectories. In *Proc. of ICDAR '03*.
8. D. Rubine. Specifying gestures by example. In *Proc. of SIGGRAPH '91*.
9. N. Sae-Bae, K. Ahmed, K. Isbister, and N. Memon. Biometric-rich gestures: a novel approach to authentication on multi-touch devices. In *Proc. of CHI '12*.
10. M. Sherman, G. Clark, Y. Yang, S. Sugrim, A. Modig, J. Lindqvist, A. Oulasvirta, and T. Roos. User-generated free-form gestures for authentication: Security and memorability. In *Proceeding of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '14, 2014.
11. J. Tian, C. Qu, W. Xu, and S. Wang. Kinwrite: Handwriting-based authentication using kinect. In *Proc. of NDSS '13*.
12. A. Whitehead and K. Fox. Device agnostic 3d gesture recognition using hidden markov models. In *Proceedings of the 2009 Conference on Future Play on @ GDC Canada*, Future Play '09, pages 29–30, New York, NY, USA, 2009. ACM.
13. J. O. Wobbrock, A. D. Wilson, and Y. Li. Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. In *Proc. of UIST '07*.

Gesture and Rapport in Mediated Interactions

Andrea Stevenson Won
Communication Department
Stanford University
450 Serra Mall, Stanford
University, Stanford, CA
94305-2050
aswon@stanford.edu

ABSTRACT
Nonverbal behavior plays a number of roles in communication. Synchrony in particular has been proposed to foster rapport, which has been linked to positive outcomes in a number of types of interaction. While these behaviors may occur naturally in face-to-face encounters, they must be designed into computer-mediated systems. This paper describes a new roadmap for assessing the effects of nonverbal behavior on outcome. We describe two published studies as examples of current and ongoing work on tracking and detecting body movements using commercially available computer vision hardware, and using these movements to predict the outcome of interpersonal interactions. This work holds implications for the design of interactive environments that foster collaboration, creativity, and learning.

Author Keywords
Gesture; rapport; synchrony, collaboration; creativity; teaching, learning, computer vision.

ACM Classification Keywords
K.4.3 Computers and Society: Organizational Impacts, Computer-supported collaborative work

INTRODUCTION
Gesture is not solely a means of communication; as has been demonstrated, individuals use gesture to inform cognition even when they are working alone [1]. However, in any interaction, gesture has the potential to facilitate the transfer of information in a number of ways. As Kirsch [2] points out, creating an external object via gestures allows it to be shared as an "object of thought". Gesture thus can augment or clarify the verbal information being conveyed [3]. Even when discussing abstract material, gestures alter how people conceptualize abstract concepts and can assist them in clearly understanding their own meanings [4],[5]. Coordinated gesture and body movements may also facilitate interactions by allowing the interactants to reach a state of rapport with one another, in which attention is focused, emotions are positive, and turn-taking shifts from one participant to the next. This paper will focus on this aspect of gesture-assisted interaction.

We will briefly describe two recently conducted studies examining the automatic detection of gesture in two types of interactions; a dyadic teaching and learning task, and a dyadic collaborative task. The results of these studies allow some speculation on how systems may be designed to support these kinds of interactions, and how other interfaces, such as those using digital agents, can leverage information on the relationships between participants' gestures in an interaction.

TEACHING AND LEARNING INTERACTIONS
Body movements are an important component of teaching and learning in several ways. Beyond the meaningful gestures that explicitly support content, body movements also relate to the attitudes of the participants and outcomes of interactions. Gesture and posture in educational contexts have thus been examined for what they may reveal about teaching and learning (for a review, see Roth, [6]). As the importance of learner control of the process gains recognition [7] the advantages of using nonverbal channels to better understand students' attitudes have become apparent. For example, students' nonverbal behaviors have been recorded and correlated with observers' reports to predict students' levels of engagement, with the goal of developing automated systems that could help predict and assist learning [8],[9].

Since successful communication between teacher and student is one critical component of the learning process, researchers have also examined the development of teacher/student rapport via synchronous nonverbal behavior. LaFrance and Broadbent [10] recorded student and teacher movements in small classrooms and had human coders note whether these students demonstrated synchronous behavior, such as mirrored or matching the teacher's movements. Researchers found a correlation with synchronous movements between teacher and student gestures and students' self reported involvement and rapport. Bernieri used a more gestalt concept of synchrony [11], asking coders to rate perceived movement synchrony (described as "simultaneous movement, tempo similarity, and smoothness") of high school students in teaching/learning dyads, and found a similar relationship between students' self-reported rapport and the observed synchrony of the interaction. In a recent study [12], reciprocal gestures (coded by humans) between teachers and students engaged in a language task not only correlated

with reported rapport, but also with higher student quiz scores.

In learning environments, can automatically detected, summary measures of tracked gestures, predict the outcome of interpersonal interactions?

