A User-Developed 3-D Hand Gesture Set for Human–Computer Interaction

Anna Pereira, University of California at Berkeley, **Juan P. Wachs**, Purdue University, West Lafayette, Indiana, and **Kunwoo Park** and **David Rempel**, University of California at Berkeley

Objective: The purpose of this study was to develop a lexicon for 3-D hand gestures for common human–computer interaction (HCI) tasks by considering usability and effort ratings.

Background: Recent technologies create an opportunity for developing a free-form 3-D hand gesture lexicon for HCI.

Method: Subjects (*N* = 30) with prior experience using 2-D gestures on touch screens performed 3-D gestures of their choice for 34 common HCI tasks and rated their gestures on preference, match, ease, and effort. Videos of the 1,300 generated gestures were analyzed for gesture popularity, order, and response times. Gesture hand postures were rated by the authors on biomechanical risk and fatigue.

Results: A final task gesture set is proposed based primarily on subjective ratings and hand posture risk. The different dimensions used for evaluating task gestures were not highly correlated and, therefore, measured different properties of the task–gesture match.

Application: A method is proposed for generating a user-developed 3-D gesture lexicon for common HCIs that involves subjective ratings and a posture risk rating for minimizing arm and hand fatigue.

Keywords: HCI, human–computer interaction, gesture, fatigue, usability

Address correspondence to David Rempel, Department of Bioengineering, University of California, Berkeley, 1301 S. 46th Street, UC RFS Building 163, Richmond, CA 94804, USA; e-mail: david.rempel@ucsf.edu.

HUMAN FACTORS
Vol. 57, No. 4, June 2015, pp. 607–621
DOI: 10.1177/0018720814559307
Copyright © 2014, Human Factors and Ergonomics Society.

INTRODUCTION

Three-dimensional hand gestures for human–computer interaction (HCI) involve using hand and finger motions or postures to signal a command or task to a computer or mobile device, whereas 2-D gestures involve signaling commands by touching a sensing surface, such as a tablet or screen. Two-dimensional gestures may involve a variety of motions, such as flicking to change windows or spreading fingertips to indicate enlarge, but are limited to the sensing of fingertips on a surface. By contrast, 3-D hand gestures offer a greater set of motions limited primarily by the image-capture technology. Recent motion-capture technology, especially high-resolution sensors designed to capture just hand gestures, is expected to lead to an increase of 3-D hand gesture input systems and gesture languages.

Such 3-D hand gesture recognition systems provide an opportunity to enhance or even bypass the keyboard, mouse, and touch screens. Other advantages of 3-D hand gesture systems are that they can be used in a sterile noncontact environment, like operating rooms (Wachs et al., 2008); allow interactions with multiple people simultaneously; can facilitate multicultural interaction when language is a barrier; allow biometrics screening; enable sign language–based communication; allow technology to be secure and prevent vandalism; and improve manipulation of objects with large 3-D screens.

The use of 3-D hand gestures for HCI has the potential to create a more intuitive, creative, and productive experience than traditional input devices or 2-D gesture input (Ni, Bowman, North, & McMahan, 2011; Wachs, Kolsch, Stern, & Edan, 2011). It also has the potential to make HCI more comfortable. Although some research has been done in this area (Farhadi-Niaki, Etemad, & Arya, 2013; Nacenta, Kamber,

Downloaded from hfs.sagepub.com at Stephen B. Thacker CDC Library on October 21, 2015

Qiang, & Kristensson, 2013; Vatavu, 2012), there is still a lack of analytical and objective methods for the optimal design of gestures for HCI (Wachs et al., 2008). Most research has involved evaluating researcher- or user-defined 2-D gestures or the hardware and software that enable 3-D gesturing interaction (Stern, Wachs, & Edan, 2006). The studies that have involved examining desktop gesturing for HCI deal with gestures for mouse or stylus-type input, not for higher-level commands and tasks (Smith & Schraefel, 2004). Researchers who have evaluated gesturing for interaction with computers have not considered the ergonomics of gestures and the physical fatigue and discomfort associated with prolonged gesturing. This is an important concern because repeated awkward hand and arm postures can lead to pain (Nielsen, Storring, Moeslund, & Granum, 2004; Rempel, Lee, & Camilleri, 2014), and unfortunately, gestures with awkward postures have been recommended for common tasks (Ni, 2011).

There are limited studies on the optimal design of 3-D hand gesture languages for HCI (Baudel & Beaudouin-Lafon, 1993; Nielsen et al., 2004). De la Barré, Chojecki, Leiner, Mühlbach, and Ruschin (2009) compared two simple select tasks. One task was holding a stationary point, and the second was tapping in toward the screen while pointing. The stationary method required less learning time, had less positioning error, required a smaller button activation size, and was associated with better usability. Another study demonstrated that precision for pointing is reduced with 3-D hand gestures compared to using a touch screen or a mouse for pointing (Vogel & Balakrishnan, 2005). This problem with 3-D gestures may be due to the lack of haptic feedback or reduced hand stability due to lack of arm support.

Mathematical models have been developed to evaluate the psychophysiologic and technical performance of gesture control systems based on posture, but the biomechanical models used were primitive (Stern et al., 2006). Menu selection interfaces have been designed for gesture interaction (Bowman & Wingrave, 2001; Ni, 2011); however, comfort and ergonomic design were not considered. The gesture set was arbitrarily decided. Karam and Schraefel (2005) developed gestures for controlling background music with the commands *previous track, next track*, and *stop*. The gesture language was created by the researchers and derived from user studies. Their final selected gestures were a left-to-right hand wave for playing the next piece, right-to-left hand wave for playing the previous piece, and an open palm facing the screen (e.g., halt gesture) for stopping playback. A simple gesture interface for navigation and entertainment system for cars was developed and tested by Alpern and Minardo (2003). Iterative cycles of testing and thinking out loud were used to refine the gesture language and onscreen menu. The final gesture lexicon was 8 symbolic gestures and number (1 through 5) and directional (up, down, left, and right) gestures. The interface was reported to have fewer errors and was preferred over the traditional touch interface. Other studies have engaged end users to suggest and evaluate the best 2-D gestures for common touch screen tasks, such as panning and zooming (Fikkert, Vet, Veer, & Nijholt, 2010; Wobbrock, Morris, & Wilson, 2009; Wright, Lin, O'Neill, Cosker, & Johnson, 2011).

Developers of an HCI gesture lexicon, that is, the assignment of specific gestures to HCI commands or tasks, need to consider the user's expectations of computer interpretation (e.g., natural language) in order to minimize learning time, maximize cultural transparency, and optimize usability and productivity. Developers of the gesture lexicon should also consider the fatigue or discomfort associated with a gesture. Common tasks should be matched to gestures that are rapid and comfortable to form with minimal biomechanical and ergonomic risk (Wachs et al., 2008). In addition, the gesture lexicon should work across a variety of platforms so that it can be widely adopted. Already existing widely used 2-D touch gestures should also be considered.

