Multi-finger Chords for Hand-held Tablets: Recognizable and Memorable

Julie Wagner[1,2]
julie.wagner@lmu.de
[1]University of Munich (LMU)
Munich, Germany

Eric Lecolinet[2]
eric.lecolinet@telecom-paristech.fr
[2]Telecom ParisTech - CNRS LTCI
Paris, France

Ted Selker[3]
ted.selker@gmail.com
[3]CITRIS, University of California
Berkeley, USA

ABSTRACT

Despite the demonstrated benefits of multi-finger input, todays gesture vocabularies offer a limited number of postures and gestures. Previous research designed several posture sets, but does not address the limited human capacity of retaining them. We present a *multi-finger chord* vocabulary, introduce a novel hand-centric approach to detect the identity of fingers on off-the-shelf hand-held tablets, and report on the detection accuracy. A between-subjects experiment comparing 'random' to a 'categorized' chord-command mapping found that users retained categorized mappings more accurately over one week than random ones. In response to the logical posture-language structure, people adapted to logical memorization strategies, such as 'exclusion', 'order', and 'category', to minimize the amount of information to retain. We conclude that structured chord-command mappings support *learning*, *short-*, and *long-term retention* of chord-command mappings.

Author Keywords

multi-finger chord; chord-command mapping; finger identification; hand-held tablet

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation: User Interfaces: Interaction styles; Input devices and strategies

INTRODUCTION

Tablet computers are typically deployed to browse content rather than to perform complex data manipulation. Still, touch-enabled tablet interfaces have been replacing many uses of the traditional keyboard and mouse. To avoid complexity, application designers reduce the number of menu items when porting applications from PC to tablet environment. For example Adobe Photoshop's[1] menu is reduced from 648 menu commands on a PC to 35 commands on a tablet. Among other reasons this is done for simplification to cope in gestural interfaces with the lack of hotkeys that simplify in conventional interfaces the access to repetitive menu use.

Through hand postures and gestures, multi-touch promises more flexible input than traditional *WIMP interfaces* (Windows-Icons-Menus-Pointer). Yet, today's commercial tablet interfaces provide a limited set of input gestures. Gestural interface design for multi-touch interfaces is still emerging; we face similar challenges and goals today as designers did for keyboard-and-mouse interfaces in the 80s: (1) increase *input expressivity*, e.g. by providing keyboard shortcuts, or adding additional buttons or wheels to mouse devices; and (2) design *task-action mappings* that facilitate learning and memorization [22, 23], e.g. using mnemonics such as holding down 'CTRL' + 'C' to 'copy' a document. This paper addresses both points in the context of multi-touch enabled hand-held tablets; (1) we extend input expressivity by proposing an approach to distinguish among a set of hand postures; and (2) investigate *posture-command mappings*, addressing the limited capacity of human memory.

By taking advantage of the rich dexterity of our hands, we can obtain large sets of postures and gestures [4, 6, 15]. Gesture designers of such multi-touch technology need to work on very limited information about the user's hand posture: e.g., capacitive touch technology provides the number and position of touch, but not the finger's identity. Existing solutions have resorted to clumsy external hardware such as gloves or cameras [24], or additional time-consuming registration gestures [3]. We propose a method for recognizing a set of what we call *multi-finger chords* on off-the-shelf hand-held tablets. We use this term to describe a positioning of particular fingers on a screen relative to the hand. Similar to playing piano-chords, some fingers touch the surface, some are lifted up. Fingers are not spread or flexed; instead they remain extended in a relaxed position. This position is based on the shape of the hand and, thus, easily detectable and reproducible. We present a novel approach to distinguishing among multi-finger chords using *hand-shape characteristics* to derive a hand model.

An effective posture language, however, must also face the limited human capacity of retaining a large number of chord-command mappings. People move back and forth between mobile device, tablet, and desktop interfaces: input gestures should be easy to remember even when not constantly performed or practiced. Some previous studies proposed *natural* gesture sets [20, 30]. 'Natural' refers to a mapping between gesture and command, which is rooted in language, culture,

[1]http://www.photoshop.com/products/photoshopexpress

CHI 2014, April 26 - May 01 2014, Toronto, ON, Canada
Copyright 2014 ACM 978-1-4503-2473-1/14/04...$15.00.
http://dx.doi.org/10.1145/2556288.2556958

or experience in the physical world. Since natural gesture-command mappings use prior knowledge, it can improve the memorization of such mappings. However, abstract gesture or posture sets and abstract domain-specific commands do not have such desirable properties: *how should we design chord-command mappings in such cases?*

George Miller [18] contributed a famous insight to the understanding of information processing in human memory: organizing and grouping items into 'chunks' can increase the capacity of human memory. We investigate the effect of *grouped chord-command* mappings on memorization: similar input multi-finger chords map to similar commands. We demonstrate that structured chord-command mappings support recall even if a given gesture had not been performed for a long period of time.