Using Gesture to Predict Learning Outcomes

We conducted a study assessing the interaction between 53 pairs of teachers and students to investigate how gestures might be used to predict the outcome of a teaching/learning task [13]. In order to capture naturalistic nonverbal behavior of two participants standing at a conversational distance, we used an inexpensive and unobtrusive active computer vision system (two wall-mounted Microsoft Kinects). The Kinect generates an approximation of a human skeleton (shown in Figure 1) for each participant. It records the positions of the nodes, which roughly approximate the major joints, at a rate of approximately 30 frames per second.

We used combinations of sixteen nodes to create 18 angles for each participant, and derived summary measures consisting of the mean, standard deviation and skewness of each of these angles over the entire period of the interaction. We then grouped these measures into five regions that roughly corresponded to the right and left arms and legs, and the head/torso. In this way, we were able to examine specific regions of the body in an anatomically meaningful way without identifying specific gestures. We could also compare the movements of one participant with those of his or her conversational partner. We broke our dataset into subsets of increasingly exclusive high and low success pairs, and then predicted whether a teaching/learning interaction would be successful or unsuccessful.

While it would be premature to draw definite conclusions from the features selected as predictive, or from the correlations between our rough summary measures, this study does offer some indications for future directions.

First, it is significant that we were able to predict success using only these summary movements. More refined models built on this type of easily captured gestural information hold out hope for providing real-time assessment and feedback to participants in naturalistic interactions, among other applications.

Second, the facts that movements of the participants correlated within pairs, and that these correlations were observable using summed measures, implies that this type of body movement tracking may provide a rich resource for further investigations of synchrony. Our second study, described below, aimed to build on these findings and investigate synchrony, asking this question: if gesture is predictive, is synchrony a factor in outcome?

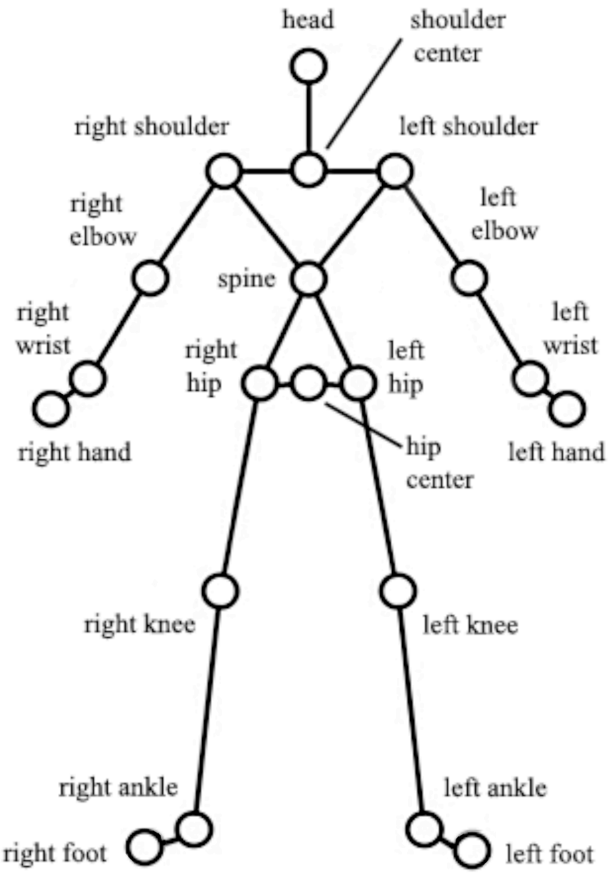


Fig. 1. The modified skeleton derived from Kinect data output in the form of a wireframe. The wireframe consists of 20 nodes with X, Y, and Z values, as well as a 21st node calculated from the other 20 that marks the overall position of the avatar.

COLLABORATIVE INTERACTIONS

In a follow-up to the teaching and learning task described above, we recorded the movements of two participants engaged in a creative collaborative task. We chose this task for several reasons. First, this task was designed to prompt equal participation from both people in the interaction. This served as a contrast to the teaching/learning task, which required the teacher to take a leadership role while the student remained relatively passive. Second, group synergy has been proposed as a key element in creative collaborations [14]. Thus, this was another domain in which we might expect to find synchronous behavior linked to rapport and cooperation. Third, mediated collaborations are a particularly important area of research. Understanding what elements of gesture are important in dyadic and group collaboration is both theoretically important and crucial to understand when considering interfaces that should facilitate such collaboration.

Rapport and Synchrony

Using the methods described above in the teaching and learning task, we analyzed gestures from 52 participant pairs who were tasked with generating ideas for conserving water and energy [15]. Transcribing and rating the responses to the task provided a measure of creative success for each dyad. Two raters scored each response for "appropriate novelty" [16]. Based on the previous study, we proposed that successful collaborations, (collaborations that jointly came up with more creative ideas) would also demonstrate greater synchrony.

In order to assess synchrony, we correlated the movements of each angle of each participant in each pair at each moment in time (similar to the relationship shown between participant 1 and participant 2 in the first two rows of Figure 2). Simple "synchrony scores" were constructed for each pair by averaging the correlations by body regions (similar to the procedure described above). Across pairs, synchrony scores for all regions correlated positively with the pair creativity score, with the head and torso regions correlating particularly highly. In other words, the more correlated a pair's movements were, the higher creativity score they tended to achieve.

