User-Defined Gestures for Surface Computing

Jacob O. Wobbrock
The Information School
DUB Group
University of Washington
Seattle, WA 98195 USA
wobbrock@u.washington.edu

Meredith Ringel Morris, Andrew D. Wilson
Microsoft Research
One Microsoft Way
Redmond, WA 98052 USA
{merrie, awilson}@microsoft.com

ABSTRACT
Many surface computing prototypes have employed gestures created by system designers. Although such gestures are appropriate for early investigations, they are not necessarily reflective of user behavior. We present an approach to designing tabletop gestures that relies on eliciting gestures from non-technical users by first portraying the *effect* of a gesture, and then asking users to perform its *cause*. In all, 1080 gestures from 20 participants were logged, analyzed, and paired with think-aloud data for 27 commands performed with 1 and 2 hands. Our findings indicate that users rarely care about the number of fingers they employ, that one hand is preferred to two, that desktop idioms strongly influence users' mental models, and that some commands elicit little gestural agreement, suggesting the need for on-screen widgets. We also present a complete user-defined gesture set, quantitative agreement scores, implications for surface technology, and a taxonomy of surface gestures. Our results will help designers create better gesture sets informed by user behavior.

Author Keywords: Surface, tabletop, gestures, gesture recognition, guessability, signs, referents, think-aloud.

ACM Classification Keywords: H.5.2. Information interfaces and presentation: User Interfaces – *Interaction styles*, *evaluation/methodology*, *user-centered design*.

INTRODUCTION
Recently, researchers in human-computer interaction have been exploring interactive tabletops for use by individuals [29] and groups [17], as part of multi-display environments [7], and for fun and entertainment [31]. A key challenge of surface computing is that traditional input using the keyboard, mouse, and mouse-based widgets is no longer preferable; instead, interactive surfaces are typically controlled via multi-touch freehand gestures. Whereas input devices inherently constrain human motion for meaningful human-computer dialogue [6], surface gestures are versatile and highly varied—almost anything one can do with one's hands could be a potential gesture. To date, most surface gestures have been defined by system designers, who personally employ them or teach them to user-testers [14,17,21,27,34,35]. Despite skillful design, this results in somewhat arbitrary gesture sets whose members may be chosen out of concern for reliable recognition [19]. Although this criterion is important for early prototypes, it is not useful for determining which gestures match those that would be chosen by users. It is therefore timely to consider the types of surface gestures people make *without* regard for recognition or technical concerns.

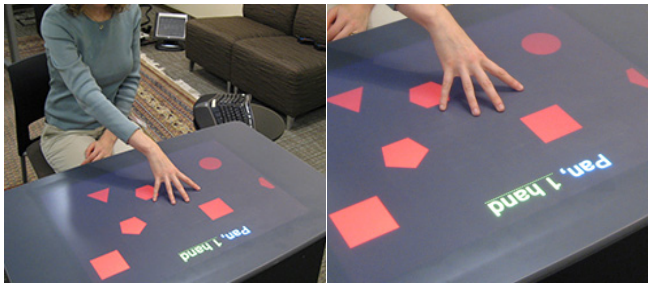

Figure 1. A user performing a gesture to pan a field of objects after being prompted by an animation demonstrating the panning effect.

What kinds of gestures do non-technical users make? In users' minds, what are the important characteristics of such gestures? Does number of fingers matter like it does in many designer-defined gesture sets? How consistently are gestures employed by different users for the same commands? Although designers may organize their gestures in a principled, logical fashion, user behavior is rarely so systematic. As McNeill [15] writes in his laborious study of human discursive gesture, "Indeed, the important thing about gestures is that they are *not* fixed. They are free and reveal the idiosyncratic imagery of thought" (p. 1).

To investigate these idiosyncrasies, we employ a guessability study methodology [33] that presents the *effects* of gestures to participants and elicits the *causes* meant to invoke them. By using a think-aloud protocol and video analysis, we obtain rich qualitative data that illuminates users' mental models. By using custom software with detailed logging on a Microsoft Surface prototype, we obtain quantitative measures regarding gesture timing, activity, and preferences. The result is a detailed picture of user-defined gestures and the mental models and performance that accompany them.Although some prior

CHI 2009, April 4–9, 2009, Boston, Massachusetts, USA.

work has taken a principled approach to gesture definition [20,35], ours is the first to employ users, rather than principles, in the development of a gesture set. Moreover, we explicitly recruited non-technical people without prior experience using touch screens (e.g., the Apple iPhone), expecting that they would behave with and reason about interactive tabletops differently than designers and system builders.

This work contributes the following to surface computing research: (1) a quantitative and qualitative characterization of user-defined surface gestures, including a taxonomy, (2) a user-defined gesture set, (3) insight into users' mental models when making surface gestures, and (4) an understanding of implications for surface computing technology and user interface design. Our results will help designers create better gestures informed by user behavior.

RELATED WORK

Relevant prior work includes studies of human gesture, eliciting user input, and systems defining surface gestures.

Classification of Human Gesture

Efron [4] conducted one of the first studies of discursive human gesture resulting in five categories on which later taxonomies were built. The categories were *physiographics*, *kinetographics*, *ideographics*, *deictics*, and *batons*. The first two are lumped together as *iconics* in McNeill's classification [15]. McNeill also identifies *metaphorics*, *deictics*, and *beats*. Because Efron's and McNeill's studies were based on human discourse, their categories have only limited applicability to interactive surface gestures.

Kendon [11] showed that gestures exist on a spectrum of formality and speech-dependency. From least to most formal, the spectrum was: *gesticulation*, *language-like gestures*, *pantomimes*, *emblems*, and finally, *sign languages*. Although surface gestures do not readily fit on this spectrum, they are a language of sorts, just as direct manipulation interfaces are known to exhibit linguistic properties [6].

Poggi [20] offers a typology of four dimensions along which gestures can differ: *relationship to speech*, *spontaneity*, *mapping to meaning*, and *semantic content*. Rossini [24] gives an overview of gesture measurement, highlighting the movement and positional parameters relevant to gesture quantification.

Tang [26] analyzed people collaborating around a large drawing surface. Gestures emerged as an important element for simulating operations, indicating areas of interest, and referring to other group members. Tang noted *actions* and *functions*, i.e., behaviors and their effects, which are like the signs and referents in our guessability methodology [33].

