User-Driven Design Principles for Gesture Representations

Erin McAweeney[1,2]
erinmc12@uw.edu

Haihua Zhang[1,3]
hh-zh14@mails.tsinghua.edu.cn

Michael Nebeling[1]
nebeling@umich.edu

[1]School of Information, University of Michigan, Ann Arbor, MI

[2]Information School, University of Washington, Seattle, WA

[3]Department of Psychology, Tsinghua University, Beijing, China

ABSTRACT
Many recent studies have explored user-defined interactions for touch and gesture-based systems through end-user elicitation. While these studies have facilitated the user-end of the human-computer dialogue, the subsequent design of gesture representations to communicate gestures *to* the user vary in style and consistency. Our study explores how users interpret, enact, and refine gesture representations adapting techniques from recent elicitation studies. To inform our study design, we analyzed gesture representations from 30 elicitation papers and developed a taxonomy of design elements. We then conducted a partnered elicitation study with 30 participants producing 657 gesture representations accompanied by think-aloud data. We discuss design patterns and themes that emerged from our analysis, and supplement these findings with an in-depth look at users' mental models when perceiving and enacting gesture representations. Finally, based on the results, we provide recommendations for practitioners in need of "visual language" guidelines to communicate possible user actions.

Author Keywords
gesture-based interfaces; gesture representations; human-computer dialogue; end-user elicitation; graphical perception

ACM Classification Keywords
H.5.2. Information interfaces and presentation: User Interfaces—Input devices and strategies, Interaction styles.

INTRODUCTION
Using graphics and illustrations to represent gestures provides a rich vocabulary of design elements for expressing the dimensions of interaction [16]. These representations are common in gesture-based systems as learning supplements for new users, as well as within academic papers communicating findings on user-defined gestures [14,20,21,25,30,31,33,34]. Even less constrained and more sophisticated interaction styles available with the latest wearable and mobile devices, such as virtual and augmented reality interfaces, will both broaden the scope of interaction and increase the need to communicate unfamiliar symbolic input of a foreign systems and/or modalities [30]. Likewise, research that explores gesture input in gesture-based systems will need to communicate new forms of gesture findings with the advent of new interactive technology. For example, recent elicitation studies have been investigating user-defined gestures using a wide range of multi-modal and mid-air interactions [4,5,17,37].

While the range of gesture-based interfaces is expanding, and despite interface design guidelines existing for popular gesture-based technologies[1], there are few broadly accepted conventions to articulate the dimensions of a gesture in a single graphical "language." Disparate design patterns are used across existing guidelines and academic papers (Figure 1) with little guarantee that the users, be it other researchers, system developers, or, as in our study, potential end-users, will interpret the depicted gesture correctly. The trouble is often that existing representations are designed by researchers and designers with the full knowledge of how the gesture was originally enacted, inevitably leading to omissions based on assumptions on what does and does not need to be communicated in the representation.

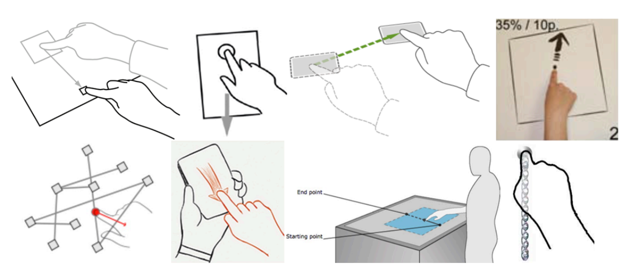

Figure 1. Various representations of the 'single finger swipe' gesture in different academic papers [27,36,10,14,35,5,13,2].

CHI 2018, April 21–26, 2018, Montreal, QC, Canada.

ACM ISBN 978-1-4503-5620-6/18/04…$15.00.
https://doi.org/10.1145/3173574.3174121

[1] https://developer.android.com/guide/practices/ui_guidelines/index.html, https://developer.apple.com/ios/human-interface-guidelines/overview/themes/

Existing elicitation studies have emphasized the user form of gestural input, but little research has been done in the way of understanding the communication of gestures as symbolic output to the user. Narrowing the articulation gap between computer and user, on both ends, leads to quicker learning of a system, faster task performing, reduced error rate, and faster recall [8]. The benefits for closing this articulation gap further are twofold: designers and researchers can focus on the goal of the communication—facilitating intended interaction with their system or expressing findings—and users can feel more directly engaged with, and in control of, interaction techniques.

As there are many options that design elements afford in the articulation of a gesture, we ask what an ideal composition of elements may consist of for seamless communication to the user. *How can gesture representations be designed to help users form the correct mental models to make the leap from symbolic communication to gesture enactment?* A better understanding of how users interpret and enact gestures based on representations would facilitate more concise articulation of available interactions and creation of a more seamless dialogue between the user and the system.

We adapted the end-user elicitation methodology from Wobbrock et al. [38] to investigate how users process gesture representations and how this may contrast with existing mental models around gesture representation interpretation. Our research is made-up of two studies. Study 1 contributes a taxonomy of the graphical elements used to communicate gestures, comprised of six dimensions and 26 categories ranging from multi-frame to finger touchpoints. Study 2 contributes quantitative and qualitative dimensions of user-elicited gesture representations refined with partners. By first analyzing and generating a taxonomy from existing representations in Study 1 similar to [38], we could use the results to systematically compare designer and end-user produced representations from Study 2.

Our observations during production, enactment, and refinement stages illuminated participants' thought processes during interpretation of gesture representations. We analyzed agreement among participants in encoding fundamental aspects of gestures, including motion, position, posture, touch, and time, possibly indicating a shared preference and interpretation of graphical presentation. From the synthesis of these findings, we provide guidance using design principles to create gesture representations with and for end-users.

RELATED WORK

To the best of our knowledge, no work has been done specifically investigating the graphical representation of gestures, but related work has been done in the way of recognition and interpretation of icons, evolving end-user elicitation methodologies, and gesture learning strategies for interactive systems.

Icon Graphical Perception

Communication with icons is widely used in the field of HCI, and researchers have been exploring the taxonomies and measurement of icon characteristics to gain a better understanding of how users perceive and interpret the icon design. Garcia et al. [11] developed a subjective metric for rating the abstractness and concreteness of icons and found that both the concreteness and context affected the identification of icons. McDougall et al. [22] extended the measurement of icon characteristics by establishing rating norms for icon concreteness, complexity, meaningfulness, familiarity, and semantic distance. Furthermore, McDougall et al. [23] also examined the effects of some icon characteristics on users' performance, indicating that it would be possible to design concrete and simple icons to facilitate learnability and efficiency. McDougall and Isherwood [24] found a dynamic pattern of how users utilize the stimulus characteristics, such as familiarity and semantic distance, to identify icons. Based on this previous research, a comprehensive and generalizable icon taxonomy was proposed by Nakamura and Zeng-Treitler [29], including axes of lexical category, semantic category and representation strategy. Their findings of the taxonomy and the three strategies factors provided a robust measurement for assessing and creating icons.

