TouchTools: Leveraging Familiarity and Skill with Physical Tools to Augment Touch Interaction

Chris Harrison　　　　**Robert Xiao**　　　　**Julia Schwarz**　　　　**Scott E. Hudson**

Human-Computer Interaction Institute

Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh PA 15213

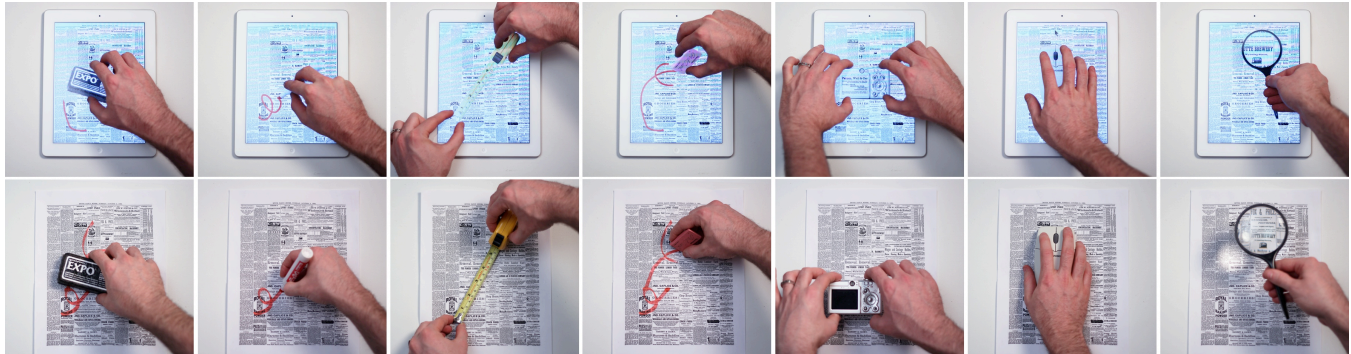{chris.harrison, brx, julia.schwarz, scott.hudson}@cs.cmu.edu

Figure 1. Our example suite of TouchTools (top row; real tools shown below). From left to right: whiteboard eraser, marker, tape measure, rubber eraser, camera, mouse, magnifying glass. Contemporary multitouch interactions, such as single finger panning and two-finger pinch are unaffected. Compare these hand poses to the gestures from prior systems shown in Figure 2.

ABSTRACT

The average person can skillfully manipulate a plethora of tools, from hammers to tweezers. However, despite this remarkable dexterity, gestures on today's touch devices are simplistic, relying primarily on the chording of fingers: *one*-finger pan, *two*-finger pinch, *four*-finger swipe and similar. We propose that touch gesture design be inspired by the manipulation of physical tools from the real world. In this way, we can leverage user familiarity and fluency with such tools to build a rich set of gestures for touch interaction. With only a few minutes of training on a proof-of-concept system, users were able to summon a variety of virtual tools by replicating their corresponding real-world grasps.

INTRODUCTION

The success of modern interactive computing is due in no small part to the efficacious porting of physical elements to digital interfaces. Buttons, tabs, sliders, folders, and even the larger desktop metaphor all draw upon our experiences in the real world to make us comfortable and agile in the virtual one. Simultaneously, the average person can skillfully manipulate an impressive array of tools, from scissors and glue sticks to wrenches and hammers.

The core idea behind *TouchTools* is to draw upon user familiarity and motor skill with tools from the real world, and bring them to interactive use on computers. Specifically, users replicate a tool's corresponding real-world grasp and press it to the screen as though it was physically present. The system recognizes this pose and instantiates the virtual tool as if it was being grasped at that position (Figure 1 and Video Figure). Users can then translate, rotate and otherwise manipulate the tool as they would its physical counterpart. For example, a marker can be moved to draw, and a camera's shutter button can be pressed to take a photograph.

Like using our hands in the real world, this approach provides fast and fluid mode switching, which is generally cumbersome in today's interactive environments. Contemporary applications often expose a toolbar that allows users to toggle between modes (e.g., pointer, pen, eraser modes) or require use of a special physical tool, such as a stylus. TouchTools can utilize the natural modality of our hands, rendering these accessories superfluous.