Morris et al. [17] offer a classification of cooperative gestures among multiple users at a single interactive table. Their classification uses seven dimensions. These dimensions address groups of users and omit issues relevant to single-user gestures, which we cover here.

Working on a pen gesture design tool, Long et al. [13] showed that users are sometimes poor at picking easily differentiable gestures. To address this, our guessability methodology [33] resolves conflicts among similar gestures by using implicit agreement among users.

Eliciting Input from Users

Some prior work has directly employed users to define input systems, as we do here. Incorporating users in the design process is not new, and is most evident in *participatory design* [25]. Our approach of prompting users with *referents*, or effects of an action, and having them perform *signs*, or causes of those actions, was used by Good et al. [9] to develop a command-line email interface. It was also used by Wobbrock et al. [33] to design EdgeWrite unistrokes. Nielsen et al. [19] describe a similar approach.

A limited study similar to the current one was conducted by Epps et al. [5], who presented static images of a Windows desktop on a table and asked users to illustrate various tasks with their hands. They found that the use of an index finger was the most common gesture, but acknowledged that their Windows-based prompts may have biased participants to simply emulate the mouse.

Liu et al. [12] observed how people manipulated physical sheets of paper when passing them on tables and designed their *TNT* gesture to emulate this behavior, which combines rotation and translation in one motion. Similarly, the gestures from the *Charade* system [1] were influenced by observations of presenters' natural hand movements.

Other work has employed a Wizard of Oz approach. Mignot et al. [16] studied the integration of speech and gestures in a PC-based furniture layout application. They found that gestures were used for executing simple, direct, physical commands, while speech was used for high level or abstract commands. Robbe [23] followed this work with additional studies comparing unconstrained and constrained speech input, finding that constraints improved participants' speed and reduced the complexity of their expressions. Robbe-Reiter et al. [22] employed users to design speech commands by taking a subset of terms exchanged between people working on a collaborative task. Beringer [2] elicited gestures in a multimodal application, finding that most gestures involved pointing with an arbitrary number of fingers—a finding we reinforce here. Finally, Voida et al. [28] studied gestures in an augmented reality office. They asked users to generate gestures for accessing multiple projected displays, finding that people overwhelming used finger-pointing.

Systems Utilizing Surface Gestures

Some working tabletop systems have defined designer-made gesture sets. Wu and Balakrishnan [34] built *RoomPlanner*, a furniture layout application for the DiamondTouch [3], supporting gestures for rotation, menu access, object collection, and private viewing. Later, Wu et al. [35] described gesture *registration*, *relaxation*, and *reuse* as elements from which gestures can be built. The gestures

designed in both of Wu's systems were not elicited from users, although usability studies were conducted.

Some prototypes have employed novel architectures. Rekimoto [21] created *SmartSkin*, which supports gestures made on a table or slightly above. Physical gestures for panning, scaling, rotating and "lifting" objects were defined. Wigdor et al. [30] studied interaction on the *underside* of a table, finding that techniques using underside-touch were surprisingly feasible. Tse et al. [27] combined speech and gestures for controlling bird's-eye geospatial applications using multi-finger gestures. Recently, Wilson et al. [32] used a physics engine with Microsoft Surface to enable unstructured gestures to affect virtual objects in a purely physical manner.

Finally, some systems have separated horizontal touch surfaces from vertical displays. Malik et al. [14] defined eight gestures for quickly accessing and controlling all parts of a large wall-sized display. The system distinguished among 1-, 2-, 3-, and 5-finger gestures, a feature our current findings suggest may be problematic for users. Moscovich and Hughes [18] defined three multi-finger cursors to enable gestural control of desktop objects.

DEVELOPING A USER-DEFINED GESTURE SET

User-centered design is a cornerstone of human-computer interaction. But users are not designers; therefore, care must be taken to elicit user behavior profitable for design. This section describes our approach to developing a user-defined gesture set, which has its basis in prior work [9,19,33].

Overview and Rationale

A human's use of an interactive computer system comprises a *user-computer dialogue* [6], a conversation mediated by a language of inputs and outputs. As in any dialogue, *feedback* is essential to conducting this conversation. When something is misunderstood between humans, it may be rephrased. The same is true for user-computer dialogues. Feedback, or lack thereof, either endorses or deters a user's action, causing the user to revise his or her mental model and possibly take a new action.

In developing a user-defined gesture set, we did not want the vicissitudes of gesture recognition to influence users' behavior. Hence, we sought to remove the *gulf of execution* [10] from the dialogue, creating, in essence, a monologue in which the user's behavior is always acceptable. This enables us to observe users' unrevised behavior, and drive system design to accommodate it. Another reason for examining users' unrevised behavior is that interactive tabletops may be used in public spaces, where the importance of immediate usability is high.

In view of this, we developed a user-defined gesture set by having 20 *non-technical* participants perform gestures on a Microsoft Surface prototype (Figure 1). To avoid bias [5], no elements specific to Windows or the Macintosh were shown. Similarly, no specific application domain was assumed. Instead, participants acted in a simple blocks world of 2D shapes. Each participant saw the effect of a gesture (e.g., an object moving across the table) and was asked to perform the gesture he or she thought would cause that effect (e.g., holding the object with the left index finger while tapping the destination with the right). In linguistic terms, the effect of a gesture is the *referent* to which the gestural *sign* refers [15]. Twenty-seven referents were presented, and gestures were elicited for 1 and 2 hands. The system did not attempt to recognize users' gestures, but did track and log all hand contact with the table. Participants used the think-aloud protocol and were videotaped. They also supplied subjective preference ratings.

The final user-defined gesture set was developed in light of the *agreement* participants exhibited in choosing gestures for each command [33]. The more participants that used the same gesture for a given command, the more likely that gesture would be assigned to that command. In the end, our user-defined gesture set emerged as a surprisingly consistent collection founded on actual user behavior.

Referents and Signs[1]

Conceivably, one *could* design a system in which all commands were executed with gestures, but this would be difficult to learn [35]. So what is the right number of gestures to employ? For which commands do users tend to guess the same gestures? If we are to choose a mix of gestures and widgets, how should they be assigned?

To answer these questions, we presented the effects of 27 commands (i.e., the referents) to 20 participants, and then asked them to invent corresponding gestures (i.e., the signs). The commands were application-agnostic, obtained from desktop and tabletop systems [7,17,27,31,34,35]. Some were conceptually straightforward, others more complex. The three authors independently rated each referent's *conceptual complexity* before participants made gestures. Table 1 shows the referents and ratings.