Compared to regular interface icons, gesture representations are usually more complex, containing both concrete body parts and abstract symbols for motion. Little attention has been paid to how various graphical elements are used to construct the representation of gestures, and our research addresses this gap.

Improved Elicitation Studies

To cope with the increasing variety of interactions, researchers have strived for more user-centered gesture design by incorporating users into the design process. Wobbrock et al. [38] proposed the user elicitation method as one way to help understand users' perception and mental models about interaction gestures. This methodology elicits design input from users and finds consensus sets of gestures among users' proposed designs. Since then, researchers have investigated users' preference for various types of gestures with this method. Moreover, Rädle et al. [32] have used elicitation to generate spatially-aware cross-device interactions, and other researches have concerned freehand gestures for Mid-air TV and living room control [7, 28, 37]. Follow-up research results showed that users preferred user-designed gesture sets compared with gestures designed by experts [27]. As gestures are getting more complex with a higher degree of freedom, it is necessary to focus on how to communicate those gestures efficiently to users as well.

Despite the usefulness of the user elicitation method for generating user input, there remains the problem of legacy bias, that is, users' previous experience with interfaces could cause bias in their creation, thus fail to explore more potential new designs. To reduce legacy bias, Morris et al.

[26] proposed three techniques to improve this method: production, priming, and partner. Production forces participants to generate more proposals than the most readily available one. Priming works by showing participants examples of new technologies to inspire creativity. Partner requires users to participate in an elicitation study in groups receiving feedback from partners and improvising. In our study, we used this improved user elicitation method to obtain more creative gesture representation designs from participants, and incorporated an iterative process between participants to better understand the users' mental models.

Gesture Representations for Learning

Novel techniques for users, especially novices, to learn gestures have also been investigated. Bragdon et al. designed GestureBar [3], a new user interface that vividly presents how to execute gestures using animations, tips and practice areas, and conducted a qualitative study to show that GestureBar is preferred by users compared to traditional cheat sheets for learning gesture commands. Freeman et al. [9] developed an in-situ learning system, ShadowGuide, for novice users to efficiently learn and enact complicated multi-touch and whole-hand gestures, by displaying the current posture as feedback, and necessary completion paths as feedforward.

Currently, most researchers and designers use their own way of representing gestures without standards or guidelines in expressing gesture articulation. Therefore, our study examines how users perceive, understand, and identify gesture representations when enacting them. This will facilitate the above-mentioned gesture learning systems by providing suggestions to design representations that can express gesture details effectively to users.

STUDY 1: A TAXONOMY FOR REPRESENTATIONS

Due to the exploratory nature of this research, our first study focused on gaining knowledge of existing processes and tools used to create gesture representations. This study was a two-step process: interviewing designers and researchers of existing representations followed by an in-depth analysis of existing gesture representations in academic papers.

Interviews with Elicitation Researchers

An author contacted six researchers that had led their own elicitation studies [14,20,21,25,30,31,33,34] for an informal interview by email and Skype. Interviewees had previously conducted gesture elicitations in various modalities and had produced gesture representations to communicate findings. Just between these six researchers, the way in which they created representations varied significantly—two utilized photograph with motion lines transposed over them, two were solely text, and the remaining two were illustrations. Researchers drew on experiences besides the paper at hand to speak to other techniques and processes they had personally used in other elicitation studies.

Interviewees were asked specifically about who designed their gesture representations, what the thought process and considerations in choosing graphical elements was, and what tool support, if any, was used or could be useful in creating future gesture representations.

Three common threads concerning the production of gesture representations surfaced in the interviews:

1. Researchers relied on the same tools when creating illustrations: Adobe Illustrator and Photoshop. The most common process was tracing over a photograph using one of these programs and a filter. These tools and techniques researchers felt comfortable and satisfied using.
2. Design elements comprising the representation of gestures were chosen as an informal, ad-hoc process amongst co-authors and often influenced by the representations in the original end-user elicitation paper [38].
3. Researchers spoke to maintaining a balance between the *ease* of production and the *communicative effectivenes*s throughout the production process.

These interviews primarily gave us a better grasp of the current production and use of representations, especially in academic papers. Overall, ad-hoc creations of representations based on past papers or researcher/designer discretion omits the deep consideration of end-user expectations, leaving room for miscommunication. Researchers were focused on communicating gesture sets to readers and/or users, but the lack of design guidelines left the burden on researchers to discern proper design elements, and shifted goals to *ease* of production, detracting from the effort and time out towards the primary goal. Taking out the *how-to* of the design process would let researchers focus back on the original intent: communication of findings. Involving the user in creating representations and deriving principles from there would lead to a clearer design process for researchers and designers when considering the expression of gestures for users in their papers and beyond.

Literature Review

Our analysis of representations began with classifying design elements used to encode gestures in representations in existing literature. We applied a grounded theory approach to account for the lack of an existing theoretical framework [18]. To capture a large breadth of current elicitation studies, our sample was collected from the ACM list of citations from the original end-user elicitation "End-User Defined Gestures for Surface Computing" [38]. We included those papers that contained representations of interactions across all modalities that appeared at CHI, ISS (formerly ITS), and UbiComp conferences. In total, this resulted in 30 papers, encompassing 3 modalities—touch, air, and object-centric. Representations included a variety of production techniques such as tracing, photographs, computer graphics, abstract lines and dots, and text.

We followed an inductive method that included an iterative process through three phases of coding: open, axial, and categorical. Two authors independently coded themes and common design elements found in our data and consolidated categories and dimensions of representations together. Through three iterations of this coding process design elements and higher-level categories of common design patterns emerged from the data. A final set of codes was agreed upon and validated through inter-rater reliability ($k = .910$, $p < .005$).

CLASSIFITACTION OF GESTURE REPRESENTATIONS

Based on the analysis of current elicitation studies we produced a taxonomy of design elements used to communicate gestural interactions (Table 1), consisting of six dimensions and 26 categories.

To the best of our knowledge, this is the first framework that classifies the design elements used in symbolic communication of gesture representations.

Dimension	Category	**Dimension**	Category
Perspective	1st Person	**Color**	Yes
	3rd Person		No
	Mirror	**Gesture Elements**	1-Sided Arrow
	Bird's eye		2-Sided Arrow
	Side Angle		Dotted Lines
Frame	Single		Ghost
	Multi		Finger Trail
Body Context	Full-Body		Other Motion Lines
	Lower-Body		Touchpoints
	Upper-Body		Numbers
	Other Body Parts		Text
Environmental Context	Physical Objects		Bending Joints
	Virtual Objects		Axis

Table 1. Our taxonomy of design elements used in gesture representations. Structural elements are denoted with shading.