Further, the gestures employed on today's touch devices are relatively simplistic. Most pervasive is the chording of the fingers [1,8]. For example, a "right click" can be triggered with a two-fingered tap. On some platforms, moving the cursor vs. scrolling is achieved with one or two finger translations respectively. On some Apple products, four-finger swipes allow users to switch between desktops or applications. Other combinations of finger gestures exist, but they generally share one commonality: the *number of fingers* parameterizes the action. This should be rather startling, as very few actions we perform in the real world rely on poking with different numbers of fingers [12,24].

RELATED RESEARCH

There has been considerable work invested in the design of novel gesture sets, both in free-space (see e.g. [3]) and on planar touchscreens, the latter of which most closely relates to this work. Figure 2 offers a wide, though not exhaustive

overview of hand gestures in the literature. Additionally, our motivations overlap with those of tangible and graspable interfaces (see e.g., [20]). However, being entirely virtual gives TouchTools a significantly different flavor, both from a user experience and technical perspective.

Researchers have proposed several touch gesture sets that go beyond finger counting. For example, it is possible to use the shape of the hand's contact area, as imaged by a planar touch sensor. For example, the RoomPlanner system [26] utilized a set of whole hand gestures, including a single finger, the hand laid flat, the hand in a corner shape, and the edge of the hand. Wigdor et al. [21] also used the edge of the hand (a "rail" gesture) along with "curved rail" and "rock" gesture for performing manipulation tasks on a large interactive surface. ShadowGuides [8] put forward an assistive visualization scheme to aid users in learning and performing a variety of complex touch gestures. Although the above gesture sets were designed by experts, it is also possible to craft gesture sets by soliciting ideas and designs from end users [5,25]. Though these systems use hand *shape*, they do not explicitly model *grasps* – hand poses humans employ to *manipulate tools*.

This work was inspired by SurfaceMouse [2], which allowed users to summon a virtual mouse by placing their hand on an interactive tabletop as though a mouse was present. More generally, [27] introduces the concept of *registration*, where an initial hand shape is used to parameterize a gesture continued on a surface, even if it is later relaxed.

Instead of using hand shape in a modal fashion, it is also possible to model the hand in a physical manner. For example, Wilson et al. [23] demonstrated the ability to pick up, pin, corral and otherwise manipulate virtual objects by modeling the hand as a collection of boundaries (e.g., a cloud of spheres) in a physics simulation. SmartSkin [17] allowed users to form barriers with their hands to push and capture objects such as balls. ShapeTouch [4] allowed users to apply small or large virtual forces to objects by varying hand contact area and shape. This could be used to pile objects together, then selectively move the whole pile or a single item, or move a collection of objects by "caging" them with the fingers. It was also possible to pin objects and even peel them back. Of note, the design of the latter systems draws more upon the *physics* of the real world than it does from use of *physical artifacts* in the real world.

There has been also considerable work on grasping virtual objects, especially in virtual or augmented reality contexts [16]. With respect to surface interactions, Wilson et al. [22] modeled the hands and fingers as a series of long rods, which could be used to pick up virtual objects. Although this enabled "grasping", it is not a modal grasp in the same manner as one might manipulate a pen or eraser. Instead, such methods generally offer six degree-of-freedom translation and rotation of virtual 3D objects.

Hand pose estimation is also related to our work. Early systems used mechanical gloves, but now it is possible to avoid instrumenting the user through e.g., computer vision [6]. In the touchscreen domain, people have used IR illumination [7,19], capacitive sensing [18], fingerprint scanners [10], ARTag instrumented gloves [13] and depth cameras [14] to capture aspects of hand and finger pose.

Finally, researchers have demonstrated many compelling uses for hybrid finger+tool systems, including pens [9] and special purpose widgets [20]. This interactive, multi-tool use is something we also strongly advocate. However, styli can be problematic to integrate into smaller mobile devices, and it is clearly not feasible for mobile devices to contain a small suite of full-sized tools.

DEVELOPING A TOOLSET
Identifying Tool Opportunities
To ground our work and better understand the full extent of tool use opportunities, we ran an hour-long brainstorming session with four graduate students studying human-computer interaction (three male, mean age 26.3). Each was paid $10 for their participation. To begin, participants individually brainstormed physical tools they used at least once a year. Each tool was written down on a post-it note. After ten minutes, these were read aloud in a round-robin fashion; unique tools were stuck on a large white board. In total, 72 unique tools were enumerated. Participants were then asked to refine this set, selecting tools with "possible and productive" digital uses. This eliminated tools such as virtual watering cans and salad spinners, reducing the set to 51 tools. From this list, we selected 18 tools we believed had the most compelling interactive uses.