Developers of a gesture lexicon should also consider cognitive load. Given that a user can remember five plus or minus two items in short-term memory (Miller, 1956), the number and type of gestures and tasks should be carefully selected. The mental workload for a lexicon for a large number of tasks can be reduced by reusing gestures (similar to switching mode). For

Downloaded from hfs.sagepub.com at Stephen B. Thacker CDC Library on October 21, 2015

example, the same gesture can have two different meanings, depending on the context (Kaiser et al., 2003). Reuse of gestures allows for a larger set of tasks with fewer defined gestures and reduces the complexity (Wu & Balakrishnan, 2003). The chance of misrecognition increases with the number of gestures. Wobbrock et al. (2009) also considered consistency and symmetry in the development of a 2-D gesture set for touch surface interaction. *Symmetry* refers to reversible gestures for opposite tasks, such as opening the hand or closing the hand for enlarging or shrinking tasks.

In addition to cognitive load, designers of a gesture lexicon need to consider the visibility and specificity of the gesture posture. The selected gestures should be distinctive, but there will be some degree of variability in the gesture shape between and within people when they perform a specific gesture (to facilitate pattern recognition and cognitive association). Differences in the duration and shape of a gesture are referred to as the spatiotemporal variability (Keskin, Erkan, & Akarun, 2003). In addition, many sensors, such as the Kinect (Zhang, 2012) and Leap Motion (2013), use a single camera view for gesture capture. Therefore, the distinctive features of a gesture should be visible from a common sensor angle (Kaiser et al., 2003).

The ultimate design of common gesture lexicons should follow natural language principles (e.g., easy to learn, easy to remember, easy to perform; Baudel & Beaudouin-Lafon, 1993; Nielsen et al., 2004). However, given past experience, these gestures are unlikely to be guided by knowledge of hand postures that are comfortable or that follow ergonomic principles. Instead, they are likely to be guided by the system's gesture recognition accuracy (also referred to as technical principles). Previous researchers have proposed to solve these conflicting goals (principles) using multiobjective optimization models (Wachs, 2006). In such cases, weights are assigned to those goals that are more important from the user's standpoint.

Repeatedly forming hand gestures may be associated with hand pain, depending on the duration of gesturing, the gesture movement patterns, and the gesture postures, based on epidemiologic and physiologic studies of the hand and studies of sign language interpreters (Bao, Howard, Spielholz, Silverstein, & Polissar, 2009; Hignett & McAtamney, 2000; Keir, Bach, & Rempel, 1998; Rempel et al., 2014). Sign language interpreters have an extensive and unique experience forming repeated hand gestures, and many suffer from hand or arm pain after gesturing for several hours (Feuerstein, Carosella, Burrell, Marshall, & Decaro, 1997; Rempel et al., 2014). This finding is especially alarming when considering the possibility that computer users may be performing gestures many hours per day (leading to the phenomenon called "gorilla arm").

Studies of sign language interpreters have identified particular hand and arm gestures that, if repeated, are comfortable and associated with little muscular stress, whereas other specific gestures can be identified as uncomfortable or even painful. Gestures that included extreme wrist flexion and asynchronous finger postures are especially uncomfortable, as are those with extreme wrist extension and forearm rotations (Rempel et al., 2014). Over time, these short-term uncomfortable or painful symptoms can develop into prolonged pain and impairment (Webster & Snook, 1994). The reason to study sign language interpreters is not to recommend sign language for an HCI lexicon but to identify the basic hand postures and motions associated with discomfort and fatigue in order to design 3-D gestures for common HCI tasks that are not fatiguing.

The purpose of this study was to create a 3-D hand gesture set for specific computer tasks guided by user-generated gestures, with the final selection based on user ratings, estimation of postural risk, and consideration of system capabilities. Thirty subjects developed gestures for 34 common HCI tasks. During the study, users were prompted to form one or more gestures that first come to mind to presented tasks. They also rated gestures on preference, match, easiness, and effort. Videos of the gestures formed were analyzed for gesture type, response time, and order. Results of user preference, match, easiness, and effort order in addition to gesture popularity and posture score were integrated into a single overall score that was used to develop a task gesture lexicon.

Downloaded from hfs.sagepub.com at Stephen B. Thacker CDC Library on October 21, 2015

TABLE 1: The 34 Tasks or Commands for Which Subjects Chose Gestures and Mean Investigator-Rated Conceptual Complexity (1 = *simple*, 5 = *complex*)

Task	*M*	*SD*	Task	*M*	*SD*
1. Move	1.00	0.00	18. Minimize	3.67	0.58
2. Select single	1.00	0.00	19. Cut	3.67	0.58
3. Rotate	1.33	0.58	20. Accept	4.00	1.00
4. Shrink	1.33	0.58	21. Reject	4.00	1.00
5. Delete	1.33	0.58	22. Menu access	4.33	0.58
6. Enlarge	1.33	0.58	23. Help	4.33	0.58
7. Pan	1.67	0.58	24. Task switch	4.67	0.58
8. Close	2.00	0.00	25. Undo	5.00	0.00
9. Zoom in	2.00	0.00	26. Gesture on	1.67	0.58
10. Zoom out	2.00	0.00	27. Gesture off	1.67	0.58
11. Select group	2.33	0.58	28. Volume up	2.67	0.58
12. Open	2.33	0.58	29. Volume down	2.67	0.58
13. Duplicate	2.67	1.53	30. Mute	2.33	0.58
14. Previous	3.00	0.00	31. Save	3.00	1.00
15. Next	3.00	0.00	32. New	4.00	1.00
16. Insert	3.33	0.58	33. Find	4.33	0.58
17. Paste	3.33	1.15	34. Control cursor	2.67	0.58

METHOD

In this laboratory study, 30 subjects performed self-derived gestures for 34 common computer tasks (e.g., move, select single, enlarge). Subjects were presented with images of the tasks, then performed the gesture(s) of their choice that they thought would be the best ones to execute the task. The study was approved by the university institutional review board, and subjects signed a consent form.

Subjects

Eligibility criteria were age between 18 to 65 years, being right hand dominant, experience with touch screen devices, and no history of upper-extremity musculoskeletal disorders in the past 6 months. Of the 30 subjects, 14 were female and the average age was 33 (±13). All subjects had some familiarity with 2-D or 3-D gesturing (e.g., Apple iOS, Nintendo Wii, Android, and Microsoft Kinect).