Taxonomy of Gesture Representations

The design elements that comprise gesture representations were classified into 6 categories: *Body context, Environmental context, Perspective, Frame, Color,* and *Gesture elements.*

Body context includes extraneous body parts in the representations that were not vital for the enactment of the gesture. An example of this would be a representation including a forearm in a surface gesture that required only the hands and fingers. Similarly, *Environmental context* expresses objects external to the body and not inherently a part of the gesture, such as a screen present in the representation that was not being directly interacted with in the gesture.

Perspective encompasses point-of-view of the representation, including the angle (Mirror, Bird's eye, Side Angle) and whether the representation is read as the user personally enacting it (1st person) or being shown the gesture (3rd person). *Frame* represents whether a single frame was used with other gesture elements to communicate aspects such as motion and time, or if multiple frames were used to communicate a single gesture. *Color* includes the use of both hue and shade for encoding gestural aspects. We grouped *Perspective, Frame,* and *Color* as **structural** elements that are necessary for the design of any representation.

Comparatively, *Gesture elements* (as well as *Body context* and *Environmental context)* is grouped as **details** of a representation, and is added to structure to further communicate aspects of the gesture. *Gesture elements* encompasses all symbols and signs added to the representation. *How* and *when* design elements are used to encode aspects of a gesture in end-user produced representations is discussed in the following sections.

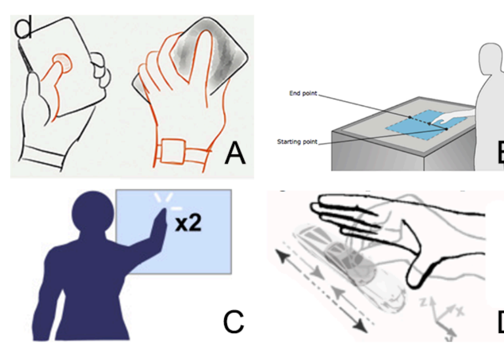

Figure 2. Examples of taxonomy elements used in gesture representations. From top left to bottom right: (A) [5] Color, Multi-frame, 1st Person, Ghost, and Touchpoints. (B) [13] Upper-body, 3rd Person, Single Frame, 1-Sided Arrows, Dotted Lines, Finger Trail, Label and Physical Object. (C) [37] Upper-Body, 3rd Person, Single Frame, Other Motion Lines, Numbers, and Physical Object. (D) [30] Side Angle, 2-Sided Arrows, dotted Lines, Ghost, Axis, Virtual Object.

Analysis

Our coding approach was intended to tease out only graphical communication elements while ensuring the taxonomy was not skewed by modality or gesture type. Through iterations of coding, any elements judged to contribute little communicative value were excluded. For example, *Color* was coded when it added emphasis to the part of the hand in motion, but was not included if the entire representation was in color and thus not used strategically for communicative purposes. Similarly, elements that were inherent to the gesture, rather than a conscious design choice, were also not included. For example, an object-centric gesture enacted with a mobile phone was not coded with the category *Physical Object* since including the mobile phone was fundamental to the gesture itself.

STUDY 2: USER-DEFINED REPRESENTATIONS

We employed a traditional elicitation method [38] and adapted additions to the approach including priming, production, and partners [26]. Our design situated one participant as a designer and the other as a system user providing feedback. The structure *encouraged* feedback, as opposed to traditional elicitation designs that *deter* feedback to keep an unrevised mental model intact. This method has been used in past elicitation studies [19] in which a person, rather than a system, provides feedback to the other participant proceeding initial elicitation. In our study, rather than acting as a recognizer, though, the partner was a first-time user of a "system" that needed to enact the represented gesture presented to them. The gesture could only be communicated to the unknowing partner through representations, as a game of empirical Pictionary. The flow of our study facilitated a more natural dialogue between system and user, in which the use of design elements, and the alteration of those elements, could be observed and talked through. Unaffected user-defined representations in the beginning stages were recorded, and designs were then used in the proceeding simulated user-computer dialogue to refine designs between partners.

Participants

We recruited 30 paid participants (20 were female, 10 were male) through a university-maintained human subjects research platform. Participants had to be 18 years or older. The average age was 23 (*sd= 3.9*). We wanted to recruit end-users that are not expert designers or researchers, but do have experience using gesture-based interfaces. 27 participants reported daily use, the remaining three once a week. Only three participants had an education and/or career history in design. For learning the symbolic input vocabulary of their gesture-based technologies, most reported that they relied on trial-and-error (29 participants), searching online (20 participants), and being taught by someone else (17 participants). Participants were compensated $20 for the approximate 1.5-hour study time.

Choosing Gestures

We chose gestures from the original 30 papers used in our sample from Study 1. Of these papers, we excluded three with text-only representations, since chosen gestures were balanced evenly across our taxonomy and these representations could solely be coded as text. This left 27 remaining papers. One gesture was chosen from each paper, and selected so the entirety of the sample covered every category of the taxonomy.

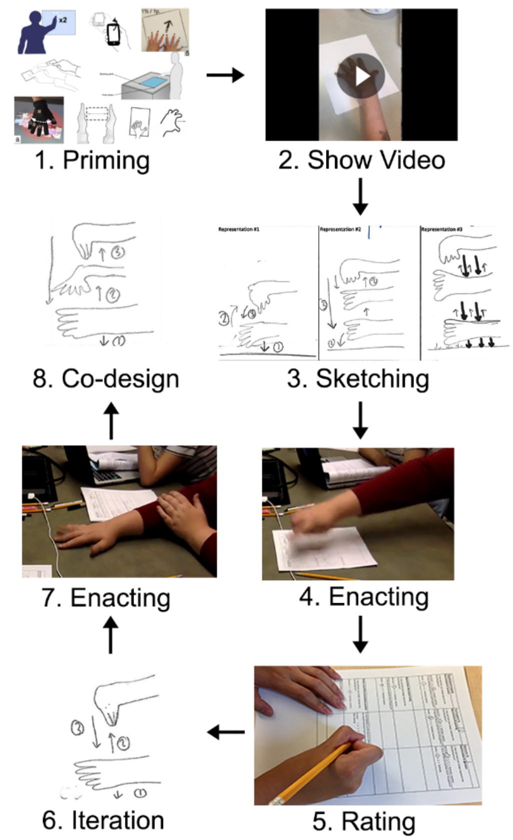

Figure 3. Our methodological process for eliciting gesture representations from users.

Study Procedure

Our procedure was an eight-step process (Figure 3) outlined in detail below.