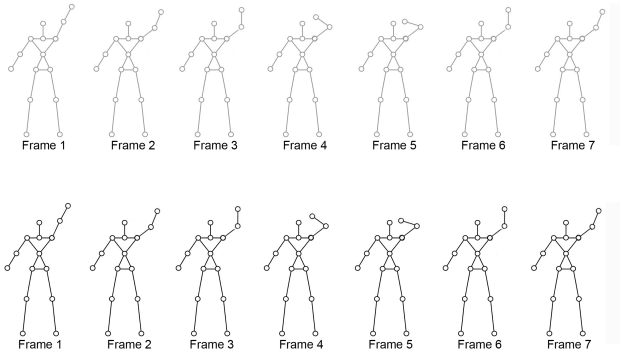


Fig. 2. The figure in gray represents the first participant, and the figure in black represents the second participant. These two rows show the pairs in an interaction that is completely synchronous, yielding a very high correlation.

CONCLUSION

In the papers described above, we were able to predict the outcome of interactions using fairly coarse measures of movement gleaned from a simple computer vision interface. This validates the importance of these body movements, as well as the method of capturing them, which allows the relationship between two participants' gestures to be easily compared. Better understanding of the role of gestures in interpersonal interactions will help in integrating this behavior into collaborative interfaces. The apparent significance of synchronous behavior also underlines the importance of taking interpersonal synchrony into account even in interfaces in which users are not fully embodied, such as touchscreens [18].

While understanding this behavior is useful to guide human interactions, it could also assist in the design of agents. Much research has examined what kinds of nonverbal behavior embodied agents should utilize for greatest effectiveness, including how this behavior should be guided by the nonverbal behavior of the human partner.

Using General Measures of Body Movement

In the two experiments described above, body movements were summed over the entire time period and did not capture individual gestures. Thus, only general conclusions could be drawn about the types of movement that predict successful outcomes. Increasing the granularity of the movements that are captured and assessed allows for understanding the roles of specific categories of gesture.

Similarly, the measure of synchrony used in prediction only reflected the extent to which participants matched their movements during the same moments in time. More subtle measures would be appropriate for capturing mirroring or matching behavior in which first one participant and then another may take the lead. Future research can help confirm in what kinds of interactions synchronous behavior is most important, and include more specific manifestations of synchrony such as mirroring and matching behavior.

In addition, it is important to understand what aspects of the interaction are affected by synchronous behavior. How are the mechanisms involved related to liking and affiliation? How are they related to cognition, increasing the ability of conversants to understand — and possibly even predict — the gestures of their conversational partners?

It will also be important to understand how behavior changes in different contexts. Culture, gender, age, role, and the type of interaction are likely to be factors in how gestures are produced and perceived [18].

One area for useful future work is to examine how body movements may be predictive in different circumstances. For example, when all participants are seated, how do conversational partners synchronize their movements? When more than two people are interacting, how does the pattern of movement pass among participants?

While gestures and body movement can inform observers about aspects of an interaction, performing such gestures can also change the person who engages in them. Detecting and offering feedback on gestures and body movements may be used to alter the outcomes of an interaction. Because tracking body movements in particular may reveal behavior that the participants themselves may not be able to easily observe, such systems may be able to provide information that can assist an interaction in real time. A number of domains could benefit from this kind of feedback. Teachers or tutors could use this to improve teaching outcomes. Physicians might use this to practice building rapport with patients. Receiving input on nonverbal behavior might help users learn to improve social skills, reduce social anxiety, and help in conflict resolution.

The information taken from body movements during a single interaction might also indicate whether a single, brief interaction between two specific individuals is likely to be successful. This information could be used to optimize partnerships by providing feedback to existing pairs, or by reassigning individuals to more compatible partners.
How humans are embodied in interactive environments is important to engagement in these environments, the success of social interactions in these environments, and user enjoyment. It also appears to provide information on whether or not these interactions will be successful. In order to design the most effective interfaces, learning what nonverbal behavior to track, and how it should most effectively be rendered in a mediated or virtual environment, is crucial. Leveraging current technologies to track and interpret the gestures of interactants will aid in this goal.

ACKNOWLEDGMENTS

The work presented herein was funded in part by grant 108084-5031715-4 from the National Science Foundation, and also by Konica Minolta as part of a Stanford Media-X grant. We thank Konica Minolta for the valuable insights provided by their visiting researchers, especially Dr. Haisong Gu. We thank the staff of the Stanford Virtual Human Interaction Lab (VHIL), especially lab manager Cody Karutz.

