Understanding Touch

Christian Holz and Patrick Baudisch
Hasso Plattner Institute
Potsdam, Germany
{christian.holz, patrick.baudisch}@hpi.uni-potsdam.de

ABSTRACT
Current touch devices, such as capacitive touchscreens are based on the implicit assumption that users acquire targets with the center of the contact area between finger and device. Findings from our previous work indicate, however, that such devices are subject to systematic error offsets. This suggests that the underlying assumption is most likely wrong. In this paper, we therefore revisit this assumption.

In a series of three user studies, we find evidence that the features that users align with the target are *visual* features. These features are located *on the top* of the user's fingers, not at the bottom, as assumed by traditional devices. We present the *projected center model*, under which error offsets drop to 1.6mm, compared to 4mm for the traditional model. This suggests that the new model is indeed a good approximation of how users conceptualize touch input.

The primary contribution of this paper is to help understand touch—one of the key input technologies in human-computer interaction. At the same time, our findings inform the design of future touch input technology. They explain the inaccuracy of traditional touch devices as a "parallax" artifact between user control based on the top of the finger and sensing based on the bottom side of the finger. We conclude that certain camera-based sensing technologies can inherently be more accurate than contact area-based sensing.

Author Keywords
Touch input, targeting, generalized perceived input point model, experiment.

ACM Classification Keywords
H5.2 [Information interfaces and presentation]: User Interfaces. Input devices & strategies.

General Terms
Design, Experimentation, Human Factors.

INTRODUCTION

Current touch technologies, such as *capacitive touchpads* [7] and *FTIR* [11], sense touch by observing the *contact area* between finger and device. These devices are thereby based on the implicit assumption that the contact area encodes the information about the desired target in the first place, i.e., that *users* somehow use the contact area to encode which target they are trying to refer to, e.g., by touching the target with the center of the finger contact area.

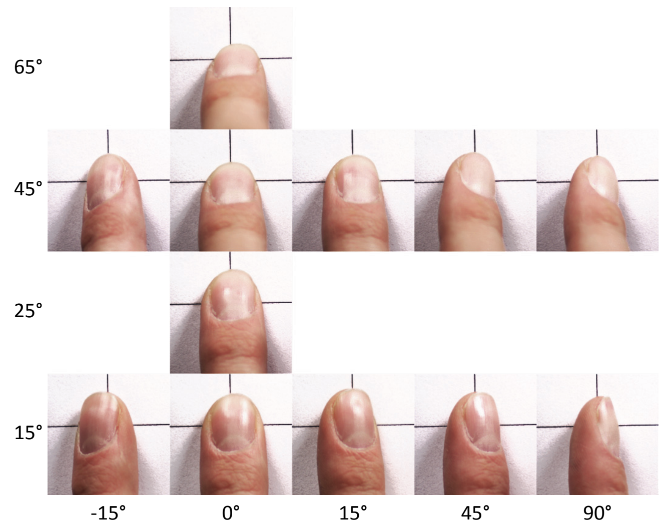


Figure 1: A study participant targeting crosshairs using different finger angles. Can you guess how this user is conceptualizing touch, i.e., what geometric relationship between finger and crosshairs the user is trying to maintain independent of how the finger is held? Our findings suggest that users indeed target as suggested by this illustration, i.e., by aligning finger features and outlines in a hypothesized top-down perspective.

Our recent findings seem to put this assumption into question. While the "contact area model" is clearly plausible on a macroscopic scale, our earlier findings with very small targets indicate that touch input on such devices is subject to systematic error offsets (*generalized perceived input point model* [12]). In fact, as much as two thirds of the overall inaccuracy of touch seem to be caused by these effects. When users target with an almost horizontal finger, for example, the target position measured by a capacitive touch device is off by as much as several millimeters—on small screens devices this is a substantial effect. In our earlier studies, the size and direction of the error offset was affected by a range of parameters, including finger posture measured in roll, pitch, and yaw [12].

In the aforementioned paper, we compensated for this effect using an elaborate schema of *corrective offsets* (user- and finger-posture specific position adjustments). The existence of these systematic offsets, however, raises much deeper questions. In particular, the existence of these offsets seems to indicate that the assumption these devices are built on, i.e., that users target based on finger contact area, *is wrong*.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

So if users do not touch targets based on contact area, how *do* they target? How do they decide whether a finger is "on target" or whether it requires corrective moves? In this paper, we are attempting to answer this question.

UNDERSTANDING TOUCH

In order to help us specify what we are trying to find out, Figure 2 illustrates the general concept of touch input: users communicate a 2D target location to a touch device. As users acquire a target with their finger, such as the shown crosshairs (a), they effectively translate the 2D target location into the 6D posture of their finger (6D = 3D rotation + 2D position + pressure). (b) Due to the lack of a better term, we will refer to this 2D-to-6D mapping as *the user's mental model of touch*, in the traditional sense of a user's mental model of the way an object operates [15], but with the "object" being the user's own finger.

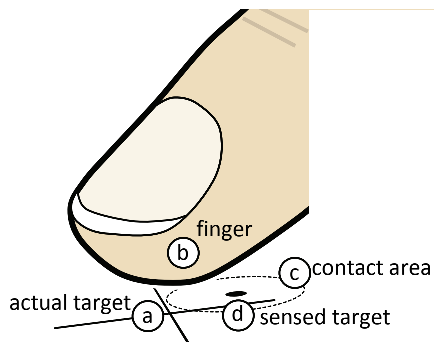


Figure 2: When acquiring a target, here marked with 2D crosshairs (a), users effectively translate the 2D position of the target into a 6D position of their finger. On traditional touch devices, the finger leaves a contact area (c), which is observed by the touch pad and reduced to its 2D centroid (d).

The objective of any touch input device is to *invert this translation*, i.e., to reconstruct the 2D target location from this 6D finger posture. We will refer to this 6D-to-2D mapping as *the device's conceptual model.* Perfect reconstruction is achieved if and only if the mapping implemented by the touch device is indeed inverse to the 2D-to-6D mapping implemented by the user, i.e., if the device's conceptual model matches the user's mental model.