Participants

Twenty paid participants volunteered for the study. Nine were female. Average age was 43.2 years (sd = 15.6). All participants were right-handed. No participant had used an interactive tabletop, Apple iPhone, or similar. All were recruited from the general public and were not computer scientists or user interface designers. Participant occupations included restaurant host, musician, author, steelworker, and public affairs consultant.

Apparatus

The study was conducted on a Microsoft Surface prototype measuring 24" × 18" set at 1024 × 768 resolution. We wrote a C# application to present recorded animations and speech illustrating our 27 referents to the user. For example, for the *pan* referent (Figure 1), a recorded voice said, "Pan. Pretend

[1]To avoid confusing "symbol" from our prior work [33] and "symbolic gestures" in our forthcoming taxonomy, we adopt McNeill's [15] term and use "signs" for the former (pp. 146-147). Thus, *signs* are gestures that execute commands called *referents*.

REFERENTS			REFERENTS		
	Mean	*SD*		*Mean*	*SD*
1. Move a little	1.00	0.00	15. Previous	3.00	0.00
2. Move a lot	1.00	0.00	16. Next	3.00	0.00
3. Select single	1.00	0.00	17. Insert	3.33	0.58
4. Rotate	1.33	0.58	18. Maximize	3.33	0.58
5. Shrink	1.33	0.58	19. Paste	3.33	1.15
6. Delete	1.33	0.58	20. Minimize	3.67	0.58
7. Enlarge	1.33	0.58	21. Cut	3.67	0.58
8. Pan	1.67	0.58	22. Accept	4.00	1.00
9. Close	2.00	0.00	23. Reject	4.00	1.00
10. Zoom in	2.00	0.00	24. Menu access	4.33	0.58
11. Zoom out	2.00	0.00	25. Help	4.33	0.58
12. Select group	2.33	0.58	26. Task switch	4.67	0.58
13. Open	2.33	0.58	27. Undo	5.00	0.00
14. Duplicate	2.67	1.53	**MEAN**	**2.70**	**0.47**

Table 1. The 27 commands for which participants chose gestures. Each command's conceptual complexity was rated by the 3 authors (1=simple, 5=complex). During the study, each command was presented with an animation and recorded verbal description.

you are moving the view of the screen to reveal hidden off-screen content. Here's an example." After the voice finished, our software animated a field of objects moving from left to right. After the animation, the software showed the objects as they were *before* the panning effect, and waited for the user to perform a gesture.

The Surface vision system watched participants' hands from beneath the table and reported contact information to our software. All contacts were logged as ovals having millisecond timestamps. These logs were then parsed by our software to compute trial-level measures.

Participants' hands were also videotaped from four angles. In addition, two authors observed each session and took detailed notes, particularly concerning the think-aloud data.

Procedure

Our software randomly presented 27 referents (Table 1) to participants. For each referent, participants performed a 1-hand and a 2-hand gesture while thinking aloud, and then indicated whether they preferred 1 or 2 hands. After each gesture, participants were shown two 7-point Likert scales concerning gesture goodness and ease. With 20 participants, 27 referents, and 1 and 2 hands, a total of $20 \times 27 \times 2 = 1080$ gestures were made. Of these, 6 were discarded due to participant confusion.

RESULTS

Our results include a gesture taxonomy, the user-defined gesture set, performance measures, subjective responses, and qualitative observations.

Classification of Surface Gestures

As noted in related work, gesture classifications have been developed for human discursive gesture [4,11,15], multimodal gestures with speech [20], cooperative gestures [17], and pen gestures [13]. However, no work has established a taxonomy of surface gestures based on user behavior to capture and describe the gesture design space.

TAXONOMY OF SURFACE GESTURES		
Form	*static pose*	Hand pose is held in one location.
	dynamic pose	Hand pose changes in one location.
	static pose and path	Hand pose is held as hand moves.
	dynamic pose and path	Hand pose changes as hand moves.
	one-point touch	Static pose with one finger.
	one-point path	Static pose & path with one finger.
Nature	*symbolic*	Gesture visually depicts a symbol.
	physical	Gesture acts physically on objects.
	metaphorical	Gesture indicates a metaphor.
	abstract	Gesture-referent mapping is arbitrary.
Binding	*object-centric*	Location defined w.r.t. object features.
	world-dependent	Location defined w.r.t. world features.
	world-independent	Location can ignore world features.
	mixed dependencies	World-independent plus another.
Flow	*discrete*	Response occurs *after* the user acts.
	continuous	Response occurs *while* the user acts.

Table 2. Taxonomy of surface gestures based on 1080 gestures. The abbreviation "w.r.t." means "with respect to."

Taxonomy of Surface Gestures

The authors manually classified each gesture along four dimensions: *form*, *nature*, *binding*, and *flow*. Within each dimension are multiple categories, shown in Table 2.

The scope of the *form* dimension is within one hand. It is applied separately to each hand in a 2-hand gesture. One-point touch and one-point path are special cases of static pose and static pose and path, respectively. These are worth distinguishing because of their similarity to mouse actions. A gesture is still considered a one-point touch or path even if the user casually touches with more than one finger at the same point, as our participants often did. We investigated such cases during debriefing, finding that users' mental models of such gestures involved only one contact point.

In the *nature* dimension, symbolic gestures are visual depictions. Examples are tracing a caret ("^") to perform *insert*, or forming the O.K. pose on the table ("👌") for *accept*. Physical gestures should ostensibly have the same effect on a table with physical objects. Metaphorical gestures occur when a gesture acts on, with, or like something else. Examples are tracing a finger in a circle to simulate a "scroll ring," using two fingers to "walk" across the screen, pretending the hand is a magnifying glass, swiping as if to turn a book page, or just tapping an imaginary button. Of course, the gesture itself usually is not enough to reveal its metaphorical nature; the answer lies in the user's mental model. Finally, abstract gestures have no symbolic, physical, or metaphorical connection to their referents. The mapping is arbitrary, which does not necessarily mean it is poor. Triple-tapping an object to delete it, for example, would be an abstract gesture.