REFERENCES

1. Kim, M. J., & Maher, M. L. The impact of tangible user interfaces on designers' spatial cognition. *Human–Computer Interaction 23*, 2 (2008), 101-137.
2. Kirsh, D. Thinking with external representations. In *Cognition Beyond the Brain* (pp. 171-194). Springer London (2013).
3. Tversky, B., Heiser, J., Lee, P. and Daniel, M. Explanations in gesture, diagram, and word. In K. R. Coventry, T. Tenbrink, & J. A. Bateman (Editors), Spatial Language and dialogue. Oxford: Oxford University Press. (2009), 119-131.
4. Jamalian, A. and Tversky, B. Gestures alter thinking about time. *In Proceedings of the 34th annual conference of the Cognitive Science Society* (2012), 551-557.
5. Kim, M. J., & Maher, M. L. The impact of tangible user interfaces on spatial cognition during collaborative design. *Design Studies 29*, 3 (2008), 222-253.
6. Roth, W. Gestures: Their Role in Teaching and Learning. *Educational Research 71*, 3 (2011), 365-392.
7. Kay, J. Learner control. *User modeling and user-adapted interaction 11*(1-2), 111-127, 2001.
8. Mota, S. and Picard, R.W. Automated Posture Analysis for detecting Learner's Interest Level. *Journal of Consumer Research*, (2003), 323-339.
9. Dragon, T., Arroyo, I., Woolf, B., Burleson, W., el Kaliouby, R. and Eydgahi, H. Viewing student affect and learning through classroom observation and physical sensors. *In Intelligent Tutoring Systems, Springer Berlin/Heidelberg,* (2008), 29-33.
10. LaFrance, M. and Broadbent, M. Group rapport: Posture sharing. as a nonverbal indicator. *Group and Organization Studies 1*, (1976), 328-333,
11. Bernieri, F.J. Coordinated movement and rapport in teacher student interactions. *Journal of Nonverbal Behavior 12*, (1988), 120-138.
12. Zhou, J., The Effects of Reciprocal Imitation on Teacher–Student Relationships and Student Learning Outcomes. *Mind, Brain, and Education 6,* 2 (2012), 66-73.
13. Won, A.S., Bailenson, J.N., Janssen, J.H. Automatic Detection of Nonverbal Behavior Predicts Learning in Dyadic Interaction. (2014) *Under review.*
14. Kurtzberg, T. R. and Amabile, T. M. From Guilford to creative synergy: Opening the black box of team level creativity. *Creativity Research Journal 13*, 3-4 (2001). 285-294.
15. Won, A. S., Bailenson, J. N., Stathatos, S. C., Dai, W. Automatically Detected Nonverbal Behavior Predicts Creativity in Collaborating Dyads 9. *Journal of Nonverbal Behavior, in press.* (2014).
16. Oppezzo, M. A. Walk for thought: the effects of taking a walk outside on creative ideation. (Unpublished doctoral dissertation). (2012), Stanford University, Stanford, CA., USA.
17. Kleinsmith, A., De Silva, P. R., & Bianchi-Berthouze, N. Cross-cultural differences in recognizing affect from body posture. *Interacting with Computers 18*, 6 (2006), 1371-13.
18. Apted, T., Kay, J., & Quigley, A. Tabletop sharing of digital photographs for the elderly. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* ACM. (2006), 781-790

Mining Expert-driven Models for Synthesis and Classification of Affective Motion

S. Ali Etemad
Department of Systems and Computer Engineering
Carleton University, Ottawa, ON, Canada
ali.etemad@carleton.ca

Ali Arya
School of Information Technology
Carleton University, Ottawa, ON, Canada
arya@carleton.ca

ABSTRACT
Modeling and perception of stylistic human motion are popular fields of research due to their wide range of applications in movies, games, and virtual environments. Currently, most existing techniques for extraction, recognition, and synthesis of affective features in motion utilize datasets of pre-recorded sequences. Nonetheless, we believe an efficient and effective approach for modeling affective features (as well as other types of motion features) is to directly collect them from animators as opposed to traditional techniques. Accordingly, we propose a novel method that allows for implementation of this approach. Using this method, we mine a set of features that can transform a neutral walk into affective variations. To evaluate the outcome, perception studies and affect classification are carried out, validating our approach and results. The findings provide a low-cost and effective alternative to complex computational methods and shed light on how we perform and perceive affective motion.

Author Keywords
Human motion; motion capture; expert-driven; affect; graphical user interface.

ACM Classification Keywords
H.1.2 User/Machine Systems: Human factors; I.3.7 Three-Dimensional Graphics and Realism: Animation.

INTRODUCTION
Detailed psychological and physiological studies on biological motion via point-light (PL) display [1] are dominant means that have lead to our understanding of human motion [2]. Due to the increased use of both interactive and non-interactive motion content in digital media, biological motion studies have become a popular field of research in recent years [3]. Accordingly, the computer science community has recently become heavily engaged in modeling stylistic motion using a variety of different techniques. For example, Rose et al. [4] used relative editing of sequences to blend motion styles successive to alignment via time warping. Lee and Popović [5] proposed a technique that learns behavior styles using a limited number of examples. Probabilistic models were proposed by Brand and Hertzmann [6] and were used along a rich motion dataset to lean interpolated/extrapolated styles using cross-entropy optimization. Arikan et al. [7] proposed a framework for synthesizing user-driven motion sequences based on a dataset of pre-annotated motion data. Physics-based or control-based methods have also been widely investigated for modeling and control of motion. In this area, Liu et al. [8] proposed dynamic simulations with motion capture data to model contact forces on characters.