Tasks

The tasks or commands evaluated were selected from common tasks used for Microsoft Office (Table 1). Tasks 1 to 25 were previously examined in Wobbrock et al. (2009) for 2-D gesture input, and we added Tasks 26 to 34. The added tasks are common tasks used in the Windows operating system. The tasks were rated for conceptual complexity by three investigators (AP, JW, DR) using a scoring system (1 = *simple* to 5 = *complex*) similar to that of Wobbrock et al. Menu interactions were not considered because important changes to menu layout are expected in future operating systems to better accommodate gesture input (Microsoft Windows 8; Ni, 2011).

Each task was displayed on the monitor, typically as before-and-after images of the task. For example, *move* images are displayed in Figure 1; the subject is to make a gesture that will move the picture of the puppy from the upper left to the lower right on the screen. Once the subject understood the task and the desired output, the initial image was shown again, and the subject demonstrated the gestures he or she would make to complete the task. Participants were encouraged to make at least two different gestures for a task. A random number generator was used to assign test order of tasks.

Downloaded from hfs.sagepub.com at Stephen B. Thacker CDC Library on October 21, 2015

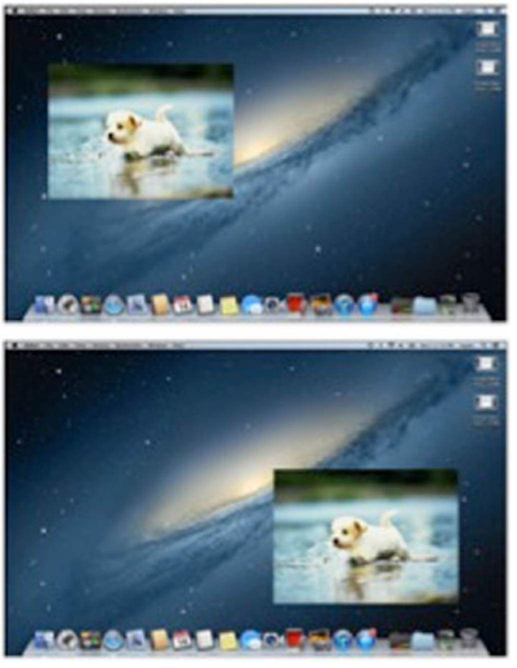

Figure 1. Example before-and-after images for the *move* tasks. The subjects proposed a gesture that moved the picture of the puppy from the upper left to the lower right on their screen.

The tasks used screen shots from Apple OS X and Microsoft 8. When appropriate, mouse shortcut cues (e.g., close button on window) were hidden to force the use of a gesture rather than pointing at a button.

Experimental Setup

The subjects were provided an adjustable chair, desk, and monitor. The work surface was adjustable in height and set approximately 2 cm below elbow height. The monitor (50 cm diagonal) location was adjusted to place the top near the subject's eye level and at an arm's length away. A keyboard and mouse were on the desk surface between the monitor and the subject but were not used.

Four video cameras were used to simultaneously record the experiment (Swann Security System, SWDVK-414002, Santa Fe Springs, CA). The first camera's view was the subject's hands and arms from the right side, the second camera's view was the subject's hands from above, the third camera's view was the interviewer's hands from the side, and the fourth camera's view was the screen. The video images were later used by research technicians to classify the different gestures generated and to count them for each task.

The before-and-after images of each task were shown while the task description was spoken by the researcher. Subjects were instructed to make the gesture that they thought best represented the task. They could make several different gestures. While making the gesture, participants were encouraged to think aloud and explain the gesture they were making. After the gestures were formed, the researcher asked open-ended, scripted questions about why they selected a specific gesture. Questions were nonleading, neutral, and in the present tense.

Subjective Ratings: Preference, Match, Ease, and Effort

After all gestures were completed for a task, subjects ranked or rated their gestures on four dimensions: preference, match, ease, and effort. First, they selected their first and second most preferred gesture. Then, using a Likert scale (1 = *low* to 7 = *high*), they rated their top two preferred gestures on two statements: "The gesture is a good match for its intended purpose" and "The gesture is easy to perform." Subjects then rank ordered all of their proposed gestures, starting with "the gesture that requires the least amount of effort." These four dimensions—preference, match, ease, and effort—were considered to be independent qualities of the gestures. Examples of gestures that are difficult to perform are found among pianists who learn complex gestures as part of their musical apprenticeship.

Gesture Classification, Popularity, Order, and Response Time

The videotapes of all of the gestures generated were analyzed by two technicians to categorize gestures that were essentially the same, record gesture order, and measure the response time. Gestures were considered to be the same if they produced the same motions or postures. Motion primarily comprised the direction of joint motion and synchronization of fingers, wrists, and forearm movements. Posture comprised the end postures for the fingers, wrists, and forearm.

Downloaded from hfs.sagepub.com at Stephen B. Thacker CDC Library on October 21, 2015

TABLE 2: Gesture Posture Scores

Gesture Posture	Score
Finger (metacarpophalangeal joint) extension	
Not full	1
Full	3
Angle of separation between two most extreme fingers	
0° to 14°	1
15° to 45°	2
>45°	3
Angle of extension/flexion of the wrist from neutral	
0° to 14°	1
15° to 45°	2
>45°	3
Angle of ulnar/radial deviation of the wrist from neutral	
0° to 9°	1
10° to 20°	2
>20°	3
Angle of forearm pronation/supination from neutral	
<9°	1
10° to 45°	2
>45°	3
External rotation of shoulder	
<0°	1
0° to 30°	2
>30°	3

Note. Higher scores represent higher risk for fatigue or musculoskeletal disorder.

Gesture popularity was the number of subjects who generated the same gesture for a task, independent of order. However, if the same gesture was performed by only one or two subjects, it was not counted. In the few cases in which the technicians disagreed on category, one of the authors (AP) made a final selection. The order of appearance of each gesture for each subject (e.g., first, second, third) was recorded. Response time was the time in seconds from when the researcher completed the explanation of the task to the subject's first gesture response. The terms for these three dimensions are *popularity, order*, and *response time*.

Gesture Posture Ratings

Gestures were assigned a biomechanical risk score based on the most extreme joint postures created during the gesture. The overall posture score for a gesture was calculated as the sum of the scores from Table 2. The posture score ratings (higher values represent increased risk) are based on existing risk assessment tools and studies of physiologic and epidemiologic risk linking postures to high biomechanical loads, fatigue, pain, or musculoskeletal disorders (Bao et al., 2009; Hignett & McAtamney, 2000; Keir et al., 1998; Rempel, Bach, Gordon, & So, 1998).