Current touch devices implement this back-translation as illustrated by Figure 2c and d: They observe the 2D contact area between finger and pad (c) and reduce it to its centroid (d). As explained earlier, however, our previous findings indicate that this *contact area model* is not correct, i.e., it *does not* reconstruct the intended input position accurately. Apparently, users do *not* aim by pointing using the center of the contact area.

Note: throughout this paper, we will use an apparatus similar to Figure 2, i.e., *crosshairs* marking the target. In our previous studies, crosshairs performed indistinguishably from a dot target [12], which suggests that the influence of crosshairs onto users' mental model of touch is reasonably small.

An example and preview of findings

As a preview of our analysis, take a look at the sequence of images shown in Figure 1. They show a user (Participant #7 from User Study 1 presented in this paper) targeting a pair of crosshairs with different finger angles. Looking across the sequence of images, we may already catch some indication of what mental model of touch this user adheres to. Certain geometric relationships between finger and crosshair seem to remain—independent of what finger posture the experimenter makes this participant assume.

The participant shown in Figure 1 is a particularly good representative of the new model of touch we propose and which we call the *projected center model*. This model says that users align certain *visual* features with the target. In the shown example, it is the horizontal center of the finger outline and the vertical center of the fingernail that the user is aligning with the target.

We chose the specific viewpoint of this image sequence with intent: even though the user's head was actually located to the bottom left of the picture during these trials, our findings suggest that users imagine this top-down perspective during touch input. Based on this perspective, they decide whether their finger is on target or whether it requires adjustment.

Under the projected center model, the error offsets of the contact area model effectively disappear (they drop to 1.6mm, compared to 4mm for traditional, contact area-based sensing), suggesting that the projected center model matches users' mental model of touch very closely.

The projected center model also explains why capacitive touch input devices are inaccurate: devices based on the contact area touch model sense features located at the *bottom* of the users' fingers, while users target based on features located *on the top/along the sides* of their fingers. The inaccuracy of touch on traditional touch devices is therefore an artifact of the parallax between the top and bottom of a finger.

Approach

In the remainder of this paper, we present a series of studies that validate the reasoning outlined above.

Many models in HCI are created by measuring a feature and fitting a function to it. Unfortunately, we do not know yet what feature to measure or even what modality (sense of touch, vision, and so on). This forces us to take a more general approach to the problem: "(1) *Guess* the law (or model in our case), (2) compute the consequences, and (3) compare the computation to experiment/observation. If it disagrees with experiment it is wrong [6]."

In this paper, we apply these three steps as follows: (1) Before we attempt to guess mental models, we narrow down the search space. We conduct a series of interviews and then consider only the subset of models that are based on features mentioned by participants. (2) The consequences we predict are that models that match the user's mental model will feature error offsets approximating zero. (3) We conduct a series of pointing studies. We measure error offsets as the average distance between the sensed input location and the actual target location for the respective user (cf. offset in our previous work [12]).

Since our primary goal is to understand touch, we require the remaining error offsets of a good candidate model to be small. Only when the remaining offsets get reasonably close to zero can we argue that the tested model indeed corresponds to the actual mental model of the respective user and thus contributes to an explanation of touch.

Procedure

We proceed in four steps.

Step 1—Interviews: We interview users to learn how they (think they) target using touch input, i.e., what features they try to align with the target.

Step 2—Model creation: Based on participants' input, we create a set of 7×7 candidate models. User input inspires us to focus on models based on a top-down perspective.

Step 3—Filtering models: We conduct two pointing studies in which we determine error offset for all candidate models under different variations of finger and head postures. We eliminate models with large offsets, as they indicate a poor representation of participants' mental models. We keep 2×3 candidate models.

Step 4—Final evaluation: We conduct a final pointing study using the combined set of independent variables from the studies in Step 3 (finger and head position) to determine the error offsets and thus the "fit" of the remaining models.

Contribution

We primarily make a scientific contribution. With *scientific* we mean: an attempt to understand an *underlying*, not directly observable mechanism—in our case touch input, one of the key technologies in human computer interaction. In particular, we explain why current touch devices are inaccurate by challenging the common assumption that touch input is about the contact area.

Still, essentially as a side effect, our findings have implications on engineering in that they inform the design of better touch input technology. They suggest that devices that observe the outline or "projection" of a finger have the potential to offer better touch precision than devices based on contact area.

RELATED WORK

The work presented in this paper is related to touch input, targeting aids, impact of finger posture on touch precision, and modeling users' input behavior.

Touch technologies and determining contact points

A wide range of touch technologies determine the input coordinates based on contact area (e.g., *capacitive sensing* [7]). Most optical technologies have adopted the same approach, such as *frustrated total internal reflection* (*FTIR*, e.g., [11], [19]). Wang and Ren examined participants' input precision on such an optical system [21].

Camera-based systems, such as the *Digital Desk* [22], typically cannot see the contact area and have therefore adopted a visual model. The *Visual Touchpad* uses two cameras to identify the user's fingertips and determine finger yaw [13]. Agarwal et al. used stereo cameras above the user's hands and trained a classifier to recognize fingertips and accurately determine when the user is touching or hovering their finger above the surface [1]. *LucidTouch* centers the contact point in the outline at a fixed offset from the fingertip [23]. *PlayAnywhere* observes fingers from above with a camera mounted at an angle and derives touch input from the distance between the user's fingers and the shadows they cast [24].

Targeting aids

The inaccuracy of touch has been attributed to the *fat finger problem* [20], i.e., the softness of the fingertip in combination with the fact that the finger occludes the target. A series of targeting aids alleviate the problem by enlarging the target, e.g., *TapTap* [18] (see also [2]) and *Dual Finger Stretch* [4]. Others avoid occlusion altogether by introducing an offset between finger and target, e.g., *offset cursor* [16], *shift* [20], *cross-lever* and *precision handle* [2], as well as the use of *styli* [17]. *Back-of-device interaction* [3] avoids occlusion by moving the user's finger to the device's back.