RELATED WORK

Increase Input Expressivity

Knowing which user is interacting where on the surface offers a powerful means to design personalized interfaces and incorporate social protocols in interface dialogs. Some tabletop interfaces make use of built-in table cameras [25, 27, 28] or additional hardware, e.g. cameras [24] or proximity sensors in the table-frame [2], to infer touch-ownership in addition to touch-position. However, to increase expressiveness of a single user, we are also interested in techniques to provide information beyond simple touch.

Previous work proposes a number of techniques to make touch more distinctive; Finger-count [5] uses the number of touches; MicroRolls [26] detects specific patterns in the trajectory of touch events while users perform small roll motions with their fingers; and SimPress [9] analyses the fingers contact area. Wang et al. [28] used constraints due to the anatomy of the human hand to identify which touches belong to the same hand. However, none of these approaches addresses *touch-to-finger* ownership.

One simple approach is the *Lift-and-Stroke* technique [15]; users place all five fingers of their hand on the surface and then lift the ones not required for a given chord. Unfortunately lifting certain fingers while simultaneously holding others down is difficult to perform [15]. Similarly, Au et al. [3] proposed a technique that requires the registration of all fingers; first, users hold down all fingers; instead of lifting some fingers, users lift the whole hand and then select items from the appearing on-screen menu. However, their approach requires visual attention and might be impractical in cases where the attention is focused on external devices, e.g. large displays.

On multi-touch tables, the built-in camera can infer the finger-ownership of touch from the hand's halo [15] by analyzing – a shadow casted by hands seen by the camera – where the touch event occurs relative to the hand. In fact, most existing approaches for user-, hand-, and finger- identification [29] require external hardware. We contribute an approach for finger-identification for off-the-shelf tablets. We present a basic posture vocabulary of nine postures and propose space- and time-multiplexed ways for extension.

Mappings in Interfaces

Compared to the diversity of investigations studying ways of augmenting mere touch input and proposing gesture sets [4, 15, 20, 30], there is relatively little literature on the choice of mapping between gestures and commands. Challenges in gesture interfaces comprise (1) discoverability and (2) memorability. Discoverability had been addressed, e.g., by feed-forward systems such as *OctoPocus* [7] or *Arpège* [13], that guide the user with visual clues to perform gestures or postures.

Memorization is addressed in the context of 'natural' gestures. Wobbrock et al. [30] applied user-elicitation methods to gather gestures with large agreement across users. They found that participants preferred user-authored gestures over those created by HCI experts [20]. Natural gestures take advantage of pre-established associations which support memorization, and even more so when those associations are personally assigned by the user [21]. However, in the absence of such cultural references, linguistic associations, or metaphorical links, *how should abstract posture sets be integrated into the interface so that they become easy to use?*

We need to account for the lack of 'natural' associations of postures. Previous work showed that the method of *organizing* menu structures has an effect on visual search performance [17]. Indeed, we are used to virtual information being structured: we group tools in palettes, categorize menu items by meta-attributes, and save documents in tree-structures. Similarly Buxton [11] presents examples of grouping input gestures in human-computer dialogs into meaningful units following two aspects: (1) gestures have a kinaesthetic connection that match the logical structure of two *tokens* and (2) the choice of the follow-up gesture is limited to a single choice reducing cognitive load. Users are forced to follow a proper syntax, e.g. select then position. Yet, we lack sufficient guidance for applying it to the design of posture sets.

In psychology, organization is assumed to be a necessary condition for memory [16]. Most gesture work in HCI investigated short-term retention of gestures [6, 7]. However, we agree with previous discussions [1, 21] that gesture sets should also be studied with respect to long-term recall in order to truly understand gesture memorability. In the following sections, we introduce a novel multi-finger chord vocabulary for off-the-shelf tablets. Postures of our vocabulary can be categorized into three families of postures due to similarities among input movements. We then explore the effect of categorized chord-command mappings on long-term memorization: *does it improve long-term retention if the structure of performed input movements matches the virtual menu structure?*

DESIGNING A MULTI-FINGER CHORD VOCABULARY

Multitouch-enabled technologies, can detect the location and number of touch contacts with our hands. Without additional technology, however, they cannot identify the specific finger, which is touching the surface. Two-finger gestures have become pervasive in tablet gesture languages. But two fingers cannot suffice for identifying fingers reliably because they only define a segment of a line. The hand has a polygonal