In the *binding* dimension, object-centric gestures only require information about the object they affect or produce. An example is pinching two fingers together on top of an object for *shrink*. World-dependent gestures are defined with respect to the world, such as tapping in the top-right

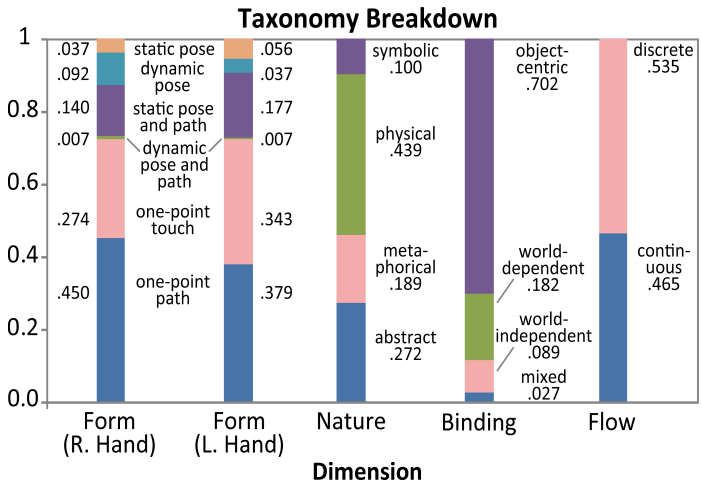

Figure 2. Percentage of gestures in each taxonomy category. From top to bottom, the categories are listed in the same order as they appear in Table 2. The *form* dimension is separated by hands for all 2-hand gestures. (All participants were right-handed.)

corner of the display or dragging an object off-screen. World-independent gestures require no information about the world, and generally can occur anywhere. We include in this category gestures that can occur anywhere *except* on temporary objects that are not world features. Finally, mixed dependencies occur for gestures that are world-independent in one respect but world-dependent or object-centric in another. This sometimes occurs for 2-hand gestures, where one hand acts on an object and the other hand acts anywhere.

A gesture's *flow* is discrete if the gesture is performed, delimited, recognized, and responded to as an event. An example is tracing a question mark ("?") to bring up help. Flow is continuous if ongoing recognition is required, such as during most of our participants' *resize* gestures. Discrete and continuous gestures have been previously noted [35].

Taxonometric Breakdown of Gestures in our Data

We found that our taxonomy adequately describes even widely differing gestures made by our users. Figure 2 shows for each dimension the percentage of gestures made within each category for all gestures in our study.

An interesting question is how the conceptual complexity of referents (Table 1) affected gesture *nature* (Figure 2). The average conceptual complexity for each nature category was: physical (2.11), abstract (2.99), metaphorical (3.26), and symbolic (3.52). Logistic regression indicates these differences were significant ($\chi^2_{(3,N=1074)}$=234.58, p<.0001). Thus, simpler commands more often resulted in physical gestures, while more complex commands resulted in metaphorical or symbolic gestures.

A User-defined Gesture Set

At the heart of this work is the creation of a user-defined gesture set. This section gives the process by which the set was created and properties of the set. Unlike prior gesture sets for surface computing, this set is based on observed user behavior and joins gestures to commands.

Agreement

After all 20 participants had provided gestures for each referent for one and two hands, we grouped the gestures within each referent such that each group held identical gestures. Group size was then used to compute an *agreement score A* that reflects, in a single number, the degree of consensus among participants. (This process was adopted from prior work [33].)

$$A = \frac{\sum_{r \in R} \sum_{P_i \subseteq P_r} \left(\frac{|P_i|}{|P_r|} \right)^2}{|R|} \quad (1)$$

In Eq. 1, r is a referent in the set of all referents R, P_r is the set of proposed gestures for referent r, and P_i is a subset of identical gestures from P_r. The range for A is $[|P_r|^{-1}, 1]$. As an example, consider agreement for *move a little* (2-hand) and *select single* (1-hand). Both had four groups of identical gestures. The former had groups of size 12, 3, 3, and 2; the latter of size 11, 3, 3, and 3. For *move a little*, we compute

$$A_{move\ a\ little} = \left(\frac{12}{20}\right)^2 + \left(\frac{3}{20}\right)^2 + \left(\frac{3}{20}\right)^2 + \left(\frac{2}{20}\right)^2 = 0.42 \quad (2)$$

For *select single*, we compute

$$A_{select\ single} = \left(\frac{11}{20}\right)^2 + \left(\frac{3}{20}\right)^2 + \left(\frac{3}{20}\right)^2 + \left(\frac{3}{20}\right)^2 = 0.37 \quad (3)$$

Agreement for our study is graphed in Figure 3. The overall agreement for 1- and 2-hand gestures was A_{1H}=0.32 and A_{2H}=0.28, respectively. Referents' conceptual complexities (Table 1) correlated significantly and inversely with their agreement (r=-.52, $F_{1,25}$=9.51, p<.01), as more complex referents elicited lesser gestural agreement.

Conflict and Coverage

The user-defined gesture set was developed by taking the largest groups of identical gestures for each referent and assigning those groups' gestures to the referent. However, where the same gesture was used to perform *different* commands, a conflict occurred because one gesture cannot result in different outcomes. To resolve this, the referent with the largest group won the gesture. Our resulting user-defined gesture set (Figure 4) is conflict-free and covers 57.0% of all gestures proposed.

Properties of the User-defined Gesture Set

Twenty-two of 27 referents from Table 1 were assigned dedicated gestures, and the two *move* referents were combined. Four referents were not assigned gestures: *insert*, *maximize*, *task switch*, and *close*. For the first two, the action most participants took comprised more primitive gestures: *insert* used dragging, and *maximize* used enlarging. For the second two, participants relied on imaginary widgets; a common gesture was not feasible. For example, most participants performed *task switch* by tapping an imaginary taskbar button, and *close* by tapping an imaginary button in the top-right corner of an open view.

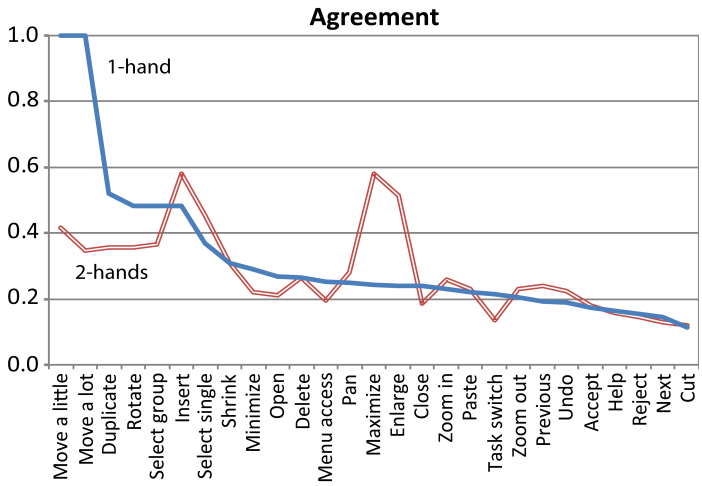

Figure 3. Agreement for each referent sorted in descending order for 1-hand gestures. Two-hand gesture agreement is also shown.