Modeling pointing

Modeling cursor-based target acquisition has a long tradition. *Fitts' Law* models the targeting time for one-dimensional targets [8]. Grossman and Balakrishnan's *probabilistic pointing* models two-dimensional target acquisition under visual control [10]. Both models assume that users can see target location and pointer, a requirement not fulfilled by touch on targets that are small enough to be occluded by the user's finger.

Early Models of Touch and the Role of Finger Posture

While touch systems traditionally reduce contact areas to a point [21], more recently researchers have proposed considering the entire contact area as input, such as *Shape-touch* [5] and *Sliding Widgets* [14].

Several researches suggested that the angle between the user's finger and the device might impact pointing. When designing the *shift* technique, Vogel and Baudisch [20] included a correction to compensate for parallax. Forlines et al. noticed that touching a target located across a tabletop results in a flat finger angle and thus a larger contact area, which leads to an offset input location [9]. Wang and Ren find that finger posture and motion impact the size of the contact area [21]. We generalized these findings in our previous work on the *generalized perceived input point model* [12].

To illustrate the error offset problem faced by devices based on the contact area model we reprint an extract of the data from this paper (see Figure 3). All 12 participants (each one shown as one column) repeatedly acquired a target on a touchpad under varying finger roll and pitch angles. Contact point distributions are shown as 65% confidence ovals each one of which represents a different roll/pitch condition. While ovals were small, they were *offset* with respect to each other, which implied the need for the aforementioned corrective offsets. Only for Participant #3 (dashed outline) ovals align, suggesting that this was the only person the targeting behavior of whom was appropriately described by the contact area model.

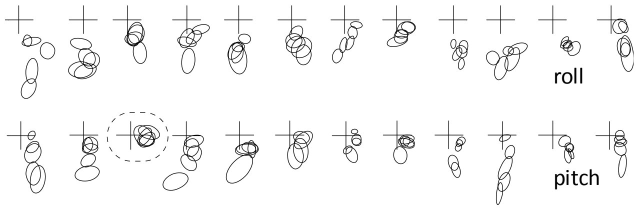


Figure 3: (from [12]) The aggregated contact locations of different finger postures are offset with respect to each other.

STEP 1: USER INTERVIEWS: PARTICIPANTS REFLECT ON THEIR TARGET ACQUISITION STRATEGIES

The purpose of this study was to learn more about users' mental models by means of an interview. While users are known to have limited ability of rationalizing low-level activities, our goal was to create a selection of *potentially* relevant models and elements that could be used to form a list of *candidate* models. We did not worry about incorrect models at this stage, as we would eliminate these in subsequent steps of our process.

Figure 4: Before being interviewed, participants acquired targets printed on a sheet of paper.

Task & Procedure

In order to help participants become aware of their mental models of touch input, participants started by repeatedly acquiring a target. As shown in Figure 4, the target was marked by crosshairs drawn on a sheet of paper and each participant acquired it 50 times. Participants were instructed to place their finger such that it would "acquire a tiny button located at the center of the crosshairs if the paper was a touch screen".

In order to help participants investigate their own mental models in more detail, participants acquired the target using the five finger postures shown in Figure 5, i.e., combinations of finger pitch and roll. All participants completed this part of the study in 10 minutes or less.

After they had completed the trials, we interviewed participants about the strategies they had used to acquire the target and based on what criteria they had decided when their fingers were "on target". We were careful not to use any terms that could bias their answer, such as "contact area", "fingertip", "finger nail", and the like. If they, however, did mention "contact area", we asked participants to draw the contact area in four figures showing stylized fingers top-down held at four different angles.

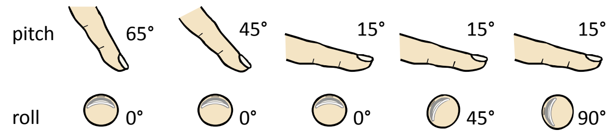


Figure 5: Finger postures. Participants assumed these five different combinations of finger pitch and roll, and then acquired the target.

Participants

We recruited 30 participants (5 female) from our institution. All participants were between 19 and 29 years old.

Results

When asked to verbalize their "targeting procedure", most participants hesitated. 4 participants insisted that they could not explain their behavior and "just touched the target intuitively without giving it too much thought".

6 participants stated right away that their experience with mobile touch-screen devices had shaped their input behavior. While two of them understood how such devices determine input coordinates, they all stated to aim based on experience with the device. They all said that the device had "taught" them how to touch small buttons over time.

Contact Area: 26 participants said that they considered contact area to be relevant to their targeting strategy. 24 of them stated that they imagined the contact area between their finger and the crosshairs and centered it on the target during the trials. One said:

> "I could not see the contact area, so I imagined where it should be located. Then I chose the center of it and positioned it on the target. That's all."

Participants' drawings of contact areas in the four figures supported our assumption that they cannot fully rationalize their behavior. Most drawings largely clashed with reality; while for a finger at a flat angle (15° pitch) the contact area extends fairly far towards the user's palm, participants always drew it too small. Similarly, for a rolled finger (90°), participants drew the contact area too large and mostly centered inside the finger, whereas in fact it is mostly offset horizontally and rather short.

Three participants mentioned a special version of contact area; they claimed to touch the target with the part of the finger that "comes down first". Five other participants explained to place their finger such that it applied the maximum amount of pressure to the target.

Visual Feedback: 13 participants reported that they positioned their finger using *visual control*. They stated that they mentally connected the crosshairs under their finger and tried to move their finger such that the target was always located under the same position inside their finger.

One said:

> "I can see my fingertip and imagine where the cross-hairs intersect. So I can visualize the bottom of my finger and always position it at the same location."

Nine participants explained that they positioned the finger, such that the target was located at a certain distance from the edge of their fingernail. Four other participants said they imagined a virtual 3D point inside their finger, which they repeatedly sought to position directly above the target. Two other participants said that they "projected" a feature in their finger down to the table and then aligned it with the target. This is an interesting observation, because such a projection is a comparably complex 3D operation that requires users to take head parallax into account.

Two participants described their targeting strategy as "cheating". Both had a visible spot on their fingernail that they vertically aligned with the target whenever finger roll was 0°. For roll different from 0° they still aligned the spot with the horizontal line of the crosshairs.