1. Before using partners [26], participants were first set up sitting at back-to-back screen. They were each primed with ten existing gesture representations presented on their screens, drawn from the sample described earlier evenly balanced across our taxonomy of design elements.
2. Next, each participant was shown five videos of *different* and randomly selected gestures enacted from our pool. Gesture videos shown to participants were all under ten seconds, performed on a tabletop to simulate surface gestures, mid-air to simulate AR/VR gestures, and used appropriate objects for object-centric gestures.
3. Participants were provided pencils, pens, and Sharpies and asked to produce three different representations for each gesture they were presented, thinking aloud as they did so. Afterwards, for each representation, they were asked to rate how well their representation communicated

the gesture in the video on a 1-5 scale, 1 being no resemblance to the original gesture and 5 being a perfect representation of the gesture.

4. One participant (A) presented the three representations that they created for their first gesture to their partner (B). As (B) tried to enact the represented gesture, they talked aloud as to what design elements and combinations of elements translated into their specific movements. While the intended gesture was the same across the production set, (B) was asked to justify each movement according to the representation elements, avoiding a learning bias as they progressed through each representation in the set.
5. After each enactment of the gesture, (A) rated on a separate sheet of paper how well they believed their partner enacted the gesture based on their representation.
6. If the enacted gesture was not determined correct by the researchers, iteration was used. (A) played the role of an adaptive system that, based on the enactment and think-aloud from their partner, chose to make changes to an existing representation, or design a new one completely, to nudge their partner towards a more accurate enactment. Every iteration, even if one of the original three were altered, was captured on video. If the gesture was enacted correctly, we skipped to step (7).
7. (B) then enacted the gesture again, thinking aloud as they did so, based on the new representation. (A) rated again how well their partner enacted this gesture using the new representation.
8. (B) was then shown the correct gesture video or confirmed that they enacted the gesture correctly. Once (B) was aware of the correct gesture, and having been a "user" of the representations, the partners co-designed a final representation, or together decided the (A) produced representation needed no further revisions, and answered follow-up questions about their final design, concerning match, complexity, and meaningfulness of symbols used.

This procedure continued for each elicited set of all five gestures that (A) was shown. Next, the roles for (A) and (B) were switched and the same procedure was followed.

Analysis

For the analysis, we used our taxonomy from Study 1 to code a priori the final 150 representations from participants using the design elements and categories that comprise our taxonomy. We employed content analysis [18] as an approach to systematically code participant produced representations. Two authors separately coded ten elicited representations, validated definitions of categories and dimensions against one another to ensure consistency, and continued coding the remainder of the elicited final representations. Final coding was validated using Cohen's Kappa ($k = .8978, p < .005$).

Dealing with Low-Fidelity End-User Representations

It became clear during the analysis that there are inherent differences in the contexts of representation production between participants and designers and/or researchers. These are:

1. Difference in background knowledge and design insight: designers/researchers had a deeper understanding of the represented gesture.
2. Difference in available time, training, and tools: our participants were given pen and paper, but our researcher interviews revealed mostly digital tools were used and representations created post-hoc.

To account for this in our analysis, we excluded design elements, such as text, that participants had used to compensate for unclear or incomplete drawings. We also differentiated the use of pens, pencils, and sharpies provided as *Color* when used strategically to emphasize various aspects of the gesture, the same as our approach to coding researcher/designer produced representations.

Finding Consensus

Once each final design had been classified in our taxonomy from Study 1, two authors grouped designs within each gesture based on similarity in design elements presented in each representation. Since our gestures were assigned randomly to participants, the number of final representations elicited range from one to nine. We chose to calculate agreement for those gestures that had four or more final representations produced for them (19 total). Although four is a relatively small sample size, given the laborious and iterative process that partners undertook in arriving at these representations, authors agreed that the respective design elements partners chose for their final representations were not coincidental.

Figure 4 illustrates the 19 selected gestures from the literature used in our study and presents agreement scores adapted from Wobbrock et al. [38].

The average agreement across all gestures was ($A = .547$). Generally speaking, this is indicates good consensus among participants in line with other elicitation studies, e.g., performed by Wobbrock [38] (A = .32/.28) and Morris [28] (A = .399). We will attempt to interpret some of the results.

The first two gestures had 100% perfect agreement amongst participants. This was probably because the two gestures were simple and the space for improved articulation was limited. The next segment of gestures had agreement scores of 50-75%. These included a variety of graphical styles and even text. Among the 12 gestures with agreement above 50%, 10 were hand gestures and only 2 were upper-body gestures. 9 contained only one simple motion such as swipe, tap, or slide. The other 3 contained multiple taps or slide with tap, which could be easily expressed in the popular norms of touchpoints and tapping motion line. The last segment is composed of representations for two gestures with agreement of less than 25%. We believe that this is because users were unfamiliar with representing lower-body motion and detailed finger movement.

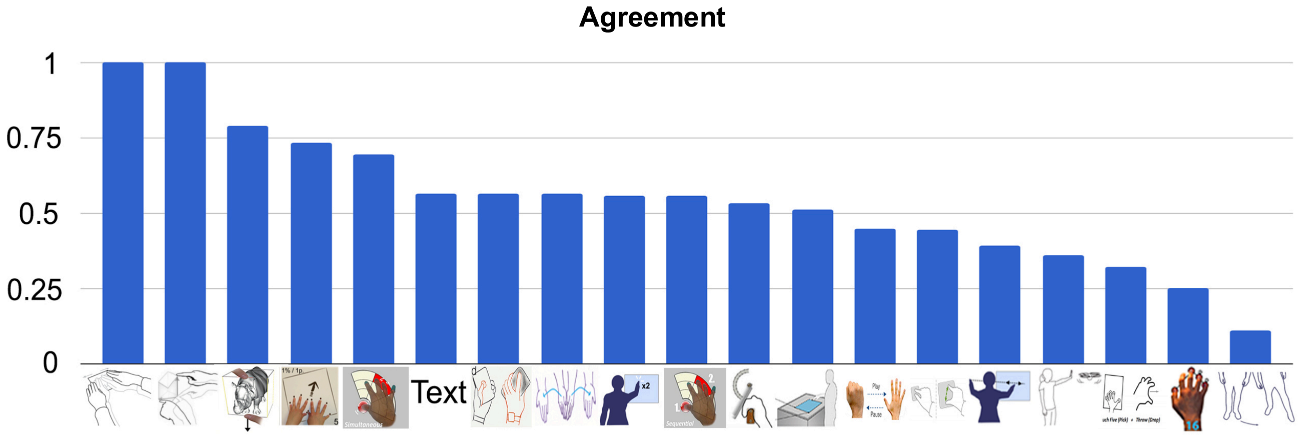

Figure 4. The agreement scores from end-user elicited representations across 19 gestures from academic papers [38,30,6,14,1,28,5,15,37,1,2,13,7,10,37,4,36,12,33].

THEMES IN END USER-ELICITED REPRESENTATIONS

Elicitations, coupled with think-aloud data, gave us a better understanding of *how* taxonomy elements are commonly applied by end-users to encode various aspects of the gesture. The data presented rich qualitative insight into the common design decisions and patterns participants incorporated throughout representations.