Our user-defined set is useful, therefore, not just for what it contains, but also for what it omits.

Aliasing has been shown to dramatically increase input guessability [8,33]. In our user-defined set, ten referents are assigned 1 gesture, four referents have 2 gestures, three referents have 3 gestures, four referents have 4 gestures, and one referent has 5 gestures. There are 48 gestures in the final set. Of these, 31 (64.6%) are performed with one hand, and 17 (35.4%) are performed with two.

Gratifyingly, a high degree of consistency and symmetry exists in our user-defined set. Dichotomous referents use reversible gestures, and the same gestures are reused for similar operations. For example, *enlarge*, which can be accomplished with four distinct gestures, is performed on an object, but the same four gestures can be used for *zoom in* if performed on the background, or for *open* if performed on a container (e.g., a folder). Flexibility exists insofar as the number of fingers rarely matters and the fingers, palms, or edges of the hands can often be used interchangeably.

Taxonometric Breakdown of User-defined Gestures

As we should expect, the taxonometric breakdown of the final user-defined gesture set (Figure 4) is similar to the proportions of all gestures proposed (Figure 2). Across all taxonomy categories, the average difference between these two sets was only 6.7 percentage points.

Planning, Articulation, and Subjective Preferences

This section gives some of the performance measures and preference ratings for gesture planning and articulation.

Effects on Planning and Articulation Time

Referents' conceptual complexities (Table 1) correlated significantly with average gesture planning time (r=.71, $F_{1,25}$=26.04, p<.0001). In general, the more complex the referent, the more time participants took to begin articulating their gesture. Simple referents took about 8 seconds of planning. Complex referents took about 15 seconds. Conceptual complexity did not, however, correlate significantly with gesture articulation time.

Effects on Goodness and Ease

Immediately after performing each gesture, participants rated it on two Likert scales. The first read, "The gesture I picked is a good match for its intended purpose." The second read, "The gesture I picked is easy to perform." Both scales solicited ordinal responses from 1 = *strongly disagree* to 7 = *strongly agree*.

Gestures that were members of larger groups of identical gestures for a given referent had significantly higher goodness ratings ($\chi^2_{(1,N=1074)}$=34.10, p<.0001), indicating that popularity does, in fact, identify better gestures over worse ones. This finding goes a long way to validating this user-driven approach to gesture design.

Referents' conceptual complexities (Table 1) correlated significantly and inversely with participants' average gesture goodness ratings (r=-.59, $F_{1,25}$=13.30, p<.01). The more complex referents were more likely to elicit gestures rated poor. The simpler referents elicited gestures rated 5.6 on average, while more complex referents elicited gestures rated 4.9. Referents' conceptual complexities did not correlate significantly with average ratings of gesture ease.

Planning time also significantly affected participants' feelings about the goodness of their gestures ($\chi^2_{(1,N=1074)}$=38.98, p<.0001). Generally, as planning time increased, goodness ratings decreased, suggesting that good gestures were those most quickly apparent to participants. Planning time did not affect perceptions of gesture ease.

Unlike planning time, gesture articulation time did not significantly affect goodness ratings, but it *did* affect ease ratings ($\chi^2_{(1,N=1074)}$=17.00, p<.0001). Surprisingly, gestures that took longer to perform were generally rated as easier, perhaps because they were smoother or less hasty. Gestures rated as easy took about 3.4 seconds, while those rated as difficult took about 2.0 seconds. These subjective findings are corroborated by objective counts of finger touch events (*down*, *move*, and *up*), which may be considered rough measures of a gesture's activity or "energy." Clearly, long lived gestures will have more touch events. The number of touch events significantly affected ease ratings ($\chi^2_{(1,N=1074)}$=21.82, p<.0001). Gestures with the fewest touch events were rated as the hardest; those with about twice as many touch events were rated as easier.

Preference for Number of Hands

Overall, participants preferred 1-hand gestures for 25 of 27 referents (Table 1), and were evenly divided for the other two. No referents elicited gestures for which two hands were preferred overall. Interestingly, the referents that elicited equal preference for 1- and 2-hands were *insert* and *maximize*, neither of which were included in the user-defined gesture set because they reused existing gestures. As noted above, the user-designed set (Figure 4) has 31 (64.6%) 1-hand gestures and 17 (35.4%) 2-hand gestures. Although participants' preferences for 1-hand gestures was strong, some 2-hand gestures had good agreement scores and nicely complemented their 1-hand counterparts.