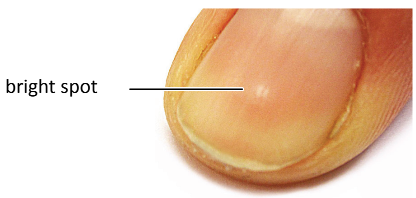


Figure 6: One participant targeted by placing a bright spot on his nail over the target.

Discussion

While the study allows proposing a wide range of possible models that explain how participants targeted, such as models based on a camera that tracks with the user's head, we decided to make an educated *guess* and limit our search for candidate models to those based on features that users can perceive *visually* and *from directly above*. This seemed plausible given that thirteen participants mentioned visual features and six of them mentioned some sort of vertical projection.

As discussed earlier when stating our 3-step approach, whether or not our intuition was right would have to be determined in the following pointing studies. If the model should perform poorly, our guess would turn out to be wrong and we would have to come back and restart the process with another model (note how this is different from phenomena that lend themselves to direct observation, in which case the interviews themselves would have already answered the question).

STEP 2: PICKING CANDIDATE MODELS

We constructed the following two (families of) candidate models (1) contact area, which had produced a good fit for one user in our previous study [12], shown as Participant #3 in Figure 3 and (2) models based on features of human fingers that are visible from above. We implemented this by tracking users' fingers using a camera placed *directly above* the crosshairs on the touchpad.

Figure 7 shows a series of features that we found to be visible from above. We classified them as *horizontal features* if they might help determine the finger's horizontal position and *vertical features* if they might help determine the finger's vertical position; some features, such as the corner of the fingernail are both (e.g., nail left and nail groove). Note the three-dimensional nature of the finger, which causes features, such as *outline center* to refer to the outline of the skin for some levels of finger roll and to the outline of the nail for other levels of finger roll.

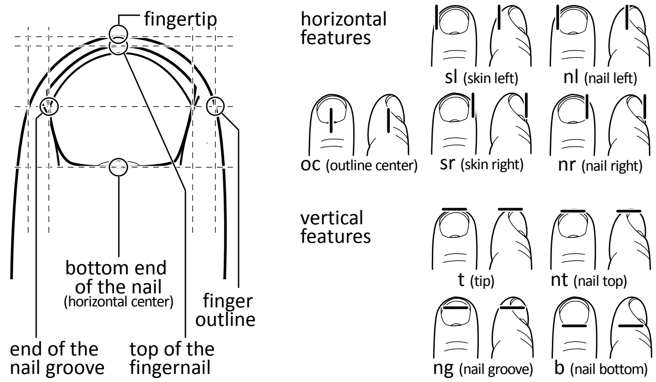


Figure 7: We found these *horizontal* and *vertical features* to be detectable and used them to construct candidate models.

In theory, users' mental models might combine any number of features in arbitrarily complex ways. We felt, however, that the effortlessness of touch pointing suggests that only simple models are truly plausible. We therefore only included models that use a single feature (such as "nr" for the mental model of users who aim relative to the right edge of their nail) and models that refer to the center point between two features, such as "sl|sr" for the mental model of users who aim relative to the center point of the finger's outline, i.e., between "*skin left*" and "*skin right*".

We will refer to terms such as "nr" or "sl|sr" as *half models*, because it takes two of them to describe the mental model of a user—one for x and one for y. To avoid the overhead of evaluating the cross product of *horizontal* and *vertical features*, however, we keep these two classes of features and models separate throughout most of this paper and will not combine them until the final study.

Figure 8 lists the horizontal and vertical half models that we created from the respective features shown in Figure 7.

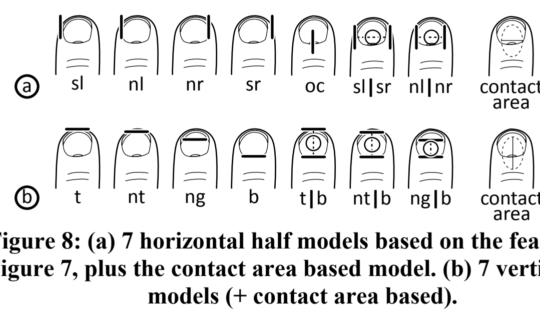


Figure 8: (a) 7 horizontal half models based on the features in Figure 7, plus the contact area based model. (b) 7 vertical half models (+ contact area based).

The idea behind a half model, such as "nail right" was not necessarily that participants would place this specific point over the target, but *some* point that is located at an offset from this feature. To include this concept, we complemented all half-models with a single user-specific *offset* (unlike our previous approach [12], which allowed for one offset *per finger angle*).

STEP 3: ELIMINATING MODELS

Next we eliminated those (half) models that did not match the mental models of any users. In order to do so, we conducted a pointing study. Using a camera above the target, we recorded participants as they repeatedly acquired a target on a capacitive touch pad. We then tried to "explain" the observed data using each of our half models and eliminated all half models that did not fit any participants.

To keep the overall number of repetitions manageable, we broke this study down into two individual studies. The first study tested twelve combinations of roll and pitch; the second tested only four combinations of roll and pitch, but varied head position in addition.

STEP 3A: ELIMINATING MODELS USING ROLL&PITCH

Task

Participants repeatedly acquired a crosshair target located on a touchpad (Figure 9). During each trial, participants first touched a 1"×1" "start" button located 2" left of the target. Participants then assumed the finger angle for the current condition with their right index finger and acquired the target. Participants committed the touch interaction by pressing a footswitch. This recorded the touch location reported by the touchpad, triggered the camera to take a picture, played a confirmation sound, and completed the trial. Participants did not receive any feedback about the location registered by the touchpad.

Figure 9: Participants acquired a crosshair target located on a touchpad and committed input using the footswitch.