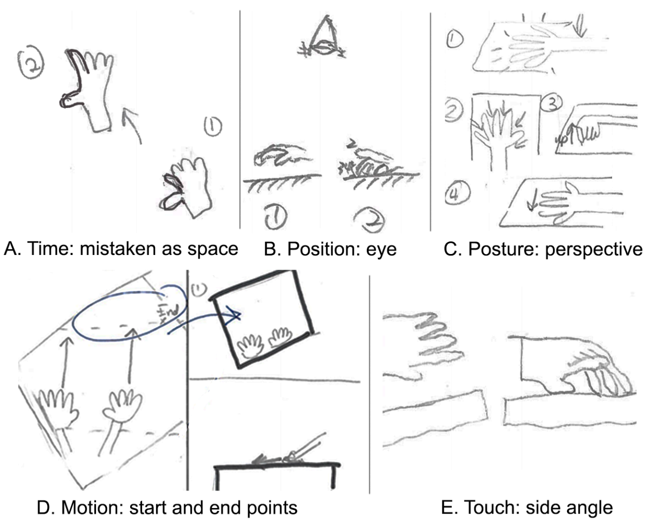

Figure 5. Design themes extracted from user-elicited gesture representations.

Our content analysis uncovered five aspects that a representation often must articulate using various graphical elements and design choices: Time, Position, Posture, Motion, and Touch. In each representation, encodings from our design element taxonomy are used to communicate some or all gesture aspects. Participants encoded these aspects with various design elements from the taxonomy, to varying degrees of success in end-user communication seen with the partner enactment portion of the study. We calculated the most commonly used categories from our taxonomy used to encode these dimensions of a *final* representation, which partners agreed upon as an accurate representation of the gesture (Figure 6). Briefly, listed below are observed common uses and misconceptions concerning these aspects:

- <u>Time</u> illustrates the *sequence* of motion and hand posture in any given gesture. Commonly, time communicated across space was often mistaken as motion. For example, an arrow used to communicate the order of frames (Figure 5, A) was interpreted as the start and end position of the hand and enacted incorrectly by 9 participants. As illustrated in Figure 6, order was most commonly represented by multi-frame and/or numbers.

- <u>Position</u>**,** specifically fixed position and stop and ending points relative to the rest of the body, were most commonly misinterpreted by participants. To remedy this, full or partial body was added for context and point of view, but participants kept extraneous body parts to a minimum to maintain focus on where the motion was taking place. Using a symbol as an eye (Figure 5, B) to illustrate point-of-view and clarify relative position was a new technique not seen in academic papers and used among 3 different participants.

- <u>Posture</u> is the shape of the hand necessary to enact a gesture correctly. Perspective and color were added to emphasize or give clarity to the correct posture of the hand in 16 elicited representations, such as changing the perspective of the representation to better illustrate the hand posture (Figure 5, C). Disrupting perspective consistency confused many participants enacting the gesture, and resulted in a delayed processing and/or wrong enactment.

- <u>Motion</u> was shown in half of final representations with points or arrows on joints specifically where bend and/or movement was happening to help make the gesture more precise. Users consistently expressed during think-aloud enactment the need for clear indication of starting and ending points of a motion, and how many times a motion should be made if double-sided arrows encoded continuous motion. 71 of 90 final representations included start and end indications, and 31 of those used a solid or start position and lighter end positions (8 final representations used the reverse and the others did not differentiate gesture start and end).

- Touch was most commonly communicated by touchpoints, used by 25 of 30 participants (Figure 6), but perspective changes were also used by 9 participants (Figure 5, E). Color was used by 8 participants for both touchpoints and literal shadows above a surface for a hovering gesture. Most users were comfortable with touchpoints and, during iteration, eliminated either touchpoints or the drawn surface, stating that having both surface and touchpoints is redundant, unless used to discern between, e.g., touch and hovering.

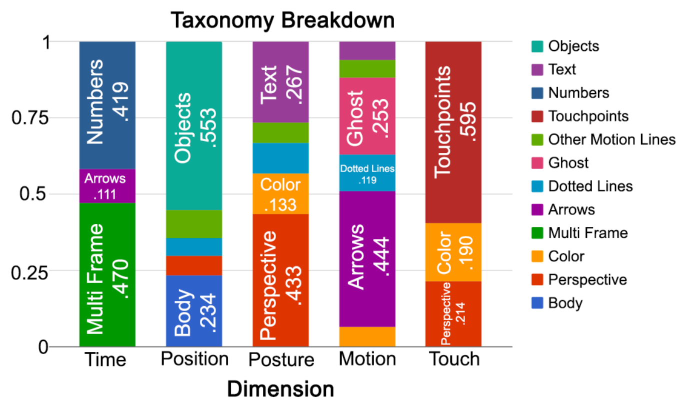

Figure 6. The most common design elements used to encode time, position, posture, motion, and touch of a gestural interaction.

Mental Model Observations

While our analysis of participant-produced elicitations demonstrates *what* and *how* end-users employed design elements in their representations, direct observations of participants and think-aloud recordings gave us more insight into *why* the user's better processed and preferred specific final designs. This section outlines our observations of how participants processed design elements, starting from the broad themes to the use of specific symbols.

Quick Learners

Since there is no agreed upon "language" for communicating gesture interaction, we observed semantic conventions that were established between partners over time when creating and interpreting gesture representations. Many participants began learning the graphical "language" of their partners, such as which symbols meant tap or "above the surface" (Figure 7, A) **(P02:** "Given that I know [my partner] uses this representation [touchpoints] for touching, this helps a lot").

This is beneficial in one-off cases, but interacting with multiple interaction technologies that have diverging and/or contradictory graphical patterns will muddle the mental models of the user. In several instances, when partners contradicted established patterns, their enactors were unable to identify the gesture or misread the gesture due to the violation of an already-established norm.

Searching for Standards, "X Marks the Spot"

Participants wondered if there was a "correct" or "conventional" way to illustrate various gestural elements, seeking standardized design patterns for gesture representations (**P017, G4**: "Is there a universal language for these?" **G2**: "I don't know the conventions for these lines"). Participants sometimes created their own based on cultural knowledge and metaphor, such as using an X for a final destination, which the participant used to communicate "X marks the spot." Commonly understood objects, such as a computer mouse (**P020, G4**), windshield wipers (**P018, G1**), book (**P012, G2**), and stop sign (**P008, G2**) were used as metaphors to communicate motion and posture of the hand (Figure 7, B).