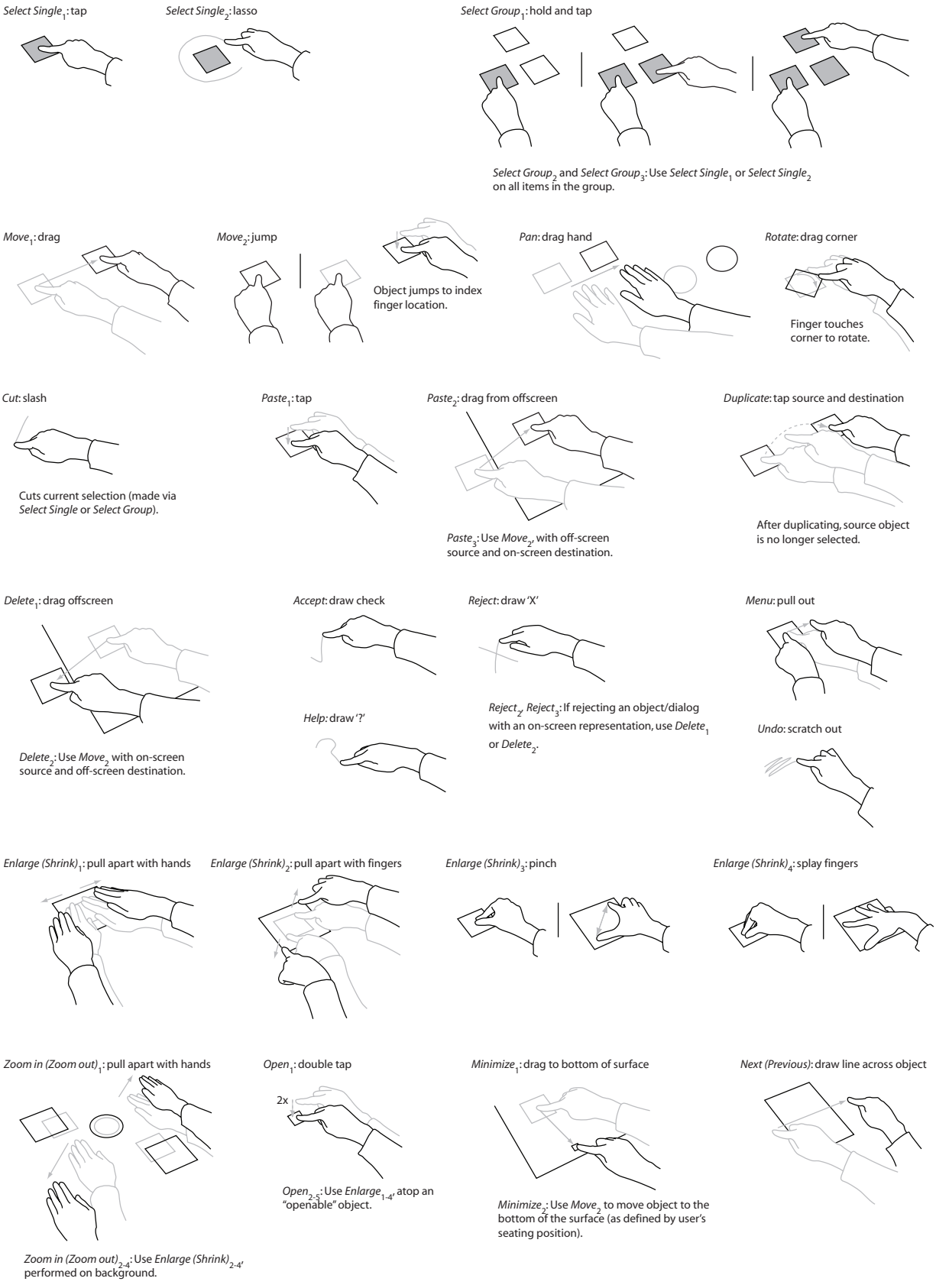

Figure 4. The user-defined gesture set. Gestures depicted as using one finger could be performed with 1-3 fingers. Gestures not depicted as occurring on top of an object are performed on the background region of the surface or full-screen object. To save space, reversible gestures (enlarge/shrink, zoom in/zoom out, next/previous) have been depicted in only one direction.

Mental Model Observations

Our quantitative data were accompanied by considerable qualitative data that capture users' mental models as they choose and perform gestures.

Dichotomous Referents, Reversible Gestures

Examples of dichotomous referents are *shrink* / *enlarge*, *previous* / *next*, *zoom in* / *zoom out*, and so on. People generally employed reversible gestures for dichotomous referents, even though the study software did not present these referents together. This user behavior is reflected in the final user-designed gesture set, where dichotomous referents use reversible gestures.

Simplified Mental Models

The rank order of referents according to conceptual complexity in Table 1 and the order of referents according to descending 1-hand agreement in Figure 3 are not identical. Thus, participants and the authors did not always regard the same referents as "complex." Participants often made simplifying assumptions. One participant, upon being prompted to *zoom in*, said, "Oh, that's the same as enlarge." Similar mental models emerged for *enlarge* and *maximize*, *shrink* and *minimize*, and *pan* and *move*. This allows us to unify the gesture set and disambiguate the effects of gestures based on where they occur, e.g., whether the gesture lands on an object or on the background.

Number of Fingers

Thirteen of 20 participants used varying numbers of fingers when acting on the surface. Of these, only two said that the number of fingers actually mattered. Four people said they often used more fingers for "larger objects," as if these objects required greater force. One person used more fingers for "enlarging actions," the effects of which had something to do with increasing size (e.g., *enlarge*, *open*). Another person felt she used more fingers for commands that executed "a bigger job." One participant said that he used more fingers "to ensure that I was pressing," indicating that to him, more fingers meant more reliable contact. This may be, at least in part, due to the lack of feedback from the table when it was being touched.

Interestingly, two participants who regularly used one-finger touches felt that the system needed to distinguish among fingers. For example, one participant tapped with his ring finger to call up a menu, reasoning that a ring-finger tap would be distinct from a tap with his index finger.

In general, it seemed that touches with 1-3 fingers were considered a "single point," and 5-finger touches or touches with the whole palm were something more. Four fingers, however, constituted a "gray area" in this regard. These findings disagree with many prior tabletop systems that have used designer-made gestures differentiated only on the basis of the number of fingers used [14,17,21,27].

It's a Windows World

Although we took care not to show elements from Windows or the Macintosh, participants still often thought of the desktop paradigm. For example, some gestured as if using a two-button mouse, tapping their index and middle fingers as if clicking. In all, about 72% of gestures were mouse-like one-point touches or paths. In addition, some participants tapped an object first to select it, then gestured on top of the very same object, negating a key benefit of gestures that couples selection and action [13]. The *close* and *task switch* referents were accomplished using imaginary widgets located at objects' top-right and the screen's bottom, respectively. Even with simple shapes, it was clear how deeply rooted the desktop is. Some quotes reveal this: "Anything I can do that mimics Windows—that makes my life easier," "I'm falling back on the old things that I've learned," and "I'm a child of the mouse."

A Land Beyond the Screen

To our surprise, multiple participants conceived of a world beyond the edges of the table's projected screen. For example, they dragged *from* off-screen onto the screen, treating it as the clipboard. They also dragged *to* the off-screen area for *delete* and *reject*. One participant conceived of *different* off-screen areas that meant different things: dragging off the top was *delete*, and dragging off the left was *cut*. For *paste*, she made sure to drag in from the left side, purposefully trying to associate *paste* and *cut*.