Reflection and Refinement
Our investigations then moved to an analysis of how these 18 tools were grasped and manipulated in the real world. Strong tool candidates had to be ergonomically sound, able to be sensed, and unique in their grip pose and orientation. We found that several tool grasps were incompatible with the planar sensing provided by capacitive touchscreens. Scissors, for instance, require motion perpendicular to the screen to simulate cutting a virtual sheet. In the future, these issues might be partially mitigated by hover-capable touchscreens [7,18] or devices equipped with depth cameras that can capture above-screen, free-space actions [11].
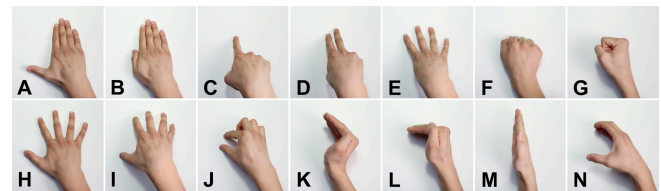

Figure 2. A comparative set of gestures from the literature. A: 'L' shape [5,8], B: flat hand with fingers together / wipe [5, 8,25,27], C: one finger [1,4,5,8,14,25,26], D: two fingers [1,8, 14], E: four fingers [1,8,14], F: grab / wipe [5,27], G: fist / rock / wipe [5,8,21,27], H: flat hand with fingers apart [4,5, 8,26], I: cage / five-finger crumple / five fingers / grab [4,8, 14,27], J: pinch / 'C' Shape / peel [4,5,25], K: curved rail [21], L: corner-shaped hand / curved hand [4,5,8,27], M: side hand / rail / vertical hand / karate chop / pile [4,5,8,21,26,27], N: cupping [4]. Gestures have been combined when similar.

Additionally, several tools utilized the same grasps. For example, scalpels and pens are both manipulated with external precision grips (also called a "writing" or "tripod" grip; see e.g., [12,15,24,28] for an extended discussion of grasp taxonomies). Pliers and rubber erasers are also manipulated with precision grips, but characterized by a larger aperture between the thumb and other digits. Thus, for example, it is possible to reliably disambiguate between a pen and rubber eraser, but not between a pen and scalpel.

To further refine our toolset, we then collected physical props for each tool of interest, and reflected on our own use and grasps. Some tools that initially violated one of the above criteria were found to have suitable planar gestural analogues. For example, we found that putting a camera directly onto a subject surface (e.g., a newspaper) did not significantly break our perception of its function. In particular, the live digital preview (LCD screen) allowed us to violate physical laws that other analog tools could not.

Ultimately, we selected seven tools to serve as our proof of concept set: whiteboard eraser, marker, tape measure, rubber eraser, digital camera, computer mouse, and magnifying glass (Figure 1). This serves only as an example set, as more grasps are possible, and alternative tools could be chosen for grips that conflicted (e.g., swapping marker for pencil). Importantly, these tools do not interfere with typical multitouch gestures, such as one-finger pan and two-finger zoom, offering backwards compatibility.

Observing Real-World Grasps
Our next step was to study how a larger cross section of people manipulated our seven selected tools. Seven participants were recruited (four female, mean age 25.9), who were each paid $10 to participate. Individually, each participant was asked to select a tool out of a box and demonstrate its use. Participants were asked to settle on an exemplar grasp, which was photographed. We then instructed participants to recreate the hand pose on top of an iPad, which recorded the touch contacts. Finally, to capture variety, participants were asked to remove their hand, relax, and recreate the hand pose two more times, which was also recorded by the iPad. This procedure was repeated for the six remaining tools in the test set, yielding a 147 training instances in total (7 participants x 7 tools x 3 repetitions).

This process had three important outcomes. Foremost, it was vital to look at grasps employed by other people to increase external validity. Fortunately, it appears that people hold objects in a relatively consistent manner. Secondly, these results directly informed the design of our gestures, which required some minimal standardization. Finally, the touch data we collected allowed us to craft a feature set for use in our classification engine, which we describe next. Importantly, our classifier is sufficiently robust that users do not have to perform the grasps identically, just similarly.

PROOF OF CONCEPT IMPLEMENTATION
Our proof of concept system, implemented on an iPad 2, records all touches during the first 100ms after an initial touch. This delay ensures that the hand has come to rest on the screen. Next, the system performs classification. If there are two or fewer touch points, the gesture is interpreted as starting a traditional tap, pan, or pinch-to-zoom action. Otherwise, the system computes a series of rotationally invariant features, allowing tools to be summoned at any angle.