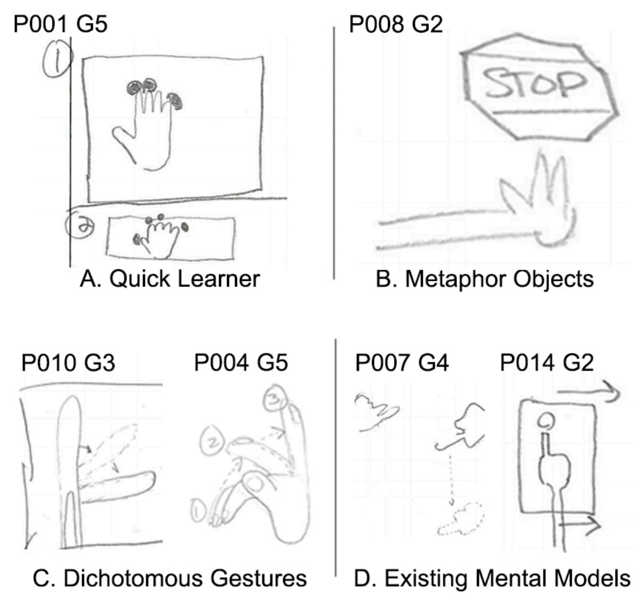

Figure 7. Examples of mental model observations while participants created and enacted gesture representations.

Dichotomous Gestures, Reversible Representations

Different participants were presented with two gestures opposite in orientation (facing left/right) and motion (palm opening/closing). The consensus set for one gesture contained the same elements to express the other gesture (Figure 7, C). As in previous elicitation studies involving gestures [38], this implies that graphical conventions for communicating gestures are intuitive in nature to the user for the fundamental gesture, regardless of orientation. This also indicates the user expectation of symmetry in graphical element use. (**P027** while encoding an outward pinch gesture: "I don't know what is conventional for that, but I know what the inward pinch looks like").

Existing Conventions

It was commonly understood among participants, even before consensus was established, that touchpoints are circles on a surface encircling the fingertips (Figure 7, D). Touchpoints seemed one of the design elements that has become standard in a "graphical lexicon" that participants did not need to establish between their partner. To the gratification of the authors, this common understanding is an indication that more graphical elements, including but not limited to the ones reported in this paper, can become standardized and second-nature for users.

While not always preordained in participants' mental models, dashed or dotted elements (arrows, hands, etc.) signified motion in air, and solid lines were interpreted as on a surface (Figure 7, D). For the ghost effect, 21 participants used solid lines to signify a start position, and lighter or dashed forms for the end position (Figure 7, D).

Modality Assumptions

Surface needed to be indicated between participants, whether that was with touchpoints or a drawn surface. About 70% of surface gesture representations had a surface added to them for the final representation to help guide incorrect partner enactment (such as assuming a mid-air gesture without context).

Constraints on modality for a specific system (i.e. touchscreen is always surface gesture) would clarify this without representing it, but mutli-modal interfaces, e.g., for augmented reality where both mid-air and surface gestures could be available, this needs to be communicated.

Text Over Abstract Symbols, "No Arrows, No Gesture"

Users tended to use straightforward elements, like ghost and text, to communicate motion more often than existing representations, which preferred arrows and other abstract symbols. In general, during the production phase, users started concrete then added necessary symbols, and partners generally agreed ($M = 4.57$ on a 1-5 scale) that added symbols were imperative to the meaning of the representations (**P011**: "No arrows, no gesture").

Although the use of abstract symbols was indispensable, they introduced the need to first establish a common meaning. Often participants labeled arrows, boxes, and circles to make sure partners understood the semantic meaning of their abstract symbols. We found that double the participants incorporated text for graphical elements than designers had in academic papers (*31/150 = .21 vs 3/30 = .10*).

DISCUSSION

What do these results imply for the future production of gesture representations? Our findings indicate disparities in the use of graphical elements between current gesture representations and those the end-users designed for themselves and their partners. Our analysis has shown that 56.5% (78/138) of final end-user elicited representations were substantially different to previous designer and researcher produced representations (12 user-defined representations were excluded as there was no basis for comparison). Although the designers and researchers creating these representations were practiced HCI professionals, studies such as ours with end-users can be more insightful that with experts, an argument substantiated in previous elicitation studies.

While every designer or researcher undoubtedly has unique circumstances to consider in a gesture representation, closing the articulatory distance is a main goal of any symbolic communication regardless of the context and design constraints for the representation [16]. For this reason, the design principles derived from this study are not intended to be a rigid "design language," but rather initial guidelines, used at a designer's discretion, to help a system meet the behaviors and preferences of an end-user. Below we will detail three examples of how our findings can be applied to improve existing gesture representations.

Three Example Applications of Our Findings

Example 1 – An Academic paper

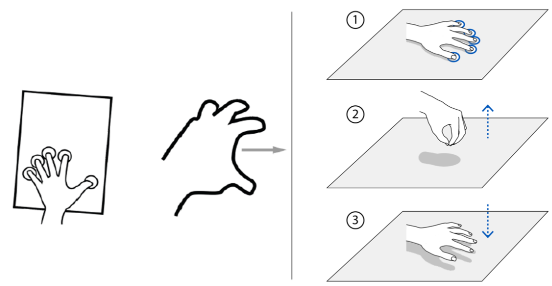

Figure 8. A gesture representation from an academic paper [36] (left) and the principle-guided redesign (right).

The image on the left of Figure 8 was included in an elicitation study paper [36] investigating multi-display environments. Our redesign includes several changes in how the representation aspects (underlined) from our Themes section are communicated:

- The side angle helped better communicate the <u>posture</u> of the hand, the <u>touch</u> of the hand on the surface, and <u>position</u> relative to the surface.
- Numbers show sequence (<u>time</u>) of the frames.
- Touchpoints indicated <u>touch</u> in the first frame, and the lack of touchpoints cues hovering <u>position</u> above surface in the third frame.
- <u>Touch</u> and <u>motion</u> are emphasized with color. This includes shadow to show a lack of touch, hovering above the surface.
- The point of view stays consistent throughout the frames and the surface does not disappear.

The original representation came with the caption "Touch Five (Pick) + Throw (Drop)." The redesigned representation can stand alone without caption or further explanation, making for a more seamless interpretation.

Example 2 – Apple AR

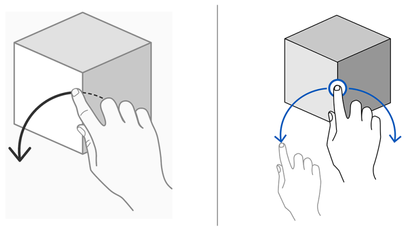

Figure 9. A gesture representation from Apple's developer guidelines[2] (left) and the principle-guided redesign (right).

The second example is from Apple's AR human interface guidelines illustrating action with a virtual object. As opposed to Example 1, the representation on the left was not explained further with a caption, and authors relied on context to infer the referent (rotate cube), and further deduce the gesture from the included design elements. Under these assumptions, the redesign included the following changes:

- Since this is a surface gesture enacted onto a virtual cube through an iPhone screen, the lines are consistently solid to indicate touch. The beginning position also is indicated with a touchpoint.
- Starting and ending points represented with a ghost effect give a concrete beginning and end position. The start position is dark and the end position is lighter.
- The hands do not occlude the motion lines of the finger paths.