In general, human motion can be used for bi-directional communication and interaction between humans and computers: an audience observes, perceives, and extracts information from displayed motion scenes or animation (computer to human) while computers can recognize gesture for gesture-based interaction (human to computer). This concept is visualized in Figure 1.

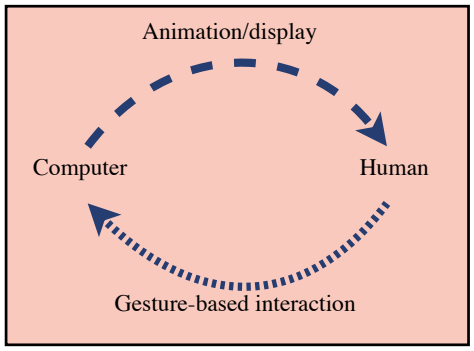


Figure 1. Application of human motion in human computer interaction is presented, where animation/display (computer to human) and gesture-based interaction (human to computer) are the two possible routes.

In this paper, we propose a new method for modeling affective (happy/sad) features in motion, which can be utilized for both of the above-mention purposes. Evidently, most existing methods rely heavily on examples [4, 5] or datasets [6, 7] of pre-recorded human motion. These techniques, despite their effectiveness and accuracy may not always provide the most efficient solutions for the problem at hand. As a result, we propose a novel approach to this problem: recording and analyzing expert animators' inputs as they generate the desired styles in motion. Our approach, we believe, is efficient and effective and can be easily scaled to generate a variety of different motion styles. Comparing the results of our method to some of the related work, we show sufficient correlation, followed by user-evaluations of the results, which validate the perceptual

quality of the features. Moreover, to demonstrate the practicality of the approach for gesture-based interaction, we utilize support vector machines (SVM) trained with the proposed features for recognition of affect from motion sequences. High recognition rates indicate the accuracy of the method and the extracted features.

BACKGROUND

In [9] it was illustrated that a motion sequence, $\mathbf{Y}$, can be described as $\mathbf{Y} = \mathbf{P} + \sum_{i=1}^{r} \mathbf{w}_i.\mathbf{S}_i$, where $\mathbf{Y}$ is the primary or main action class present in the sequence, $\mathbf{S}_i$ are the sets of style features (also referred to as secondary features), and $\mathbf{w}_i$ are the weights or impacts of each secondary feature set. Accordingly, our goal is to take a neutral sequence ($\mathbf{S} = \emptyset$) and generate the set of $\mathbf{S}$ which would convert the initial sequence into an affective/stylistic one. To simplify the problem, we assume that $r = 1$ so that combinational styles such as happy-feminine or angry-young are not considered.

In this study, we used motion capture data as our data type. These data are 3-dimensional spatial recordings of particular markers placed on a body-suit, tracked and recorded using high resolution infrared cameras. In Cartesian space, a recorded motion matrix can be represented by $\mathcal{D} = [\mathbf{R}_1\ \mathbf{R}_2 \ldots \mathbf{R}_n]$, where $\mathbf{R}_1$ through $\mathbf{R}_n$ are the trajectories of the degrees of freedom (DOF) of the representation. Each $\mathbf{R}$ vector has a temporal length of m, representing the total number of frames or time instances in the sequence.

To generate the neutral input sequence, we used a publically available dataset, the HDM05 [10]. In this dataset, multiple motion captures files of actions performed by 5 actors are available. Neutral walks are among them. To increase neutrality of the input sequence and eliminate personal walking styles, we segmented as many 2-step neutral walks from the dataset as possible. This left us with 16 short walking sequences. These sequences were then aligned through time warping [11] and averaged, resulting in a very neutral walk cycle. The accompanying video illustrates the average neutral input. A posture of this sequence is illustrated in Figure 2.

RADIAL BASIS FUNCTIONS FOR AFFECT AND STYLE

In typical animation software such as Autodesk Maya, animators can use free-form transformations for altering motion trajectories. Analyzing and unifying free-form transformations, however, are extremely difficult if not impossible. As a result we propose the use of predefined mathematical functions as constructs for affective features.

In [12] it was illustrated that Gaussian radial basis functions (RBF) are powerful and accurate means for modeling affective/stylistic features in human motion. Moreover, these functions are easy to control and comprehend. A radial function $\phi(r), \phi: \mathbb{R}^s \rightarrow \mathbb{R}$ is defined as a univariate function, where $r = \|t\|_2$, and $\|.\|_2$ is a norm operator such as the Euclidean norm. Consequently, a Gaussian RBF is defined by:

$$\varphi(t; \mu, \sigma^2) = \phi(\|t - \mu\|_2) = exp\left\{\frac{-\|t - \mu\|_2^2}{2\sigma^2}\right\},$$

where μ is the mean and σ^2 is the variance. Accordingly, we can model the affective feature of a given DOF with a weighted sum of M RBFs, resulting in:

$$\sum_j \alpha_j \varphi\left(t; \mu_j, \sigma_j^2\right),$$

where j denotes the number of RBFs used for each DOF, α_j is the set of height or intensity values of the jth RBF, μ_j is the set of means for the jth RBF, and σ_j^2 is the set of variances for the jth RBF. Each parameter set is a multi-dimensional vector due to existence of multiple DOFs and multiple RBFs per DOF.