Overall Score and Agreement Score

An overall score for task–gesture combinations was calculated by combining the values of the six variables: preference, match, ease, effort, gesture popularity, and posture. Before summation, the six variables were each normalized to a mean value of 10 and a standard deviation of 1 and adjusted so that high values were positive traits. The overall score was the sum of the six normalized variables.

Downloaded from hfs.sagepub.com at Stephen B. Thacker CDC Library on October 21, 2015

To compare our findings to prior analyses in the literature, an agreement score, A, using gesture popularity, was also calculated (Wobbrock et al., 2009).

$$A = \frac{\sum_{r \subset R} \sum_{Pi \subseteq Pr} \left(\frac{|Pi|}{|Pr|}\right)^2}{|R|}, \qquad (1)$$

where r is a referent task in the set of all tasks R, Pr is the set of proposed gestures for referent task r, and Pi is the popularity of the subset i of identical gestures from Pr. The range for A is from 0 to 1.

An example of the agreement score calculation for the *accept* command is shown in Equation (2).

$$A = \left(\frac{19}{60}\right)^2 + \left(\frac{12}{60}\right)^2 + \left(\frac{10}{60}\right)^2 + 19 * \left(\frac{1}{60}\right)^2 = 0.43. \quad (2)$$

A total of 60 different gestures were proposed by subjects for *accept*. The popularity scores for the top three gestures were 19, 12, and 10. In addition, 19 other gestures were proposed by only one subject and were, therefore, not tallied on popularity. The agreement score, which reflects the degree of consensus among participants, was calculated for all tasks and gestures.

Selection of Task Gesture Set

The selection of a final task gesture set was primarily driven by the overall score with adjustment, if necessary, to avoid repeated gestures and to reduce cognitive load. In addition, for the final set, we considered the potential effectiveness of gesture capture by a single camera system based on gesture distinctiveness and shadowing. Gesture shadowing occurs when an important gesture feature is hidden from a camera by the hand or arm. Gesture distinctiveness is based on selecting hand gestures that are different enough that they would be readily distinguished from each other by an image-capture system.

RESULTS

Overall, subjects generated over 1,300 different gestures for the 34 tasks. Only the 84 gestures that were selected by three or more subjects were analyzed further. The 84 gestures were associated with the 34 tasks, with a total of 160 different task–gesture combinations. Agreement scores for the tasks are plotted in descending order (solid line) in Figure 2. Values are between 0 and 1, with 1 being *complete agreement*. The average response time to first gesture for a task is plotted with the gray bars. There is a moderate inverse correlation between agreement score and response time ($r = -.59$).

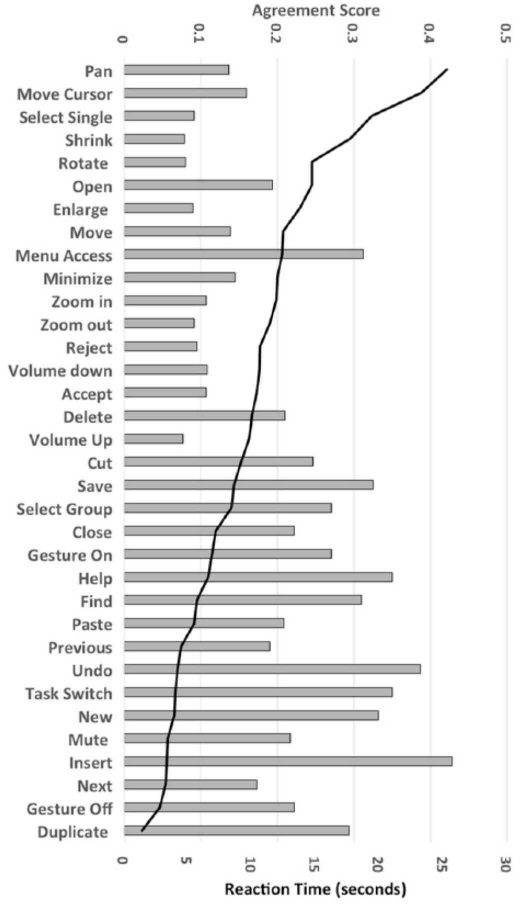

Figure 2. Agreement score (solid line) and mean response time to first gesture (gray bar) for each task.

It is beyond the scope of the paper to present all 1,300 gestures or all 160 task–gesture combinations. Instead, a gesture set using just 13 gestures is proposed for the 34 tasks based on overall score. The 13 gestures are presented in Figure

Downloaded from hfs.sagepub.com at Stephen B. Thacker CDC Library on October 21, 2015

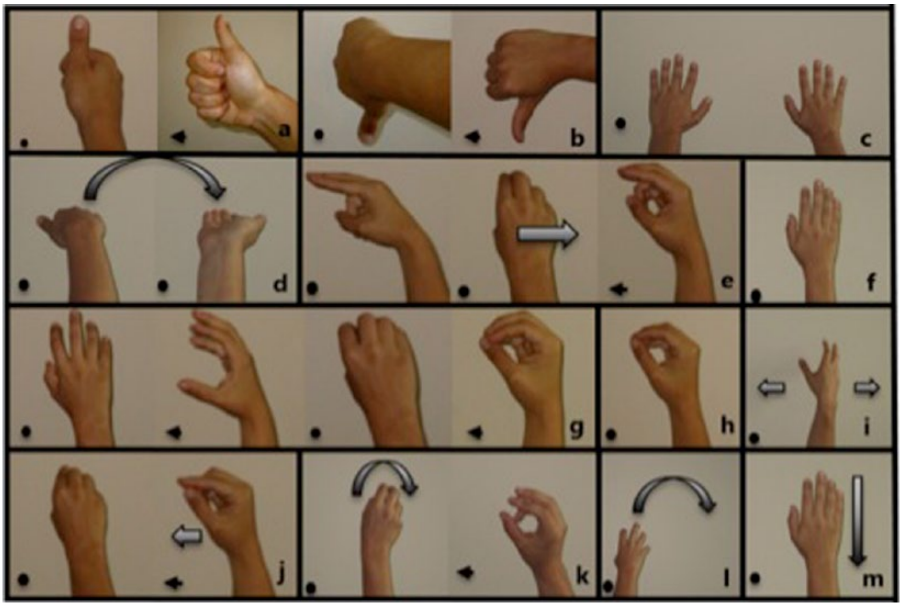

Figure 3. Proposed 13 primary gestures. These gestures are linked to tasks in Table 3 by the lowercase letters. Some gestures are more complex than shown here but are described in more detail in Table 3. Gray arrows indicate the direction of hand motion. Black arrows point toward the screen, which is to the left, and round black dots are arrows pointing toward the screen away from the reader.