Example 3 – Microsoft HoloLens

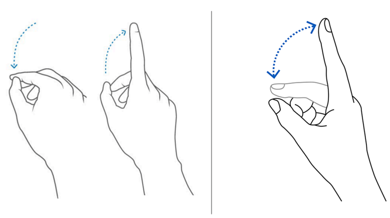

Figure 10. A gesture representation from Microsoft's HoloLens "how-to" online gesture guide[3] (left) and the principle-guided redesign (right).

The final example from Microsoft HoloLens gesture guide is a two-frame GIF that we printed into two distinct frames. This original representation on the left offers a good example of an existing representation in line with our findings since:

- Dotted arrows are used for mid-air motion.
- Side angle emphasizes the posture of the hand.
- Color is used to emphasize motion.
- Two frames show start and end position.

[2] https://developer.apple.com/ios/human-interface-guidelines/technologies/augmented-reality/

[3] https://support.microsoft.com/en-us/help/12644/hololens-use-gestures

There are many elements that align with our findings, but our redesign for this representation focuses specifically on limited animation contexts, such as being printed in an academic paper or an interface that does not support animation and/or cannot load animation in a low-resource setting. We created an example of how this could be best communicated in a still graphic. This would include:

- A single frame avoids redundant and excessive motion expression. We rely on implied motion instead.
- The ghost effect shows a clear beginning and end position, like the original, but conserves space and requires no animation.

Future Work

The study findings open opportunities to test and challenge our design guidelines both in practice and in future studies. Representations following these evidence-based principles still need to be implemented into a UI and validated as a more recognizable and ultimately more effective means of communicating gestures than ad-hoc production. Future use of this study design could also provide participants with more sophisticated tools for production to better understand the use of how end-users would produce digital and animated renderings of gestures and enact them.

Future studies could build on our results and investigate automated tools for producing symbolic communication of interaction to users. This study was a step towards the systematic and formulaic production of gesture representations, which could greatly ease the burden of producing these representations for researchers and designers.

CONCLUSION

Our analysis of user-defined gesture representations has defined guidelines and principles for designing symbolic communication of interaction in both academic papers and gesture-based systems. This study has furthered conversation on human-computer communication and the intentional use of graphical elements to improve clarity and consistency in symbolic communication. These findings can be used by researchers and designers alike to guide future creations of representations that will take end-user behavior and preference into consideration, refining the less explored system-side of the human-computer dialogue.

ACKNOWLEDGEMENTS

We thank all the researchers who volunteered their time and knowledge during interviews, study participants, and all members of the Information Interaction Lab. This project was made possible by the Research Experience for Master's Students program at University of Michigan, the Institute for Museum and Library Services grant RE-01-15-0086-15.

REFERENCES

1. Nikola Banovic, Frank Chun Yat Li, David Dearman, Koji Yatani, and Khai N. Truong. 2011. Design of unimanual multi-finger pie menu interaction. In

Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS '11). ACM, New York, NY, USA, 120-129. DOI: https://doi.org/10.1145/2076354.2076378

2. Andrew Bragdon, Arman Uguray, Daniel Wigdor, Stylianos Anagnostopoulos, Robert Zeleznik, and Rutledge Feman. 2010. Gesture play: motivating online gesture learning with fun, positive reinforcement and physical metaphors. In *ACM International Conference on Interactive Tabletops and Surfaces* (ITS '10). ACM, New York, NY, USA, 39-48. DOI: https://doi.org/10.1145/1936652.1936661
3. Andrew Bragdon, Robert Zeleznik, Brian Williamson, Timothy Miller, and Joseph J. LaViola, Jr.. 2009. GestureBar: improving the approachability of gesture-based interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '09). ACM, New York, NY, USA, 2269-2278. DOI: https://doi.org/10.1145/1518701.1519050
4. Jessica R. Cauchard, Jane L. E, Kevin Y. Zhai, and James A. Landay. 2015. Drone & me: an exploration into natural human-drone interaction. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '15). ACM, New York, NY, USA, 361-365. DOI: https://doi.org/10.1145/2750858.2805823
5. Xiang 'Anthony' Chen, Tovi Grossman, Daniel J. Wigdor, and George Fitzmaurice. 2014. Duet: exploring joint interactions on a smart phone and a smart watch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14). ACM, New York, NY, USA, 159-168. DOI: https://doi.org/10.1145/2556288.2556955
6. Aurélie Cohé, Fabrice Dècle, and Martin Hachet. 2011. tBox: a 3d transformation widget designed for touch-screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11). ACM, New York, NY, USA, 3005-3008. DOI: https://doi.org/10.1145/1978942.1979387
7. Nem Khan Dim, Chaklam Silpasuwanchai, Sayan Sarcar, and Xiangshi Ren. 2016. Designing Mid-Air TV Gestures for Blind People Using User- and Choice-Based Elicitation Approaches. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (DIS '16). ACM, New York, NY, USA, 204-214. DOI: https://doi.org/10.1145/2901790.2901834
8. James D. Foley, Andries van Dam, Steven K. Feiner, and John F. Hughes. 1990. *Computer Graphics: Principles and Practice (2nd Ed.).* Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
9. Alex Forsythe, Noel Sheehy, and Martin Sawey. 2003. Measuring icon complexity: An automated analysis. *Behavior Research Methods, Instruments, & Computers,* 35, 2 (May. 2013), 334-342. DOI:https://doi.org/10.3758/BF03202562
10. Mathias Frisch, Jens Heydekorn, and Raimund Dachselt. 2009. Investigating multi-touch and pen gestures for diagram editing on interactive surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces* (ITS '09). ACM, New York, NY, USA, 149-156. DOI: https://doi.org/10.1145/1731903.1731933
11. Mariano García, Albert N. Badre, and John T. Stask. 1994. Development and validation of icons varying in their abstractness. *Interacting with Computers*, 6, 2 (Jun. 1994), 191-211. DOI:https://doi.org/10.1016/0953-5438(94)90024-8
12. Emilien Ghomi, Stéphane Huot, Olivier Bau, Michel Beaudouin-Lafon, and Wendy E. Mackay. 2013. Arpège: learning multitouch chord gestures vocabularies. In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces* (ITS '13). ACM, New York, NY, USA, 209-218. DOI=http://dx.doi.org/10.1145/2512349.2512795
13. Tobias Hesselmann and Susanne Boll. 2010. SCiVA: a design process for applications on interactive surfaces. In *ACM International Conference on Interactive Tabletops and Surfaces* (ITS '10). ACM, New York, NY, USA, 265-266. DOI: https://doi.org/10.1145/1936652.1936708
14. Uta Hinrichs and Sheelagh Carpendale. 2011. Gestures in the wild: studying multi-touch gesture sequences on interactive tabletop exhibits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11). ACM, New York, NY, USA, 3023-3032. DOI: https://doi.org/10.1145/1978942.1979391
15. Lynn Hoff, Eva Hornecker, and Sven Bertel. 2016. Modifying Gesture Elicitation: Do Kinaesthetic Priming and Increased Production Reduce Legacy Bias?. In *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction* (TEI '16). ACM, New York, NY, USA, 86-91. DOI: https://doi.org/10.1145/2839462.2839472
16. Kasper Hornbæk and Antti Oulasvirta. 2017. What Is Interaction?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17). ACM, New York, NY, USA, 5040-5052. DOI: https://doi.org/10.1145/3025453.3025765
17. Edwin L. Hutchins, James D. Hollan, and Donald A. Norman. 1985. Direct manipulation interfaces. *Hum.-Comput. Interact.* 1, 4 (December 1985), 311-338. DOI=http://dx.doi.org/10.1207/s15327051hci0104_2
18. Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2010. *Research Methods in Human-Computer Interaction*. Wiley Publishing.