After introducing RBFs as building blocks for affect features, the goal is for expert animators to generate parameter sets (α, μ, σ^2) which can successfully transform the neutral walk into affective ones.

INTERFACE

In order for expert users to generate the stylistic themes, we developed a graphical interactive interface in MATLAB. The interface allows for users to load a motion capture file in *bvh* format. The motion can immediately be animated in a frame with labeled axis. Different body joints can be selected which are shown in the color red as opposed to all other joints being shown in blue. A particular axis (*x*, *y*, or *z*) of that joint, along which the user intends to modify the joint, needs to be selected as well. The animator can then add up to 3 RBFs to the selected DOF and interactively see the resulting sequences. Each RBF is synthesized using three sliding handles, one for each parameter α, μ, σ^2. The RBFs are generated and added to the sequence in Cartesian space. Once manipulation of a particular joint is concluded, the user can select a different joint. Not all DOFs need to be modified and not all 3 RBFs need to be used for style synthesis to occur. The modified sequence can be animated at any point. Finally, when the task is complete, the generated parameters can be saved for future analysis. Figure 2 illustrates a snapshot of the interface.

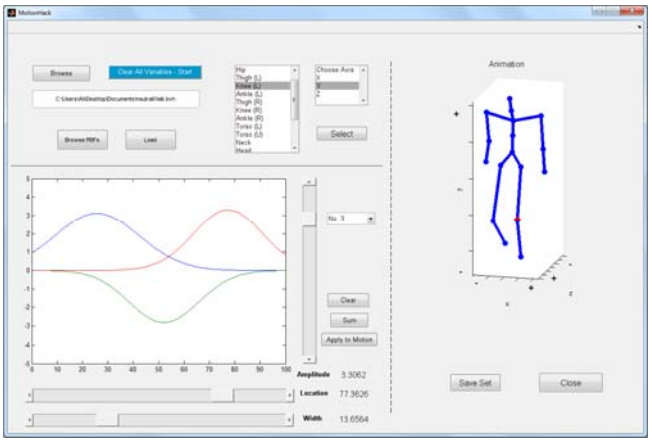


Figure 2. A snapshot of the interface used by animators to generate affective features using Gaussian RBFs.

METHOD

Participants

Two groups of human subjects participated in this study. Ethics approval was secured. The first group consisted of 11 experienced or expert animators, who were either graduate students with related experience or employees of the private sector, working for related companies. Their average age was 25.8 with a standard deviation of 4.2, 9 were males and 2 were females. These participants generated the style features using the system. They were compensated for their time. The second group consisted of 16 participants who were naïve towards motion studies. Their average age was 29.5 with a standard deviation of 4.1, 5 were females and 11 were males.

Process

The animators were asked to use the average neutral walk as input and generate happy and sad variants. They were instructed to alter as many DOFs and use as many RBFs (0, 1, 2, or 3) as they felt needed to complete the task. They were provided with practice time prior to data acquisition. The task took approximately 40 minutes for both happiness and sadness to be completed. The MATLAB based interface was run on a desktop computer with 3 GB of RAM, a 2.8 GHz processor, and a 23.6 inch 1080 HD LED screen. For perception studies of the results, a paper-based force-choice questionnaire was used. More details about this phase of the study is provided in following sections.

RESULTS

The collected features were averaged for all animators. Only modified DOFs were included in the averaging process, meaning not all values were divided by 11. If a particular DOF attracted only $n < 11$ animators, the result was divided by n. This was done to preserve the features added to all DOFs regardless of the frequency of use. As a result, the average parameters $\bar{\alpha}, \bar{\mu}, \overline{\sigma^2}$ were acquired.

Common Shapes

From the raw data, we observe that despite allowing for a variety of combinations to be formed by the basis functions, 4 major types of features are created by the animators. These common shapes are illustrated in Figure 3. The spatially inverted versions of these features were also quite frequently used. An interesting observation was that three-RBF features were almost never used. Other shapes of features were observed, but were very rare and did not demand for a new category of shapes to be introduced.

Posture vs. Motion

It is known that two general types of features compose affective styles in human motion: posture and motion [3]. Posture features are those that often stay unchanged through the sequence. Rather, they are changes to the initial *posture* of the body with which the motion is carried out. Motion features are the changes to the *motion* trajectories. With respect to the utilized model, feature type 1 from Figure 3, and its inverted version, represent posture features since they make a constant change to a particular DOF. Feature types 2, 3, and 4, and their inverted versions, all represent possible motion features. Table 1 presents the percentage of posture and motion features applied to different DOFs by the animators for happy and sad classes. We observe that interestingly, for happiness, the majority of synthesized features are motion-based. For sadness, however, the majority are posture-based.