3. Some tasks are indicated by the same gesture, in which case, task selection would be context sensitive. The proposed gesture task lexicon and the nine dimensions on which different aspects of matching are rated for each task–gesture combination are presented in Table 3. For example, the *accept* and *save* commands are both assigned the thumbs-up gesture (Figure 3a), and their overall ratings are 65 and 63, respectively.

In Table 3, the number of different gestures considered for each task is listed in the column "Count"; the range is 2 to 8. Each of these unique gestures had to be proposed for the task by at least three participants. For 22 of the 34 tasks, the final selected gesture was readily identified based on the highest overall score. For the 12 other tasks (marked with an asterisk in Table 3), there was some problem with using the gesture with the highest overall score.

For example, the highest-scored gestures for four of the tasks (i.e., *cut, paste, duplicate*, and *insert*) would not have been visible to existing gesture-capture systems. To highlight the problem further, the highest-scored gesture for *paste*, a motioning of pressing down glued paper across the palm of the left hand, may have been obscured from a camera by the left hand.

Recommended gestures for *duplicate* and *insert* had overall scores of zero and very low agreement scores, 0.02 and 0.06, respectively. The highest overall scored gesture for *insert* was point and draw a caret; however, disambiguation between pointing and drawing a caret and just pointing led to the selection of the cut-and-paste gesture for *insert*. The highest overall score for *duplicate* was a combination of three gestures: grab, pull toward user, and place back on screen. However, disambiguation with the grab-and-manipulate gesture was a concern.

The highest overall score for *enlarge* was an opening-pinch gesture. However, because the highest overall scored gestures for the tasks *shrink, zoom in*, and *zoom out* were to grab and either close or open grip, and grabbing and opening grip was still highly rated for *enlarge*, the grab-and-open gesture was also selected for *enlarge*. If resources exist, designers may consider developing both the grab and pinch

Downloaded from hfs.sagepub.com at Stephen B. Thacker CDC Library on October 21, 2015

TABLE 3: Proposed Set of 34 Task Gestures With Ratings on Nine Dimensions

Task	Figure 3	Gesture	Count	Gesture Agree-ment	Task Agree-ment	Prefer-ence	Match	Effort	Ease	Posture	Popular-ity	Overall
Accept	a	Thumbs-up	3	0.43	0.17	10.2	11.2	11.1	10.4	10.8	11.7	65.0
Save	a	Thumbs-up	4	0.43	0.18	10.0	10.5	11.9	11.2	10.8	9.2	63.0
Reject	b	Thumbs-down	3	0.73	0.14	10.5	11.1	10.8	10.4	8.1	11.3	62.0
Gesture on	c	Both palms toward screen, no movement	6	0.38	0.11	10.7	10.0	10.2	10.5	10.2	11.6	63.0
Gesture off	c	Both palms towards screen, no movement	7	0.38	0.05	9.2	10.5	9.9	10.4	10.2	9.7	59.0*
New	d	Flip over, making arch left to right	6	1.00	0.07	12.0	8.3	11.9	10.5	10.2	9.2	62.0
Cut	e	Cutting scissors movement (two fingers)	6	0.63	0.15	10.5	10.0	9.9	9.3	9.6	11.6	60.0*
Paste	e	Pressing down glued paper (two fingers)	4	1.00	0.09	10.6	9.7	9.6	10.4	11.1	10.5	61.0*
Insert	e	Finger scissor movement, then press down	3	0.00	0.02	0.0	0.0	0.0	0.0	0.0	0.0	0.0*
Duplicate	e	Finger scissor movement, then press down	2	0.00	0.06	0.0	0.0	0.0	0.0	0.0	0.0	0.0*
Help	f	Raise one hand	5	1.00	0.11	11.5	9.8	11.1	10.1	11.1	9.8	63.0
Enlarge	g	Grab and open grip to scale	6	0.30	0.23	9.1	10.2	9.6	9.7	10.8	11.4	60.0*
Shrink	g	Grab and close grip to scale	5	0.32	0.29	10.0	10.4	9.9	10.4	10.8	14.0	65.0
Zoom in	g	Grab and open grip to scale	7	0.30	0.20	11.4	10.9	10.5	10.5	10.8	12.0	66.0
Zoom out	g	Grab and close grip to scale	5	0.32	0.19	10.3	10.4	10.5	10.4	10.8	12.2	64.0
Close	g	Grab and close grip entirely	7	0.32	0.12	11.2	11.2	9.9	11.0	10.8	9.8	63.0
Rotate	g	Grab and rotate	7	1.00	0.25	10.2	10.9	10.8	10.2	10.2	12.2	64.0
Move	g	Grab, pull from screen, move, push to screen, and release	5	0.57	0.21	9.4	10.7	9.6	10.1	10.8	11.7	62.0*
Delete	g	Grab, pull from screen, place off screen, and release	5	1.00	0.17	10.1	10.4	8.5	9.6	10.8	10.1	59.0*

(continued)

Downloaded from hfs.sagepub.com at Stephen B. Thacker CDC Library on October 21, 2015

TABLE 3: (continued)

Task	Figure 3	Gesture	Count	Gesture Agree-ment	Task Agree-ment	Prefer-ence	Match	Effort	Ease	Posture	Popular-ity	Overall
Find	h	Hand shaped like telescope	6	1.00	0.09	10.5	8.4	10.5	9.1	11.1	10.9	60.0*
Pan	i	Hand flat with thumb on top and moved left and right	2	0.16	0.42	11.1	11.4	10.8	11.0	10.8	13.6	68.0
Task switch	i	Hand flat with thumb on top and moved left and right	3	0.16	0.07	10.6	10.4	10.8	10.9	10.8	10.5	63.0
Control cursor	j	Point	4	0.15	0.39	11.0	10.9	11.1	10.2	9.6	13.2	65.0
Select single	j	Point and tap in, thumb down	3	0.15	0.32	11.0	11.1	11.9	11.2	9.6	13.0	67.0
Select group	j	Point and tap in, thumb up	4	0.15	0.14	11.2	10.4	11.1	9.9	9.6	11.6	63.0
Open	j	Point and tap in twice	3	0.15	0.24	10.6	10.9	10.8	11.2	9.6	12.7	65.0
Menu access	j	Point and gesture at top of screen	5	1.00	0.21	9.9	9.3	10.8	10.2	10.2	12.0	62.0
Volume up	k	Turn knob clockwise	5	0.16	0.05	10.5	9.7	11.1	11.3	10.8	9.3	62.0
Volume down	k	Turn knob counterclockwise	5	0.38	0.07	10.2	10.5	10.2	10.7	9.3	9.5	61.0*
Mute	k	Turn knob counterclockwise and bring fingertips together	8	0.38	0.07	10.5	9.7	11.1	9.9	9.3	10.1	60.0*
Next	l	Palm down rotated to palm up	5	0.16	0.05	10.5	9.7	11.1	11.3	10.8	9.3	62.0
Previous	l	Palm up rotated to palm down	4	0.38	0.07	10.2	10.5	10.2	10.7	9.3	9.5	60.0*
Undo	l	Palm up rotated to palm down	3	0.38	0.07	10.5	9.7	11.1	9.9	9.3	10.1	60.0
Minimize	m	Pan hand down	4	0.38	0.2	10.2	10.5	10.5	10.1	11.1	12.2	64.0

Note. Count is the number of different gestures considered for a task; each of these gestures was proposed by at least three participants. An asterisk (*) denotes a problem with using the gesture with the highest overall score; see text.