Our features are as follows: the number of touch points, the estimated total touch area, and the magnitude of the principal components of the point cloud. We also compute simple statistical features (mean, median, min, max, standard deviation) over four sets of data: distances between each pair of points, distance from each point to the centroid of the point cloud, angles between consecutively-clockwise points as measured from the centroid, and the size of each point. These features are then fed into a support vector machine classifier trained on previously recorded data (see previous section). The resulting grasp classification is used to instantiate the appropriate virtual tool. Each tool has its own operational logic and draw code, allowing for tool-specific actions.

An artifact of prototyping with an iPad 2 is the inability to detect touches with fingernails. This appears to be a purposeful calibration decision by Apple. Other researchers have circumvented this by building their own hardware [17,18]. Some newer devices (e.g. Samsung Galaxy S4), can detect nails, so this particular limitation may soon be moot.

EVALUATION
We recruited 6 new participants with short nails (one female; mean age 25.7), who were paid $10 each for their time. After a brief explanation of the projects goals, participants were given physical versions of the seven test tools to handle. Participants were then asked to replicate corresponding grasps on an iPad 2 running our classifier (trained on 7 independent participants, as previously noted). If needed, the experimenter demonstrated the necessary triggering grasp. No participant took more than three minutes before they indicated they were confident to proceed to the study, after which the experimenter rendered no assistance.

Participants were then instructed to perform one of nine possible gestures: our seven tools, plus one-finger pan and two-finger pinch. These were announced verbally one at a time in a random order (each gesture was requested three times in total). The system performed real-time classification: if the system instantiated the wrong tool, this was counted as an error. Out of 162 trials, there were 7 errors: a nine-class accuracy of 95.7%. Of these seven errors, three were due to a participant performing an incorrect gesture.

DISCOVERABILITY AND SCALABILITY
Anecdotally, we found that our participants were able to discover the mouse, camera, and whiteboard eraser tools with ease. The marker, tape measure, magnifying glass, and rubber eraser often required the experimenter to demonstrate an example grasp to ensure reliable operation. It is possible that additional training data (containing a more diverse set of grasps) could mitigate this problem. Alternatively, users could provide example grasps to the system to help train it. Importantly, after a single demonstration by the experimenter, users were essentially fluent.

Overall, we believe that TouchTools has the promise to be discoverable with good design and sufficient training data. If you tell users there is a mouse function, or a dry erase marker, we believe they are likely to find it through trial and error, and upon discovery, immediately know *what it does* and *how to use it*. This goes beyond most previous gesture approaches. A quick glance at Figure 2 illustrates this – without instruction, these gestures would require considerable trial and error. Moreover, even when discovered, these gestures have little or no iconic meaning for users to guess what action each performs. For example: which gesture in Figure 2 means erase? We believe that designing gestures around real world tools improves discoverability, intelligibility and makes gestures memorable.

The scalability of TouchTools is primarily limited by unique hand poses able to be captured by a touchscreen. As the gesture set increases in size, gesture space "collisions" become more likely, and without good design, the gesture set could become less discoverable, intelligible and memorable. Nonetheless, our prototype implementation offers nine hand-pose gestures, which compares favorably to other interactive systems demonstrated to date ([1,4,5,8,14,21,25, 26,27] average 5.1 gestures each, SD=2.6). TouchTools, above all else, is a gesture design approach, and there is no shortage of tools that have digital uses and unique hand poses. We hope this work offers a new lens through which the HCI community can craft novel touch experiences.