We took the following four measures to minimize the impact of other potential factors. First, participants kept their head in a fixed position above the touchpad, as shown in Figure 9. This controlled for parallax. Second, the crosshairs marking the target extended beyond participants' fingers, allowing participants to maintain a certain amount of visual control during targeting. Third, the use of a footswitch to commit input allowed us to avoid artifacts common with other commit methods, such as inadvertent motion during take-off. And finally, participants were told to use as much time as necessary and that task time would not be recorded. Fourth, participants rested their elbow on the adjacent table (Figure 9) to preclude fatigue.

Note that there was no need to include distracter targets. Distracters have a major effect on *adaptive* input techniques, but not on unmodified touch.

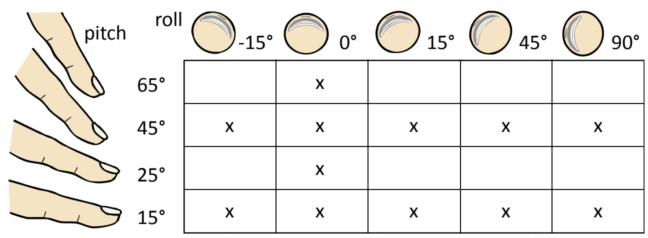
The purpose of using crosshairs was to reduce noise, thus helping us observe the underlying mental models more clearly. While the use of visible crosshairs may in theory impact participants' targeting behavior, we did not observe any such effect in our studies.

Independent variable: finger posture

As shown in Figure 10, we used the same combinations of finger pitch and finger roll as we did in previous work [12]. However, we dropped the 90° pitch condition, because the camera located directly above the pad could not capture the participant's nail in this condition (Figure 9).

Procedure

Overall, participants performed a sequence of 12 angles × 2 repetitions totaling 24 trials. The order of pitch-roll combinations was counterbalanced across participants. Then participants filled out a post-study questionnaire. All participants completed the study in 15 minutes or less.



pitch \ roll	-15°	0°	15°	45°	90°
65°		x			
45°	x	x	x	x	x
25°		x			
15°	x	x	x	x	x

Figure 10: Study conditions. Participants assumed these combinations of finger pitch and finger roll during the study.

Apparatus

The study apparatus recorded contact area using a capacitive touch pad and captured a picture of the participant's finger using an overhead camera. The capacitive pad was a 6.5" × 4.9" *FingerWorks iGesture*; the camera was a *Canon EOS 450D*, capturing participants' fingers at 140dpi. Participants committed trials using a *Boss FS-5U* footswitch. All components were connected to an Intel Core 2 Duo machine running Windows XP.

Participants

We recruited a new set of 30 participants (10 female) from places around our institution. Participants were students from a range of different disciplines and were between 20 and 32 years old. We offered a €20 incentive for the most accurate participant.

Data preparation

During a pilot study, we attached markers to participants' fingers in order to allow for automated tracking. However, participants mentioned that the markers distracted them and some participants had started to include them as features into their targeting model. We therefore dropped the markers and instead annotated visual features in the photographs by hand.

Results

Vertical half models

Figure 11a shows which vertical half models produced the best fit (i.e., the lowest error offsets) for how many participants. We consider a half model to produce the best fit if the systematic offsets produced by this half model for each condition have the smallest distance from the center of mass of all error offsets produced by that model, across all half models.

No single model offers the best fit for all users, suggesting that different users may have *different* mental models. The half model t|b, i.e., the vertical center of the fingernail performs best here—it offered the best fit for 9 participants. It is followed by ng|b a slightly different version of the vertical center of the fingernail, with another 6 participants. As expected based on [12] (see also Figure 3), the contact area model offers the best fit for a very small number of participants, here only 1 out of 30.

Figure 11b shows how well different subsets of vertical half models combined fit the data. Bars represent the average vertical error offset if participants' data is processed using only the respective half models. The red bar on the left represents the error offsets produced by the capacitive baseline condition contact area.

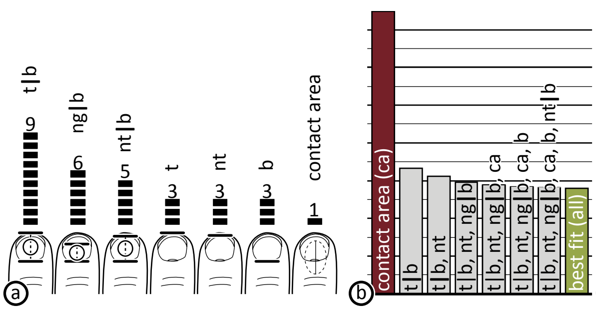


Figure 11: (a) Number of participants for which each vertical half model produced the lowest error. (b) Error produced when using only a subset of the half models to analyze participants. Switching from the model contact area to t|b reduces error offsets to 44%; using all models listed in (a) reduces the error to 37% compared to the contact area model.

The green bar on the right in Figure 11b represents the error offsets produced if every participant's input were processed using their personal best-fit model from Figure 11a; the switch to the best fit model reduces the error offsets by about 63%. We will refer to this best-fit case as "*per-participant models*" in the remainder of this paper.

Some of the half models listed as best fit model are similar. As a result it may not be necessary to maintain all of them. The gray bars in the middle of Figure 11b show how error rate increases if we drop some of the half models. We see that dropping all but three models (t|b, nt and ng|b) incurs a penalty of only 5% compared to using all vertical half models. Dropping all models but t|b incurs a penalty of 18.5% over the best-fit case. The contact area model alone, however, leads to large error offsets (averaging 3.75mm across participants).

Horizontal half models

Figure 12a shows the corresponding data for the horizontal half models. The half model oc (center of the finger outline) produced the lowest error for over a third of the participants (11 of 30). The contact area model offered the best fit for 5 of the 30 participants.

The benefit of using per-participant models is only a factor of 1.4 and, thus, by far not as large as in the vertical case (Figure 12b). This is a result that we expected based on our previous results [12], because of the smaller horizontal extent of clusters in Figure 3. Horizontal error has always been smaller, thus there is less potential for improvement.