Downloaded from hfs.sagepub.com at Stephen B. Thacker CDC Library on October 21, 2015

gestures for these four tasks. Linking multiple gestures to a single task has been previously proposed and may be the optimal solution, but it increases the gestures that the system must recognize and may increase users' cognitive load (Wobbrock et al., 2009).

The highest overall scored gesture for *mute* was not selected. This gesture was pinching of all fingers together while the fingertips were horizontal, commonly known as a "be quiet" gesture. However, because the selected gesture for *volume down*, turning knob counterclockwise, was highly rated and did not add to user learning effort and memory, it was selected for *mute*.

For *delete*, the highest overall scored gesture was panning. However, the panning gesture may be difficult to distinguish from panning and task switch. Therefore, another highly rated gesture—to grab, move off screen, and release—which did not increase user learning and memory, was selected for *delete*.

For *find*, the highest-rated gesture was placing a flat hand above the eyes, such as when shielding one's eyes from overhead sun. However, system recognition was a concern, so the hand-shaped-like-telescope gesture was selected. Again, if systems are able to interpret the hand-over-the-eyes gesture, designers may want to recognize both gestures for this task.

For the task *previous*, the highest overall rated gesture, turning a knob counterclockwise, was not selected because it was used for *mute* and *volume down*. Pointing in toward the screen was also highly rated, but it was not selected because of difficultly disambiguating from pointing to control cursor and selecting.

For the task *gesture off*, the highest-rated gesture was grabbing and closing the hand, followed by holding both palms toward the screen with no movement. The gesture of holding palms toward the screen was selected because it was more easily disambiguated from the grab-and-manipulate gesture.

A post hoc analysis was conducted of the correlations between the different rating dimensions for the proposed task gesture set (Table 4). In general, subjective dimensions were inversely related to response time to first gesture. That is, the higher the rating on popularity, preference, match, and task agreement, the shorter the initial response time to first gesture. Also, the response time to first gesture was moderately correlated (r = .57) with complexity. There was little correlation between response time and ease or gesture agreement. Preference was correlated with gesture order (.52) and effort (.54). Gesture order and effort were also correlated (.56).

DISCUSSION AND CONCLUSIONS

A gesture set for 34 common computer tasks was developed based on a task–gesture overall score that combined subject ratings of preference, ease, match, effort, and popularity and investigator ratings of posture risk for fatigue. The final selection of gestures for tasks was also guided by reducing gesture-capture conflicts and minimizing cognitive load. Several tasks were assigned the same gesture; therefore, a context-based selection process would be required.

In general, the six dimensions used for calculating the overall score (e.g., preference, match, ease, effort, gesture popularity, and posture) were not highly correlated; therefore, they measured different properties of the task–gesture match. Low correlation is a desirable feature since it allows the overall score to express the strengths and weaknesses of each dimension in a combined fashion, with little redundancy. Most of these dimensions were inversely correlated with response time to first gesture, especially task agreement (r = –.59). Therefore, in general, the less time it takes for users to propose a gesture for a task, the more likely that many subjects would pick the same gesture for the task. Tasks that required greater time for subjects to propose a gesture were *next, mute, previous, paste*, and *menu access*. The selected gestures for three of these tasks (*paste, mute*, and *previous*) were not based on the overall rating. This result suggests that other factors, such as cognitive load, should be considered when the response time is long and agreement is low. It has been shown that response time is commonly used for cognitive complexity, so those gestures that required less time are also less cognitively burdening (Horsky, Kaufman, Oppenheim, & Patel, 2003).

Interestingly, a number of users used both hands for gesturing and used a left fist to assign the same right-hand gesture to different tasks.

Downloaded from hfs.sagepub.com at Stephen B. Thacker CDC Library on October 21, 2015

TABLE 4: Correlation Matrix (*r*) for Various Dimensions of the Proposed Task Gesture Set

	Popularity	Preference	Match	Ease	Effort	Posture	Gesture agreement	Task agreement	Response time	Gesture order	Complexity	Overall Score
Popularity	—											
Preference	.14	—										
Match	.21	.22	—									
Ease	.12	.22	.20	—								
Effort	.14	.54	.09	.14	—							
Posture	.11	.06	.00	.04	.10	—						
Gesture agreement	−.12	−.11	.00	−.18	−.22	−.11	—					
Task agreement	.38	.05	.23	.10	.01	.01	.05	—				
Response time	−.26	−.14	−.27	−.04	−.22	−.12	.00	−.59	—			
Gesture order	.11	.52	.19	−.13	.56	−.08	.05	.20	−.31	—		
Complexity	−.10	−.18	−.08	−.11	−.26	.03	.05	−.36	.57	−.28	—	
Overall score	.30	.38	.31	.31	.43	.20	.07	.22	.02	−.48	.17	—

For example, a left fist while pointing in and out with the right hand would mean *duplicate*, whereas without the left fist meant *open*. All subjects in the study were right hand dominant. Designers may want to capture the left-hand gesture to allow for multiple meanings of the same right-hand gesture. This design allows the left hand to act as a "modifier" gesture. This finding is consistent with the three principles reported (Guiard, 1987; Guiard & Ferrand, 1996) about the asymmetric division of labor between hands. In this case, the nondominant hand (left) establishes a new frame of reference for the right hand by performing the fist gesture. In addition, as previously recommended, current menu options can be used to disambiguate gestures (Alpern & Minardo, 2003). Alpern & Minardo (2003) limited the gesture lexicon to 8 symbolic gestures and simple directional (up, down, left, and right) and numeric gestures (1 through 5), which were used while driving a car during menu selections or on screen interfaces, such as volume control.