Acting above the Table

We instructed participants to touch the table while gesturing. Even so, some participants gestured in ways few tables could detect. One participant placed a hand palm-up on the table and beckoned with her fingers to call for *help*. Another participant put the edges of her hands in an "X" on the table such that the top hand was about 3" off the table's surface. One user "lifted" an object with two hands, placing it on the clipboard. Acting in the air, another participant applied "glue" to an object before pasting it.

DISCUSSION

In this section, we discuss the implications of our results for gesture design, surface technology, and user interfaces.

Users' and Designers' Gestures

Before the study began, the three authors independently designed their own gestures for the 27 referents shown in Table 1. Although the authors are experts in human-computer interaction, it was hypothesized that the "wisdom of crowds" would generate a better set than the authors. Indeed, each author individually came up with only 43.5% of the user-defined gestures. Even combined, the authors only covered 60.9% of the users' set. This suggests that three experts cannot generate the scope of gestures that 20 participants can. That said, 19.1% of each author's gestures were gestures never tried by any participant, which indicates that the authors are either thinking creatively or are hopelessly lost! Either way, the benefit of incorporating users in the development of input systems is clear [9,25,33].

That our participatory approach would produce a coherent gesture set was not clear *a priori*; indeed, it reflects well on our methodology that the proposed gestures seem, in hindsight, to be sensible choices. However, it is worth

noting that the gestures are not, in fact, "obvious"—for example, as mentioned above, each author proposed only 43.5% of the gestures in their own designs. Additionally, the user-defined gesture set differs from sets proposed in the literature, for example, by allowing flexibility in the number of fingers that can be used, rather than binding specific numbers of fingers to specific actions [14,17]. Also, our user-defined gestures differ from prior surface systems by providing multiple gestures for the same commands, which enhances guessability [8,33].

Implications for Surface Technology

Many of the gestures we witnessed had strong implications for surface recognition technology. With the large number of physical gestures (43.9%), for example, the idea of using a *physics engine* [32] rather than a traditional recognizer has support. Seven participants, for example, expected intervening objects to move out of the way when dragging an object into their midst. Four participants "threw" an object off-screen to delete or reject it. However, given the abundance of symbolic, abstract, and metaphorical gestures, a physics engine alone will probably not suffice as an adequate recognizer for all surface gestures.

Although there are considerable practical challenges, tabletop systems may benefit from the ability to look down or sideways at users' hands, rather than just up. Not only does this increase the range of possible gestures, but it provides robustness for users who forget to remain in contact with the surface at all times. Of course, interactive systems that provide feedback will implicitly remind users to remain in contact with the table, but users' unaltered tendencies clearly suggest a use for off-table sensing.

Similarly, systems might employ a low-resolution sensing boundary beyond the high-resolution display area. This would allow the detection of fingers dragging to or from off-screen. Conveniently, these gestures have alternatives in the user-defined set for tables without a sensing boundary.

Implications for User Interfaces

Our study of users' gestures has implications for tabletop user interface design, too. For example, Figure 2 indicates that agreement is low after the first seven referents along the *x*-axis. This suggests that referents beyond this point may benefit from an on-screen widget as well as a gesture. Moreover, enough participants acted on imaginary widgets that system designers might consider using widgets along with gestures for *delete*, *zoom in*, *zoom out*, *accept*, *reject*, *menu access*, and *help*.

Gesture reuse is important to increase learnability and memorability [35]. Our user-designed set emerged with reusable gestures for analogous operations, relying on the target of the gesture for disambiguation. For example, splaying 5 fingers outward on an object will *enlarge* it, but doing so in the background will *zoom in*.

In our study, object boundaries mattered to participants. Multiple users treated object corners as special, e.g., for *rotate*. Hit-testing within objects will be necessary for taking the right action. However, whenever possible, demands for precise positioning should be avoided. Only 2 of 14 participants for *2-hand enlarge* resized along the diagonal; 12 people resized sideways, unconcerned that doing so would perform a non-uniform scale. Similarly, only 1 of 5 used a diagonal "reverse pinch" to resize along the diagonal, while 4 of 5 resized in other orientations.

Gestures should not be distinguished by number of fingers. People generally do not regard the number of fingers they use in the real world, except in skilled activities such as playing the piano, using a stenograph, or giving a massage. Four fingers should serve as a boundary between a few-finger single-point touch and a whole-hand touch.

Limitations and Next Steps

The current study removed the dialogue between user and system to gain insight into users' behavior without the inevitable bias and behavior change that comes from recognizer performance and feedback. But there are drawbacks to this approach. For instance, users could not change previous gestures after moving on to subsequent ones; perhaps users would have performed differently if they first saw *all* referents, and then picked gestures in an order of their choosing. Application context could also impact users' choice of gestures, as could the larger contexts of organization and culture. Our participants were all non-technical literate American adults; undoubtedly, children, Eastern, or uneducated participants would behave differently. These issues are worthy of investigation, but are beyond the scope of the current work. Thankfully, even with a lack of application context and upfront knowledge of all referents, participants still exhibited a substantial level of agreement in making their gestures, allowing us to create a coherent user-defined gesture set.

An important next step is to validate our user-defined gesture set. Unlabeled video clips of the gestures can be shown to 20 *new* participants, along with clips of designers' gestures, to see if people can guess which gestures perform which commands. (This, in effect, reverses the current study to go from signs to referents, rather than from referents to signs.) After, the user-defined gesture set can be implemented with a vision-based gesture recognizer so that system performance and recognition rates can be measured.

CONCLUSION

We have presented a study of surface gestures leading to a user-defined gesture set based on participants' agreement over 1080 gestures. Beyond reflecting user behavior, the user-defined set has properties that make it a good candidate for deployment in tabletop systems, such as ease of recognition, consistency, reversibility, and versatility through aliasing. We also have presented a taxonomy of surface gestures useful for analyzing and characterizing gestures in surface computing. In capturing gestures for this study, we have gained insight into the mental models of non-technical users and have translated these into implications for technology and design. This work

represents a necessary step in bringing interactive surfaces closer to the hands and minds of tabletop users.

REFERENCES

[1] Baudel, T. and Beaudouin-Lafon, M. (1993) Charade: Remote control of objects using free-hand gestures. *Communications of the ACM 36* (7), 28-35.

[2] Beringer, N. (2002) Evoking gestures in SmartKom - Design of the graphical user interface. *Int'l Gesture Workshop 2001, LNCS vol. 2298*. Heidelberg: Springer-Verlag, 228-240.