19. Sang-Su Lee, Jeonghun Chae, Hyunjeong Kim, Youn-kyung Lim, and Kun-pyo Lee. 2013. Towards more natural digital content manipulation via user freehand gestural interaction in a living room. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing* (UbiComp '13). ACM, New York, NY, USA, 617-626. DOI=http://dx.doi.org/10.1145/2493432.2493480

20. Nicolai Marquardt, Johannes Kiemer, and Saul Greenberg. 2010. What caused that touch?: expressive interaction with a surface through fiduciary-tagged gloves. In *ACM International Conference on Interactive Tabletops and Surfaces* (ITS '10). ACM, New York, NY, USA, 139-142. DOI: https://doi.org/10.1145/1936652.1936680

21. Nicolai Marquardt, Johannes Kiemer, David Ledo, Sebastian Boring, and Saul Greenberg. 2011. Designing user-, hand-, and handpart-aware tabletop interactions with the TouchID toolkit. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces* (ITS '11). ACM, New York, NY, USA, 21-30. DOI: https://doi.org/10.1145/2076354.2076358

22. Siné J. P. Mcdougall, Martin B. Curry, and Oscar de Bruijn. 1999. Measuring symbol and icon characteristics: Norms for concreteness, complexity, meaningfulness, familiarity, and semantic distance for 239 symbols. *Behavior Research Methods*, 31, 3 (Sep. 1999), 487-519. DOI:https://doi.org/10.3758/BF03200730

23. Siné J. P. McDougall, Oscar de Bruijn, and Martin B. Curry. 2000. Exploring the effects of icon characteristics on user performance: the role of icon concreteness, complexity, and distinctiveness. *Journal of Experimental Psychology: Applied*, 6, 4 (Jan. 2001), 291. DOI:http://dx.doi.org/10.1037/1076-898X.6.4.291

24. Siné McDougall and Sarah Isherwood. 2009. What's in a name? The role of graphics, functions, and their interrelationships in icon identification. *Behavior research methods*, 41, 2 (May. 2009) 325-336. DOI:https://doi.org/10.3758/BRM.41.2.325

25. Mark Micire, Munjal Desai, Amanda Courtemanche, Katherine M. Tsui, and Holly A. Yanco. 2009. Analysis of natural gestures for controlling robot teams on multi-touch tabletop surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces* (ITS '09). ACM, New York, NY, USA, 41-48. DOI: https://doi.org/10.1145/1731903.1731912

26. Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, m. c. schraefel, and Jacob O. Wobbrock. 2014. Reducing legacy bias in gesture elicitation studies. *Interactions* 21, 3 (May 2014), 40-45. DOI: https://doi.org/10.1145/2591689

27. Meredith Ringel Morris, Jacob O. Wobbrock, and Andrew D. Wilson. 2010. Understanding users' preferences for surface gestures. In *Proceedings of Graphics Interface 2010* (GI '10). Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 261-268.

28. Meredith Ringel Morris. 2012. Web on the wall: insights from a multimodal interaction elicitation study. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces* (ITS '12). ACM, New York, NY, USA, 95-104. DOI=http://dx.doi.org/10.1145/2396636.2396651

29. Carlos Nakamura and Qing Zeng-Treitler. 2012. A taxonomy of representation strategies in iconic communication. *Int. J. Hum.-Comput. Stud.* 70, 8 (August 2012), 535-551. DOI=http://dx.doi.org/10.1016/j.ijhcs.2012.02.009

30. Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-defined gestures for augmented reality. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '13). ACM, New York, NY, USA, 955-960. DOI: https://doi.org/10.1145/2468356.2468527

31. Dmitry Pyryeskin, Mark Hancock, and Jesse Hoey. 2012. Comparing elicited gestures to designer-created gestures for selection above a multitouch surface. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces* (ITS '12). ACM, New York, NY, USA, 1-10. DOI=http://dx.doi.org/10.1145/2396636.2396638

32. Roman Rädle, Hans-Christian Jetter, Mario Schreiner, Zhihao Lu, Harald Reiterer, and Yvonne Rogers. 2015. Spatially-aware or Spatially-agnostic?: Elicitation and Evaluation of User-Defined Cross-Device Interactions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15). ACM, New York, NY, USA, 3913-3922. DOI: https://doi.org/10.1145/2702123.2702287

33. Jaime Ruiz and Daniel Vogel. 2015. Soft-Constraints to Reduce Legacy and Performance Bias to Elicit Whole-body Gestures with Low Arm Fatigue. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15). ACM, New York, NY, USA, 3347-3350. DOI: https://doi.org/10.1145/2702123.2702583

34. Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11). ACM, New York, NY, USA, 197-206. DOI: https://doi.org/10.1145/1978942.1978971

35. Sebastian Schmidt, Miguel A. Nacenta, Raimund Dachselt, and Sheelagh Carpendale. 2010. A set of multi-touch graph interaction techniques. *In ACM*

International Conference on Interactive Tabletops and Surfaces (ITS '10). ACM, New York, NY, USA, 113-116. DOI: https://doi.org/10.1145/1936652.1936673

36. Teddy Seyed, Chris Burns, Mario Costa Sousa, Frank Maurer, and Anthony Tang. 2012. Eliciting usable gestures for multi-display environments. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces* (ITS '12). ACM, New York, NY, USA, 41-50. DOI=http://dx.doi.org/10.1145/2396636.2396643

37. Radu-Daniel Vatavu. 2012. User-defined gestures for free-hand TV control. In *Proceedings of the 10th European Conference on Interactive TV and Video* (EuroITV '12). ACM, New York, NY, USA, 45-48. DOI: https://doi.org/10.1145/2325616.2325626

38. Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '09). ACM, New York, NY, USA, 1083-1092. DOI: https://doi.org/10.1145/1518701.1518866