Our study primarily involved subjective dimensions of preference, popularity, match, ease, and effort plus investigator posture ratings to create a task gesture set guided by a systematic and analytic approach with consideration of other factors. Previous studies have involved subsets of these dimensions and, to some extent, have overlapping findings. De la Barré et al. (2009) also selected pointing with the finger for the *select* task. However, on the basis of our methods, pointing in toward the screen (pointing and moving the finger toward the screen) was used instead of an extended stationary pointing. More users performed pointing in than they performed stationary pointing, and pointing in had higher overall ratings. However, it is possible that with usability testing with a functional system, stationary pointing may provide a better user experience. But usability is likely to depend on the performance of the gesture recognition system. Karam and Schraefel (2005) and Alpern and Minardo (2003) examined gesture sets for music control. Similar to our results, Karam and Schraefel selected a left-to-right hand wave for *next piece* and a right-to-left hand wave for *previous piece*. Though the specifics of the gestures differed, many subjects used a left-to-right hand movement for *next* and a right-to-left hand movement for *previous*.

Wobbrock et al. (2009) used a similar interview method to develop a touch surface (2-D) gesture set. Many of Wobbrock et al.'s proposed

Downloaded from hfs.sagepub.com at Stephen B. Thacker CDC Library on October 21, 2015

task gestures were similar to our set. For example, we propose *minimize* as panning downward with an open hand, and Wobbrock et al. proposed pointing and dragging the object downward. Our *minimize* gesture is optimized for a 3-D interaction space and posture. Almost all of the Wobbrock et al. proposed gestures involved one finger pointing. Our task gesture set includes only four gestures with one finger pointing. The differences are likely due to the ability of 3-D systems to capture an increased number of different hand postures compared to a 2-D touch screen and our inclusion of a posture score in gesture selection. Wobbrock et al. did not consider postural risk for fatigue for gesture selection. A number of the gestures we proposed, such as thumbs-up and thumbs-down for *accept* and *reject*, would not have been possible for a 2-D touch screen interaction. Interestingly, Wobbrock et al. proposed use of drawing a check mark and *X* for *accept* and *reject*; these were the second-highest-scored gestures for these tasks in our study. In spite of the differences in capabilities between 2-D and 3-D systems, the proposed task gesture sets are surprisingly similar.

In our approach to developing a task gesture language, we considered user cognitive load and practical limitations of 3-D gesture recognition capabilities. Similar to previous studies, we attempted to limit cognitive load through consistency and symmetry in order to reduce learning and memory demands (Alpern & Minardo, 2003; Kaiser et al., 2003; Wobbrock et al., 2009; Wu & Balakrishnan, 2003). Our gestures are meant to be distinguished from a single camera angle. The gestures are also meant to be distinctive, allowing for recognition of differences in spatiotemporal variability (Keskin et al., 2003).

Several limitations of the study should be noted. First, the subjects were male and female English speakers from Northern California. Computer users from other countries or cultures may propose a different set of task gestures and may assign different subjective ratings along the dimensions we tested. Second, the gestures were proposed by subjects without the aid of a functional gesture recognition system. This approach is preferred initially because a functional system would have required training, and the system limitations would have constrained the gestures that subjects could propose. On the other hand, our approach permitted subjects to propose some gestures that are unlikely to be interpretable by gesture recognition systems, and these gestures were rejected on review. Third, in order to best use subjects' time during the experiment, the ratings of preference, match, and easiness were collected only for the two most preferred gestures for a task. This method may have prevented some worthwhile gestures from being considered for tasks. The overall score, which drove the gesture selection process, was correlated with the dimensions of preference, effort, and gesture order, but the correlations were not high (Table 4). Using a different dimension to drive selection would have led to a different selected lexicon. For example, using just response time would have led to a gesture set more correlated with task agreement and complexity. However, this method would have reduced correlation with preference, effort, and posture. We considered the six dimensions included in the overall score as the most important, but others may disagree.

In conclusion, we present an approach for developing a 3-D hand gesture lexicon for common HCI tasks that incorporates previously used subjective factors plus a new investigator rating of hand posture for estimating risk of fatigue. The method led to a 3-D task gesture lexicon with 13 primary gestures assigned to 34 tasks. The proposed lexicon should be tested against other lexicons, especially designer-developed lexicons that also take into account the limitations of image-capture technology. These comparisons should consider productivity, usability, and upper-extremity fatigue. Also, 3-D hand gesture input should be compared to the traditional method of HCI using a mouse, touch pad, or keyboard.

ACKNOWLEDGMENTS

This study was funded in part by a graduate training grant from the National Institute for Occupational Safety and Health in the U.S. Centers for Disease Communication (T42 OH008429).

KEY POINTS

- New technologies are expanding the options for human–computer interaction to include 3-D hand gestures.

Downloaded from hfs.sagepub.com at Stephen B. Thacker CDC Library on October 21, 2015

- A user-centered process was used to evaluate the fit between 34 common human–computer tasks and self-generated hand gestures. The process involved 11 different dimensions of gesture–command match, both cognitive and biomechanical.
- Based on an overall score that incorporated six of the dimensions, a 3-D hand gesture lexicon is proposed for 34 common human–computer tasks.
- The proposed 3-D hand gesture lexicon should be compared to other lexicons and to traditional human–computer interaction methods on productivity, usability, and upper-extremity fatigue.

REFERENCES

Alpern, M., & Minardo, K. (2003). Developing a car gesture interface for use as a secondary task. In *CHI 2003 Extended Abstracts on Human Factors in Computing Systems* (pp. 932–933). New York, NY: ACM.

Bao, S., Howard, N., Spielholz, P., Silverstein, B., & Polissar, N. (2009). Interrater reliability of posture observations. *Human Factors*, *51*, 292–309.

Baudel, T., & Beaudouin-Lafon, M. (1993). Charade: Remote control of objects using freehand gestures. *Communications of the ACM*, *36*, 28–35.

Bowman, D., & Wingrave, C. (2001). Design and evaluation of menus systems for immersive virtual environments. In *Proceedings of IEEE Virtual Reality 2001* (pp. 149–156). Piscataway, NJ: IEEE.

de la Barré, R., Chojecki, P., Leiner, U., Mühlbach, L., & Ruschin, D. (2009). Touchless interaction: Novel chances and challenges. In M. Kurosu (Ed.), *Human–computer interaction: Novel interaction methods and techniques* (pp. 161–169). Berlin, Germany: Springer.

Farhadi-Niaki, F., Etemad, S. A., & Arya, A. (2013). Design and usability analysis of gesture-based control for common desktop tasks. In M. Kurosu (Ed.), *Human–computer interaction: Interaction modalities and techniques* (pp. 215–224). Berlin, Germany: Springer.

Feuerstein, M., Carosella, A. M., Burrell, L. M., Marshall, L., & Decaro, J. (1997). Occupational upper extremity symptoms in sign language interpreters: Prevalence and correlates of pain, function, and work disability. *Journal of Occupational Rehabilitation*, *7*, 187–205.