[3] Dietz, P. and Leigh, D. (2001) DiamondTouch: A multi-user touch technology. *Proc. UIST '01*. New York: ACM Press, 219-226.

[4] Efron, D. (1941) *Gesture and Environment*. Morningside Heights, New York: King's Crown Press.

[5] Epps, J., Lichman, S. and Wu, M. (2006) A study of hand shape use in tabletop gesture interaction. *Ext. Abstracts CHI '06*. New York: ACM Press, 748-753.

[6] Foley, J.D., van Dam, A., Feiner, S.K. and Hughes, J.F. (1996) The form and content of user-computer dialogues. In *Computer Graphics: Principles and Practice*. Reading, MA: Addison-Wesley, 392-395.

[7] Forlines, C., Esenther, A., Shen, C., Wigdor, D. and Ryall, K. (2006) Multi-user, multi-display interaction with a single-user, single-display geospatial application. *Proc. UIST '06.* New York: ACM Press, 273-276.

[8] Furnas, G.W., Landauer, T.K., Gomez, L.M. and Dumais, S.T. (1987) The vocabulary problem in human-system communication. *Communications of the ACM 30* (11), 964-971.

[9] Good, M.D., Whiteside, J.A., Wixon, D.R. and Jones, S.J. (1984) Building a user-derived interface. *Communications of the ACM 27* (10), 1032-1043.

[10] Hutchins, E.L., Hollan, J.D. and Norman, D.A. (1985) Direct manipulation interfaces. *Human-Computer Interaction 1* (4), 311-388.

[11] Kendon, A. (1988) How gestures can become like words. In *Crosscultural Perspectives in Nonverbal Communication*, F. Poyatos (ed). Toronto: C. J. Hogrefe, 131-141.

[12] Liu, J., Pinelle, D., Sallam, S., Subramanian, S. and Gutwin, C. (2006) TNT: Improved rotation and translation on digital tables. *Proc. GI '06*. Toronto: CIPS, 25-32.

[13] Long, A.C., Landay, J.A. and Rowe, L.A. (1999) Implications for a gesture design tool. *Proc. CHI '99*. New York: ACM Press, 40-47.

[14] Malik, S., Ranjan, A. and Balakrishnan, R. (2005) Interacting with large displays from a distance with vision-tracked multi-finger gestural input. *Proc. UIST '05*. New York: ACM Press, 43-52.

[15] McNeill, D. (1992) *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press.

[16] Mignot, C., Valot, C. and Carbonell, N. (1993) An experimental study of future 'natural' multimodal human-computer interaction. *Conference Companion INTERCHI '93*. New York: ACM Press, 67-68.

[17] Morris, M.R., Huang, A., Paepcke, A. and Winograd, T. (2006) Cooperative gestures: Multi-user gestural interactions for co-located groupware. *Proc. CHI '06*. New York: ACM Press, 1201-1210.

[18] Moscovich, T. and Hughes, J.F. (2006) Multi-finger cursor techniques. *Proc. GI '06*. Toronto: CIPS, 1-7.

[19] Nielsen, M., Störring, M., Moeslund, T.B. and Granum, E. (2004) A procedure for developing intuitive and ergonomic gesture interfaces for HCI. *Int'l Gesture Workshop 2003, LNCS vol. 2915*. Heidelberg: Springer-Verlag, 409-420.

[20] Poggi, I. (2002) From a typology of gestures to a procedure for gesture production. *Int'l Gesture Workshop 2001, LNCS vol. 2298*. Heidelberg: Springer-Verlag, 158-168.

[21] Rekimoto, J. (2002) SmartSkin: An infrastructure for freehand manipulation on interactive surfaces. *Proc. CHI '02*. New York: ACM Press, 113-120.

[22] Robbe-Reiter, S., Carbonell, N. and Dauchy, P. (2000) Expression constraints in multimodal human-computer interaction. *Proc. IUI '00*. New York: ACM Press, 225-228.

[23] Robbe, S. (1998) An empirical study of speech and gesture interaction: Toward the definition of ergonomic design guidelines. *Conference Summary CHI '98*. New York: ACM Press, 349-350.

[24] Rossini, N. (2004) The analysis of gesture: Establishing a set of parameters. *Int'l Gesture Workshop 2003, LNCS vol. 2915*. Heidelberg: Springer-Verlag, 124-131.

[25] Schuler, D. and Namioka, A. (1993) *Participatory Design: Principles and Practices*. Hillsdale, NJ: Lawrence Erlbaum.

[26] Tang, J.C. (1991) Findings from observational studies of collaborative work. *Int'l J. Man-Machine Studies 34* (2), 143-160.

[27] Tse, E., Shen, C., Greenberg, S. and Forlines, C. (2006) Enabling interaction with single user applications through speech and gestures on a multi-user tabletop. *Proc. AVI '06*. New York: ACM Press, 336-343.

[28] Voida, S., Podlaseck, M., Kjeldsen, R. and Pinhanez, C. (2005) A study on the manipulation of 2D objects in a projector/camera-based augmented reality environment. *Proc. CHI '05*. New York: ACM Press, 611-620.

[29] Wellner, P. (1993) Interacting with paper on the DigitalDesk. *Communications of the ACM 36* (7), 87-96.

[30] Wigdor, D., Leigh, D., Forlines, C., Shipman, S., Barnwell, J., Balakrishnan, R. and Shen, C. (2006) Under the table interaction. *Proc. UIST '06*. New York: ACM Press, 259-268.

[31] Wilson, A.D. (2005) PlayAnywhere: A compact interactive tabletop projection-vision system. *Proc. UIST '05*. New York: ACM Press, 83-92.

[32] Wilson, A.D., Izadi, S., Hilliges, O., Garcia-Mendoza, A. and Kirk, D. (2008) Bringing physics to the surface. *Proc. UIST '08*. New York: ACM Press, 67-76.

[33] Wobbrock, J.O., Aung, H.H., Rothrock, B. and Myers, B.A. (2005) Maximizing the guessability of symbolic input. *Ext. Abstracts CHI '05*. New York: ACM Press, 1869-1872.

[34] Wu, M. and Balakrishnan, R. (2003) Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. *Proc. UIST '03*. New York: ACM Press, 193-202.

[35] Wu, M., Shen, C., Ryall, K., Forlines, C. and Balakrishnan, R. (2006) Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. *Proc. TableTop '06*. Washington, D.C.: IEEE Computer Society, 185-192.