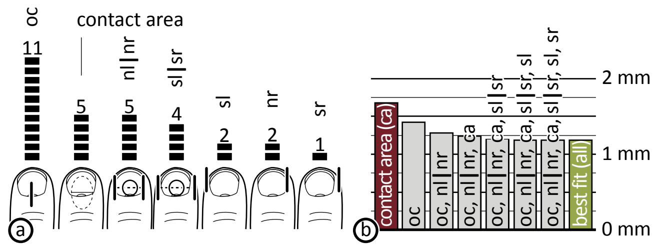


Figure 12: (a) Number of participants for which each horizontal half model produced the lowest-error offsets. (b) Using per-participant half models reduces the error offsets on average to 84% compared to using merely the contact area model.

Discussion

These findings suggest that the horizontal half-model nl (Figure 12a) and the vertical half-model ng (Figure 11a) can be eliminated, as they did not produce a best-fit for any participant. However, we will postpone the decision until we have seen results of the next study.

STEP 3B: ELIMINATING MODELS: HEAD PARALLAX

In contrast to the study in Step 3a, we added head parallax as an independent variable in this study. Again, the purpose of this study was to eliminate candidate half models.

Task

The task was the same as in the previous study, except that in addition to assuming a specific finger posture, participants also assumed one out of four predefined head positions (Figure 13).

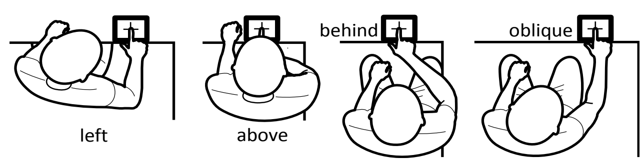


Figure 13: Participants acquired the target from these four different head positions.

Procedure

Overall, each participant completed a sequence of 2 pitch angles (15° and 45°, see Figure 10) × 2 roll angles (0° and 45°) × 4 head positions (Figure 13) × 2 repetitions = 32 trials, in four blocks, one for each head position. Finger angles as well as head positions were counterbalanced across participants. Participants filled out a post-study questionnaire. All participants completed the study in about 20 minutes.

Participants

We recruited a fresh set of 12 participants (5 female). All participants were between 19 and 24 years old. Similar to the previous study, we encouraged participants to be accurate throughout all conditions. Again, we offered a €20 incentive for the most accurate participant.

Apparatus

The apparatus was the same as in the Study 3a (Figure 9).

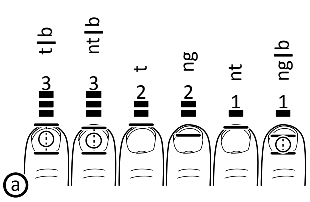
Results

Vertical half models

As in the previous study, the half models describing the vertical center of the fingernail together (t|b and nt|b) produced the best fit for half of all participants (6 of 12, Figure 14a). The contact area model, in contrast, never produced the lowest-error offsets for any participant in this study.

Again, we can reduce the number of half models without sacrificing too much precision; keeping only the vertical model t|b incurs a penalty of 6.5% over using all vertical half models. It still reduces the error of the contact area model by a factor of 2.5.

As shown in Figure 14b, the use of per-participant best-fit half models reduced the error offsets to 1.8mm, from 5mm of the contact area model to 40% of that value.



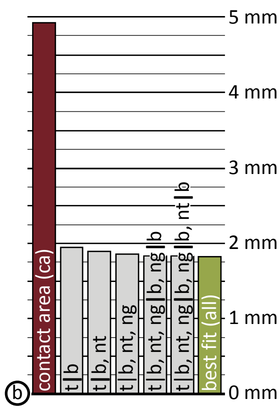


Figure 14: (a) Number of participants for which each vertical half model produced the lowest error. (b) The mean error of input drops to 39% when using t|b instead of contact area.

The contact area model incurs even bigger error offsets than in the previous study (5mm compared to 3.75mm). This suggests that the contact area model is sensitive to changes in head parallax, but more data is required to know for sure.

Horizontal Half models

As in the previous study, the horizontal half model oc produced the best fit for the largest number of participants (7 of 12, see Figure 15a). Compared to the previous study, the capacitive half model produced the best fit for a similar fraction of all participants (here 1 of 12, compared to 5 of 30).

As in the previous study, the use of per-participant half models produced only a moderate reduction of error offsets over the contact area model (about 15%).

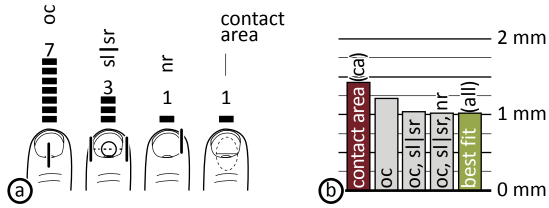


Figure 15: (a) Number of times each of the horizontal models produced the lowest-error offsets per participant. (b) The average error produced by the horizontal best-fit models sinks to 72% compared to contact area. The three horizontal half models oc, sl|sr, and t|b alone account for this improvement.

Discussion

Figure 16 lists the remaining candidate half models. We obtained this list by eliminating all models that did not produce at least one best fit in either one of the two studies. In addition, we eliminated all models the addition of which would have decreases error only marginally—the benefits of including nr, ng|b, and nt|b in the last study, for example, were less than 1%. However, we did maintain the two halves of the contact area model as well as the model *t* (absolute distance from the fingertip, which also only reduced error by less than 1%): both have been implemented in products and related work, so we wanted to see how they do in the final study.

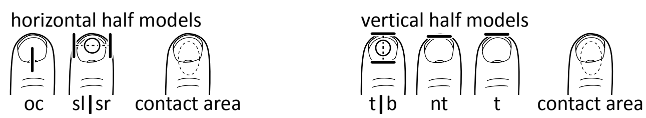


Figure 16: The remaining candidate models.

If we take a closer look at the three remaining horizontal half models, we notice that all of them are versions of the center of the finger outline, only sampled at different locations, i.e., at the bottom of the nail (oc), at the location of the nail grooves (sl|sr), and at the horizontal center of the contact area.

The three remaining vertical half models describe the target location either in relation to the fingernail (t|b) or as an offset for the top of the fingertip (t and nt).