Fikkert, W., Vet, P. van der, Veer, G. van der, & Nijholt, A. (2010). Gestures for large display control. In S. Kopp & I. Wachsmuth (Eds.), *Lecture notes in computer science: Vol. 5934. Gesture in embodied communication and human–computer interaction* (pp. 245–256). Berlin, Germany: Springer.

Guiard, Y. (1987). Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of Motor Behavior*, *19*, 486–517.

Guiard, Y., & Ferrand, T. (1996). Asymmetry in bimanual skills. In D. Elliott & E. A. Roy (Eds.), *Manual asymmetries in motor performance* (pp. 175–195). Boca Raton, FL: CRC Press.

Hignett, S., & McAtamney, L. (2000). Rapid entire body assessment (REBA). *Applied Ergonomics*, *31*, 201–205.

Horsky, J., Kaufman, D. R., Oppenheim, M. I., & Patel, V. L. (2003). A framework for analyzing the cognitive complexity of computer-assisted clinical ordering. *Journal of Biomedical Informatics*, *36*(1), 4–22.

Kaiser, E., Olwal, A., McGee, D., Benko, H., Corradini, A., Li, X., & . . . Feiner, S. (2003). Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality. In *Proceedings of the 5th International Conference on Multimodal Interfaces* (pp. 12–19). New York, NY: ACM.

Karam, M., & Schraefel, M. (2005). A study on the use of semaphoric gestures to support secondary task interactions. In *Proceedings of CHI '05 Extended Abstracts on Human Factors in Computing Systems* (pp. 1961–1964). New York, NY: ACM.

Keir, P. J., Bach, J. M., & Rempel, D. M. (1998). Effects of finger posture on carpal tunnel pressure during wrist motion. *Journal of Hand Surgery*, *23*, 1004–1009.

Keskin, C., Erkan, A., & Akarun, L. (2003). Real time hand tracking and 3D gesture recognition for interactive interfaces using HMM. In *ICANN/ICONIPP* (pp. 26–29). Berlin, Germany: Springer.

Leap Motion. (2013). Leap Motion to bring 3-D MOTION CONTROL to HP customers through strategic collaboration [Press release]. Retrieved from https://www.leapmotion.com/press_releases/leap-motion-partners-with-HP-on-strategic-collaboration

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

Nacenta, M. A., Kamber, Y., Qiang, Y., & Kristensson, P. O. (2013). Memorability of pre-designed and user-defined gesture sets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1099–1108). New York, NY: ACM.

Ni, T. (2011). *A framework of freehand gesture interaction: Techniques, guidelines, and applications* (PhD dissertation). Virginia Polytechnic Institute and State University, Blacksburg.

Ni, T., Bowman, D., North, C., & McMahan, R. (2011). Design and evaluation of freehand menus selection interfaces using tilt and pinch gestures. *Interaction Journal of Human–Computer Studies*, *69*, 551–662.

Nielsen, M., Storring, M., Moeslund, T. B., & Granum, E. (2004). A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In *Gesture-based communication in human–computer interaction: 5th International Gesture Workshop* (pp. 409–420). Berlin, Germany: Springer.

Rempel, D., Bach, J. M., Gordon, L., & So, Y. (1998). Effects of forearm pronation/supination on carpal tunnel pressure. *Journal of Hand Surgery*, *23*, 38–42.

Rempel, D., Lee, D. L., & Camilleri, M. (2014). The design of gestures for human–computer interaction: Lessons from sign language interpreters. *International Journal of Human–Computer Studies*, *72*, 728–735.

Smith, M., & Schraefel, M. (2004). The radial scroll tool: Scrolling support for stylus- or touch-based document navigation. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*. New York, NY: ACM.

Stern, H., Wachs, J., & Edan, Y. (2006). Optimal hand gesture vocabulary design using psycho-physiological and technical factors. In *Proceedings from 7th International Conference on Automatic Face and Gesture Recognition* (pp. 257–262). Piscataway, NJ: IEEE.

Vatavu, R. D. (2012, July). User-defined gestures for free-hand TV control. In *Proceedings of the 10th European Conference on Interactive TV and Video* (pp. 45-48). New York, NY: ACM.

Vogel, D., & Balakrishnan, R. (2005). Distant freehand pointing and clicking on very large, high-resolution displays. In *UIST '04: Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology* (pp. 137–146). New York, NY: ACM.

Downloaded from hfs.sagepub.com at Stephen B. Thacker CDC Library on October 21, 2015

Wachs, J. (2006). *Optimal hand gesture vocabulary design methodology for virtual robotic control* (PhD dissertation). Ben-Gurion University of the Negev, Beer-Sheva, Israel.

Wachs, J. P., Kolsch, M., Stern, H., & Edan, Y. (2011). Vision-based hand-gesture applications. *Communications of the ACM*, *54*, 60–71.

Wachs, J. P., Stern, H. I., Edan, Y., Gillam, M., Handler, J., Feied, C., & Smith, M. (2008). A gesture-based tool for sterile browsing of radiology images. *Journal of the American Medical Informatics Association*, *15*, 321–323.

Webster, B. S., & Snook, S. H. (1994). The cost of compensable upper extremity cumulative trauma disorders. *Journal of Occupational Medicine*, *36*, 713–717.

Wobbrock, J. O., Morris, M. R., & Wilson, A. D. (2009). User-defined gestures for surface computing. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI '09) (pp. 1083-1092). New York, NY: ACM.

Wright, M., Lin, C. J., O'Neill, E., Cosker, D., & Johnson, P. (2011). 3D gesture recognition: An evaluation of user and system performance. In *Proceedings of Pervasive 2011* (pp. 294-313). Berlin, Germany: Springer.

Wu, M., & Balakrishnan, R. (2003). Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In *Proceedings of the 16th Annual ACM symposium on User Interface Software and Technology* (pp. 193–202). New York, NY: ACM.

Zhang, Z. (2012). Microsoft Kinect sensor and its effects. *IEEE MultiMedia*, *19*, 4–10.

Anna Pereira is a user researcher at Microsoft in Redmond, Washington. She obtained her PhD in ergonomics from the University of California, Berkeley in 2013.

Juan P. Wachs is an assistant professor in the School of Industrial Engineering at Purdue University. He obtained his PhD in industrial engineering and management from the Ben-Gurion University in Israel in 2008.

Kunwoo Park is an undergraduate student at the University of California, Berkeley.

David Rempel is a professor in the Department of Bioengineering at the University of California at Berkeley. He obtained his MD from the University of California, San Francisco, in 1982.

Date received: April 17, 2014
Date accepted: October 13, 2014

Downloaded from hfs.sagepub.com at Stephen B. Thacker CDC Library on October 21, 2015