REFERENCES

1. Apple Computer, Inc. Multi-touch gesture dictionary (2008). US Patent 20070177803.
2. Bartindale, T. Harrison, C., Olivier, P. L., and Hudson, S. SurfaceMouse: Supplementing Multi-Touch Interaction with a Virtual Mouse. In *Proc. TEI '11*. 293-296.
3. Baudel, T., and Beaudouin-Lafon, M. Charade: remote control of objects using free-hand gestures. *Comm. ACM,* 36(7). 28–35.
4. Cao, X., A. Wilson, R. Balakrishnan, K. Hinckley, S. Hudson. ShapeTouch: Leveraging contact shape on interactive surfaces. In *Proc. ITS '08*. 129–136.
5. Epps, J., Lichman, S. and Wu, M. A study of hand shape use in tabletop gesture interaction. *CHI EA '06*. 748-753.
6. Erol, A., Bebis, G., Nicolescu, M., Boyle, R., and Twombly, X. Vision-based hand pose estimation: A review. *Computer Vision and Image Understanding*, 108(1), 2007, 52-73.
7. Ewerling, P., Kulik, A., and Froehlich, B. Finger and hand detection for multi-touch interfaces based on maximally stable extremal regions. In *Proc. ITS '12*. 173-182.
8. Freeman, D., Benko, H., Morris, M. and Wigdor, D. ShadowGuides: Visualizations for in-situ learning of multi-touch and whole-hand gestures. *Proc. ITS '09*. 165–172.
9. Hinckley, K., Yatani, K., Pahud, M., Coddington, N., Rodenhouse, D., Wilson, A., Benko, H. and Buxton, B. Pen + touch = new tools. In *Proc. UIST '10*. 27-36.
10. Holz, C. and Baudisch, P. The generalized perceived input point model and how to double touch accuracy by extracting fingerprints. In *Proc. CHI '10*. 581-590.
11. Kratz, S., Rohs, M., Guse, D., Müller, J., Bailly, G. and Nischt, M. PalmSpace: continuous around-device gestures vs. multitouch for 3D rotation tasks on mobile devices. In *Proc. AVI '12*. 181-188.
12. MacKenzie, C. L., and Iberall, T. (1994). The Grasping Hand. Advances in Psychology, Vol. 104. Elsevier Science B.V., The Netherlands.
13. Marquardt, N., Kiemer, J., Ledo, D., Boring, S. and Greenberg, S. Designing user-, hand-, and handpart-aware tabletop interactions with the TouchID toolkit. In *Proc. ITS '11*. 21-30.
14. Murugappan, S., Vinayak, Elmqvist, N. and Ramani, K. Extended multitouch: recovering touch posture and differentiating users using a depth camera. In *Proc. UIST '12*. 487-496.
15. Patkin, M. (1981). Ergonomics and the Surgeon. Clinical Science for Surgeons. Butterworth. London, UK.
16. Poupyrev, I., Weghorst, S., Billinghurst, M., Ichikawa, T., A framework and testbed for studying manipulation techniques for immersive VR. In *Proc. VRST '97*. 21-28.
17. Rekimoto, J. SmartSkin: An infrastructure for free-hand manipulation on interactive surfaces. In *Proc. CHI '02*. 113–120.
18. Rogers, S., Williamson, J., Stewart, C., and Murray-Smith, R. FingerCloud: uncertainty and autonomy handover in capacitive sensing. In *Proc. CHI '10*. 577-580.
19. Schmidt, D., Chong, M.K. and Gellersen, H. HandsDown: hand-contour-based user identification for interactive surfaces. In *Proc. NordiCHI '10*. 432-441.
20. Weiss, M., Wagner, J., Jansen, Y., Jennings, R., Khoshabeh, R., Hollan, J.D. and Borchers, J. SLAP widgets: bridging the gap between virtual and physical controls on tabletops. In *Proc. CHI '09*. 481-490.
21. Wigdor, D., Benko, H., Pella, J., Lombardo, J. and Williams, S. Rock & rails: extending multi-touch interactions with shape gestures to enable precise spatial manipulations. In *Proc. CHI '11*. 1581-1590.
22. Wilson, A. Simulating grasping behavior on an imaging interactive surface. In *Proc. ITS '09*. 125-132.
23. Wilson, A., Izadi, S., Hilliges, O., Garcia-Mendoza, A. and Kirk, D. Bringing physics to the surface. In *Proc. UIST '08*. 67–76.
24. Wilson, F. (1998). The Hand: How Its Use Shapes the Brain, Language, and Human Culture. Pantheon Books, New York.
25. Wobbrock, J., Morris, M. and Wilson, A. User-defined gestures for surface computing. *Proc. CHI '09*. 1083-1092.
26. Wu, M. and Balakrishnan, R. Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In *Proc. UIST '03*. 193–202.
27. Wu, M., Shen C., Ryall, K., Forlines, C. and Balakrishna, R. Gesture Registration, Relaxation, and Reuse for Multi-Point Direct-Touch Surfaces, *Proc. Tabletop '06*. 185-192.
28. Wynn-Parry, C. B. (1981). Rehabilitation of the Hand. Butterworths, London, UK.