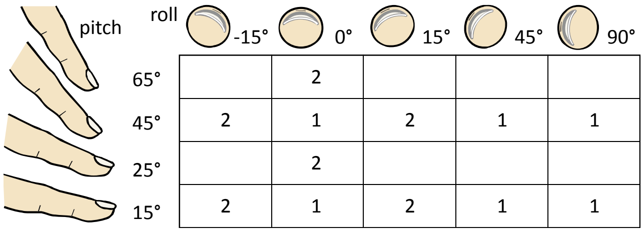
STEP 4: EVALUATING THE MODELS

The purpose of this final study was to evaluate the remaining half models. We could not re-use the data from the previous studies, because we had already used this data for learning and eliminating the very same half models. More importantly, though, participants had performed only few trials per condition; thus the data did not allow us to distinguish error offsets from random noise (i.e., spread by the fat finger problem [12]).

In this final study, we addressed this by increasing the number of repetitions to 4 trials per condition and two blocks. Aggregating these eight trials substantially reduced fat-finger noise and thus revealed the systematic error offsets we were looking for more clearly. These offsets provided us with a more reliable estimate sense of how far a change in mental model could reduce offsets and thus how closely the respective models were actually matching participant's mental models of touch.

Task

The task was the same as in the previous study (3b); we included all 12 levels of finger pitch and roll from Study 3a and all head positions. To keep the number of repetitions per participant manageable, we subdivided the roll/pitch variables between subjects, as shown in Figure 17.



pitch \ roll	-15°	0°	15°	45°	90°
65°		2			
45°	2	1	2	1	1
25°		2			
15°	2	1	2	1	1

Figure 17: To keep the number of trials per participant manageable, we ran roll/pitch between subjects. The table shows the assignment of conditions to the two groups.

Study design

Each participant completed 6 combinations of finger angles (Figure 17) × 4 head positions (Figure 13) × 2 blocks × 4 repetitions = 192 trials. All participants completed the study in 40 minutes or less. Participants filled out a questionnaire afterwards.

Apparatus

The apparatus was the same as in Study 3b (Figure 9).

Participants

We recruited a fresh set of 12 participants (6 female) from places around our institution. All participants were between 21 and 32 years old. As in the previous studies, we encouraged participants to be accurate and offered a €20 incentive for the most precise participant.

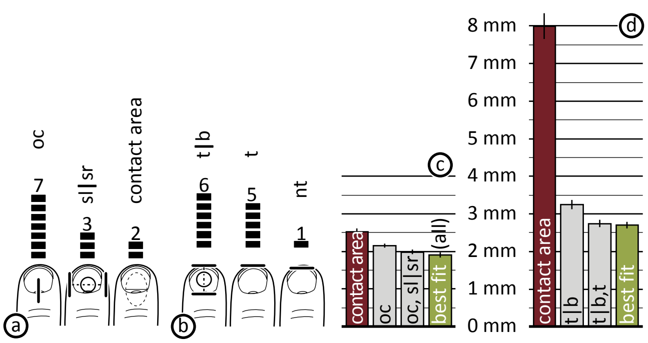


Figure 18: (a) Number of times each of the horizontal models produced the lowest error. (b) Same for the vertical models. (c) Using the best-fit horizontal half models instead of contact area **reduces the error by 15%. (d) Vertical per-participant models reduce the error produced by** contact area **by 60%.**

Results

The final study did not produce anything unexpected as the performance of the participating half models was similar to the previous two studies. As shown in Figure 18a, both oc and sl|sr produced the best fit for 5 of the 12 participants; t and t|b produced the best fit for 5 participants each. While the use of all three horizontal half models yields only 15% less error in offsets (per-participant models), vertically, per-participant models reduce the error offsets substantially by 60%. t|b alone reduces error offsets by a factor of 2.5 compared to contact area (Figure 18b).

Merging Models

Finally, we rejoined half models into full models. Figure 19 shows which half models went together well: the combination oc & t|b produced the best overall fit for 4 of the 12 participants, followed by oc & t, sl|sr & t|b, and contact area & t|b for another 2 participants each.

vertical half models \ horizontal half models	oc	sl|sr	contact area
t|b	4	2	2
t	2	1	
nt	1		
contact area			

Figure 19: Number of times a combination of half models together produced the best fit for a participant.

Figure 20 shows error offsets for the eight full models from Figure 19. A one-way ANOVA with participant as a random variable found a significant main effect of model on error offsets (F_7=38.662, p<0.001). Post-hoc t-tests using Bonferroni correction found that all six other models and the per-participant best-fit aggregate produced significantly lower error offsets than contact area (all p<0.003).

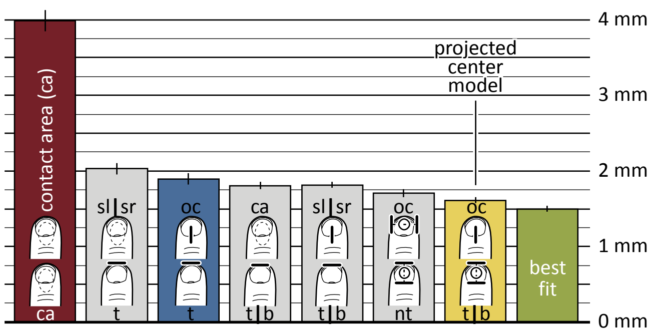


Figure 20: Remaining offset for the full models from Figure 19 compared to contact area **model (+/- std. error of the mean).**

The best individual model was oc & t|b. It says that participants target by placing the horizontal center of their finger outline and the vertical center of fingernail over the target. In the introduction to this paper, we already referred to this model using the name projected center model and the images shown in Figure 1 are best explained using this model.

In summary, the projected center model performed best out of all the models tested. Under the projected center model, the large systematic offsets of 4mm observed by the contact area model shrink down to 1.6mm, an improvement by a factor of 2.5. At the same time, the remaining offsets are close enough to zero to suggest that this model approximates participants' mental model indeed well.

This is the main finding of this paper.

IMPLICATIONS FOR ENGINEERING

Switching to an engineering perspective for a moment, we might be interested in how devices built on these different models will perform, as this might inform which models to base future devices on.