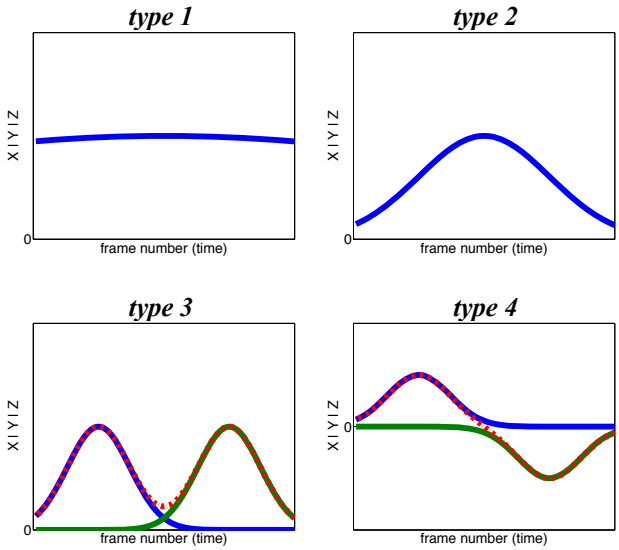


Figure 3. Four common shapes used by animators for generating happy and sad features. Spatially inverted versions of these features were also common.

	Happy	Sad
Motion	71.4%	24.5%
Posture	28.6%	75.5%

Table 1: Percentage of posture and motion features for generating happy and sad walks.

Common Features

Our system uses the 10 most commonly used features to generate affective walks. The accompanying video illustrates the output. A posture from each output is provided in Figure 4. Nevertheless, we hypothesize that by utilizing only a subset of these features, perceptual shortcuts can be taken to generate affective motion. Table 2 presents the 4 most frequently used features. In this table, the term *swing* refers to motion features and the term *tilt* refers to posture features. We observe that all four common features for happiness are motion-based while the first three features for sadness are posture-based.

Perception

The 16 subjects from group two of the participants took part in observing the system-generated affective walks and answering questions regarding the type and amount of affect on a 7-point Likert scale. When happy and sad walks generated using 10 features were presented, the sequences scored a normalized average rating of 0.66 and 0.79

respectively. When only 4 (less than half) of the features were used, however, the ratings showed only a small decline, scoring 0.47 and 0.59 for happy and sad respectively. The results are shown in Figure 5 where error bars denote standard errors. It is important to note that a neutral walk scores 0, and so, our results show acceptable affect synthesis despite using only 4 features. Due to their frequent use, these features are deemed by animators as the most critical. They can therefore, be utilized as very low cost and simple perceptual shortcuts for affect synthesis. The accompanying video shows the happy and sad walks generated using these 4 features. We observe that despite the low cost and simplicity, they are extremely effective, and the intended emotions are visible in the video.

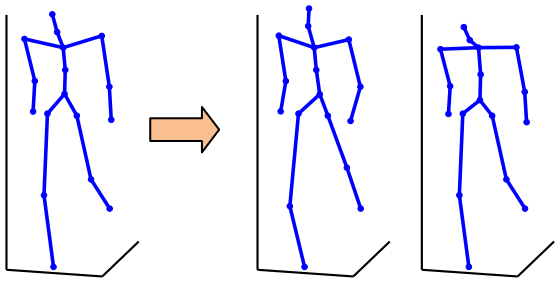


Figure 4. A posture from the neutral input converted to happy and sad using only 10 frequent features.

	1	2	3	4
Happy	Shoulders: increased swing along Z	Wrists: increased swing along Z	Knees: increased swing along Y	Head: increased swing along X
Sad	Shoulders: tilted along –Y	Head: tilted along –Y	Neck: tilted along –Y	Shoulders: decreased swing along Z

Table 2: Three most commonly used features for synthesis of happy and sad walks from a neutral input.

Previous works, mostly from the fields of psychology, have widely studied the features that appear in and result in the perception of affective motion. In [13] the significance of different body parts in perception of affect from motion has been discussed. Other works such as [14] and [15] cite many features, among which those similar to ours can be found. However, to the best of our knowledge, the order in which these features contribute, or are thought by animators to contribute, to perception of affect have not been presented before. Moreover, the direct use of expert-driven data for generation of style has not been widely explored, and our approach, we believe, can provide simple and valuable solutions for the problem at hand.

Feedback

Animators where provided with the goal and methodology of the project after data acquisition was concluded. All 11 indicated significant potential for the method, providing a normalized average 74.0% approval score for the *potential helpfulness of the findings* and 75.3% for *applicability of an autonomous system based on the results*. In regards to the interface itself, however, most found the RBF-based approach to be *unusual* as they were accustomed to free-form transformations. However, we encountered no problems during the data acquisition process after a short (~10 min.) training period.