So far, we have discussed models based on systematic offsets, as that is a good metric for testing the quality of mental models. To answer questions about device performance, we add the other error variable, i.e., spread, back in. The resulting *minimum button sizes* for 95% reliable touch input are shown in Figure 21. These values specify how accurately a device based on the respective model can be expected to perform.

The chart shows that devices based on the projected center model allow users to acquire targets of 4.9mm with 95% accuracy, compared to 7.9mm target size for the corresponding contact area model (8.6mm in [12], corrected for yaw to match conditions in this paper). In terms of target surface this difference amounts to a factor of 2.6. This means that a device implementing the projected center model could pack 2.6 times more targets into the same screen space or, alternatively, that a device could be reduced to less than half its size and still allow users to operate it reliably.

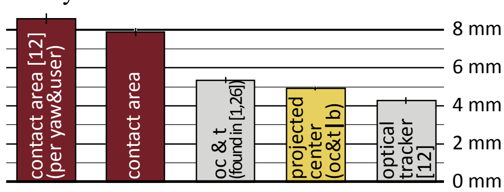


Figure 21: Results from this study in contrast to results of our previous studies in [12].

For reference, in Figure 21 we also included the 4.3mm minimum target size that we previously achieved by attaching retro-reflective markers to the users' finger and tracking using an optical tracking system [12]. Based on 600 repetitions of training data, it removes all known offsets, so that this model can be considered a current lower bound for touch accuracy. In comparison, the 4.9mm minimum button size of the *calibration-free* projected center model gets surprisingly close.

High-precision input with a much simpler device

In order to implement the projected center model a device needs to be able to reliably locate a user's fingernail, which is technically challenging.

Figure 21 points out an alternative. At 5.35mm minimum button size, the oc & t model does not quite reach the 4.9mm of the projected center model. However, it is comparably easy to manufacture, as this approach only requires locating finger outlines in a camera positioned above the target, namely the outlines of the sides and the top of the finger. This approach has already been explored in a number of research prototypes, such as *CSlate* [1] and *LucidTouch* [23]. Our findings thus suggest that this track of engineering, combined with sufficiently accurate cameras, could be a more promising approach to high-precision touch sensing than the currently more widely available devices based on contact area.

CONCLUSIONS

In this paper, we conducted an exploration of users' mental models of touch. Our primary contribution is scientific. The fact that under the proposed projected center model the error offsets found by our previous work essentially disappear suggests that this model is likely to closely match how users proceed while acquiring a target on a touch device.

On the other hand, our findings suggest that systems that track fingers using cameras from above have the potential for substantially better pointing accuracy than capacitive sensing as currently implemented, even simpler devices that track fingers based on finger outline alone.

REFERENCES

1. Agarwal, A., Izadi, S., Chandraker, M., and Blake, A. High Precision Multi-touch Sensing on Surfaces using Overhead Cameras. *Proc. Tabletop '07*, 197-200.
2. Albinsson, P. Zhai, S. High precision touch screen interaction. *Proc. CHI '03*, 105-112.
3. Baudisch, P. and Chu, G. Back-of-device interaction allows creating very small touch devices. *Proc. CHI '09*, 1923-1932.
4. Benko, H., Wilson, A. D., and Baudisch, P. Precise selection techniques for multi-touch screens. *Proc. CHI '06*, 1263-1272.
5. Cao, X., Wilson, A. D., Balakrishnan, R., Hinckley, K., and Hudson, S. E. Shapetouch: Leveraging contact shape on interactive surfaces. *Proc. Tabletop '08*, 129-136.
6. Feynman, R. The Messenger Series. Lecture 7: Seeking New Laws. *BBC Science&Nature.*
7. FingerWorks SDK, http://fingerworks.com
8. Fitts, P.M. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47, 381-391.
9. Forlines, C., Wigdor, D., Shen, C., Balakrishnan, R. Direct-touch vs. mouse input for tabletop displays. *Proc. CHI '07*, 647-656.
10. Grossman, T. and Balakrishnan, R. A probabilistic approach to modeling two-dimensional pointing. *TOCHI* 12, 3(2005):435-459.
11. Han, J. Y. Low-cost multi-touch sensing through frustrated total internal reflection. *Proc. UIST '05*, 115-118.
12. Holz, C. and Baudisch, P. The Generalized Perceived Input Point Model and How to Double Touch Accuracy by Extracting Fingerprints. *Proc. CHI '10*, 581-590.
13. Malik, S. and Laszlo, J. Visual touchpad: a two-handed gestural input device. *Proc. ICMI '04*, 289-296.
14. Moscovich, T. Contact area interaction with sliding widgets. *Proc. UIST '09*, 13-22.
15. Norman, D. The Psychology of Everyday Things. *Basic Books,* 1988.
16. Potter, R., Weldon, L., Shneiderman, B. Improving the accuracy of touch screens: an experimental evaluation of three strategies. *Proc. CHI '88,* 27-32.
17. Ren, X., Moriya, S. Improving selection performance on pen-based systems: a study of pen-based interaction for selection tasks. *TOCHI*, 7(3):384-416.
18. Roudaut, A., Huot, S., and Lecolinet. E. TapTap and MagStick: Improving One-Handed Target Acquisition on Small Touch-screens. *Proc. AVI '08*, 146-153.
19. Smart Tables, SmartTech. http://www.smarttech.com
20. Vogel, D. and Baudisch, P. Shift: A Technique for Operating Pen-Based Interfaces Using Touch. *Proc. CHI '07*, 657-666.
21. Wang, F. and Ren, X. Empirical evaluation for finger input properties in multi-touch interaction. *Proc. CHI'09*, 1063-1072.
22. Wellner, P. Interacting with paper on the DigitalDesk. In *Commun. ACM* 36, 7 (Jul. 1993), 87-96.
23. Wigdor, D., Forlines, C., Baudisch, P., Barnwell, J., and Shen, C. Lucid touch: a see-through mobile device. *Proc. UIST'07*, 269-278.
24. Wilson, A. D. PlayAnywhere: a compact interactive tabletop projection-vision system. *Proc. UIST '05*, 83-92