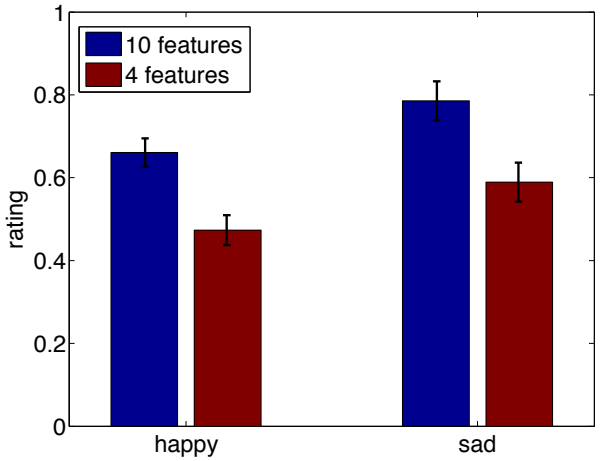


Figure 5. Normalized average perception ratings for affect in the system outputs where 10 and 4 features are used. Error bars represent standard errors.

Classification

To evaluate the accuracy and quality of the extracted features, we trained an SVM classifier with the set of modeled features. 10 input sequences (5 happy and 5 sad) where subsequently classified using the classifier. Table 3 presents the recognition rates, where all 10 sequences were classified successfully. This experiment demonstrates that the proposed method and the modeled features can be utilized for gesture-based interaction. Classifiers can be trained to look for these features and recognize the emotion of humans from their motion inputs. The benefit of training classifiers using the developed set of features as opposed to training using datasets is the higher efficiency and low computational cost of the proposed set of features. Furthermore, the features can readily be used for animation purposes.

	Happy	Sad
Happy	100%	0%
Sad	0%	100%

Table 3: Recognition rates for an SVM classifier. The modeled features are used to train the system.

CONCLUSION

In this paper, we developed a graphical user interface using which animators can manipulate input motion sequences to achieve alternative styles. Gaussian radial basis functions were proposed and utilized as constructs for affective

features. Animators utilized up to three RBFs per DOF to generate happy and sad sequences based on a neutral input. We illustrated that the technique is highly efficient and effective. Moreover, we showed that using a subset of the collected feature set, perceptual shortcuts can be utilized for generating affect. We investigated this notion by using the first four most commonly used features. The results showed that naïve audience perceived the outcome with the intended affect. Accurate classification using SVM also indicated the practicality of the features for gesture-based interaction.

FUTURE WORK

For future work, we intend to expand the research by including a wider range of styles, for example, feminine, masculine, energetic, tired, and others. Moreover, further experiments on perceptual implications of utilizing shortcuts instead of complete feature sets seem necessary. Finally, the impact of the weight parameter for features can be studied; meaning scaled feature sets can be utilized for generating stylistic motion.

ACKNOWLEDGMENTS

This work was supported in part by the Natural Sciences and Engineering Council of Canada (NSERC) and Ontario Centers of Excellence (OCE).

Some of the data used in this project was obtained from HDM05 [10].

REFERENCES

1. Johansson, G. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics 14*, (1973), 195-204.
2. Blake, R. and Shiffrar, M. Perception of human motion. *Annual Review of Psychology 58*, (2007), 47-73.
3. Normoyle, A., Liu, F., Kapadia, M., Badler, N. I. and Jörg, S. The effect of posture and dynamics on the perception of emotion. In *Proc. Symposium on Applied Perception*, ACM Press (2013), 91-98.
4. Rose, C., Cohen, M. F., and Bodenheimer, B. Verbs and adverbs: Multidimensional motion interpolation. *IEEE Computer Graphics and Applications 18*, 5 (1998), 32-40.
5. Lee, S. J. and Popović, Z. Learning behavior styles with inverse reinforcement learning. *ACM Transactions on Graphics 29*, 4 (2010), 122.
6. Brand, M. and Hertzmann, A. Style machines. In *Proc. SIGGRAPH*, ACM Press (2000), 183-192
7. Arikan, O., Forsyth, D. A., and O'Brien, J. F. Motion synthesis from annotations. *ACM Transactions on Graphics 22*, (2003), 3402-408.
8. Liu, L., Yin, K., van de Panne, M., Shao, T., and Xu, W. Sampling-based contact-rich motion control. *ACM Transactions on Graphics 29*, 4 (2010), 128.
9. Etemad, S. A. and Arya, A. Modeling and transformation of 3D human motion. In *Proc. 5th Int. Conf. Computer Graphics Theory and Applications*, (2010), 307-315.
10. Müller, M., Röder, T., Clausen, M., Eberhardt, B., Krüger, B., and Weber, A. Documentation mocap database HDM05. In *Technical report*, (2007), CG-2007-2.
11. Etemad, S. A. and Arya, A. Customizable time warping method for motion alignment. In *Proc. 7th IEEE Int. Conf. Semantic Computing,* (2013), 387-388.
12. Etemad, S. A. and Arya, A. Classification and translation of style and affect in human motion using RBF neural networks. *Neurocomputing 129*, (2014), 585-595.
13. Etemad, S. A., Arya, A., and Parush, A. Additivity in perception of affect from limb motion. *Neuroscience Letters 558*, (2014), 132-136.
14. Wallbott, H. G. Bodily expression of emotion. *European Journal of Social Psychology 28*, (1998), 879-896.
15. Roether, C. L., Omlor, L., Christensen, A., and Giese, M. A. Critical features for the perception of emotion from gait. *Journal of Vision 9*, 6 (2009), 1-32.