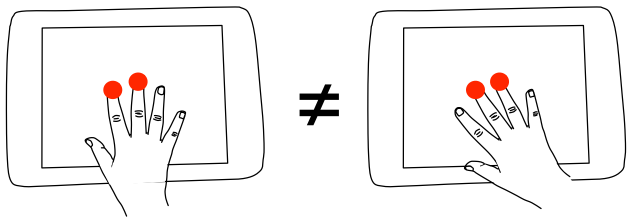

Figure 1. Minimum requirement of three touches: an example of two identical touch events triggered by two hand postures.

shape whose vertices are the fingertips. Depending on the orientation of the hand (see Fig. 1), a two-finger segment can match many segments of the hand-shape polygon. By contrast, three fingers define two segments which relative geometric properties guarantee a unique match in well chosen-cases. At least three fingers are needed for the simplest determination of the hand posture of the hand. Three fingers determine postures for a reasonably large set of noncontorted postures. The hand could curl fingers under or raise one up to increase the size of this vocabulary, but this would generate variability and decrease accuracy. We would require a more complex model of the hand to be accurate.

In the following, we present hand measurements for a classifier enabling us to identify the fingers performing a chord. We present three observations (O1, O2, O3) of human hand-shape characteristics [19] and simple ways of measuring them. These measurements are entirely based on *relative* measurements designed to be insensitive to variations in the actual size of users' hand.

We propose three *families* of simple-to-sense, hand-size insensitive input postures: (1) *Neighboring Fingers (NF)*, (2) *Thumb-Pinky Base (T-Pinky)*, (3) *Thumb-Index Base (T-Index)*. In the process of designing classifiers for the posture families, each hand-shape characteristic contributes one relative measurement that distinguishes that particular posture in its posture family. Used together, these angle and geometric measurements can distinguish all postures. In the following sections, we introduce the posture families and summarize the rationale behind our design decisions.

O1: Relative Distance between Neighboring Fingers

One observation is that index, middle, and ring finger have a relative position to their neighboring fingers; e.g. the index finger (see NF[INDEX] in Fig. 2) is positioned towards the middle finger, thus more distant to the thumb (D2 < D1). We call the three chords in Figure 2 elements of the *neighboring fingers (NF)* family. We suggest the relation between *D1* and *D2* as relative measurement independent of actual hand sizes.

O2: Relative Length of Fingers

The *thumb-index base (T-Index)* family addresses the observation that fingers of our hand have common patterns in length; e.g., the middle finger is usually longer than the index finger and the pinky is usually shorter than the ring finger. We suggest the *angle* between (1) the line going through thumb and index basis and (2) the line between index and the third

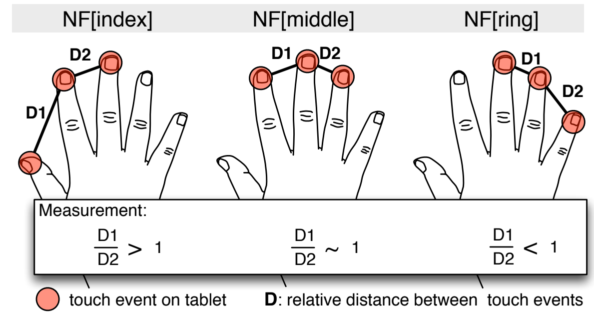

Figure 2. Neighboring fingers (NF) family: relative distances are different between the three touches.

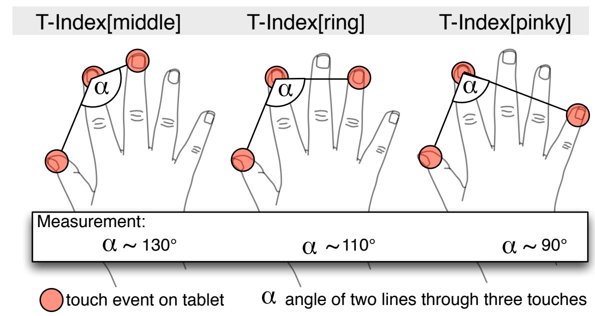

Figure 3. Thumb-index base family (T-Index): relative angle to the basis between thumb and index shrinks from middle to pinky finger.

identifying finger as relative measurement to identify the appropriate posture (see Fig. 3).

To distinguish T-INDEX[MIDDLE] from NF[INDEX], we designed this as a two-step posture: users have to simultaneously hold down the thumb-index basis first and then add the third finger touch in a second step. Compared to previous approaches, e.g. holding down and lifting up all five fingers [31], steps are rapidly performed in sequence ($\approx$ 150ms apart). Moreover, this process highlights the structure of the command and the fact that it relies on the completion of the thumb-index basis.

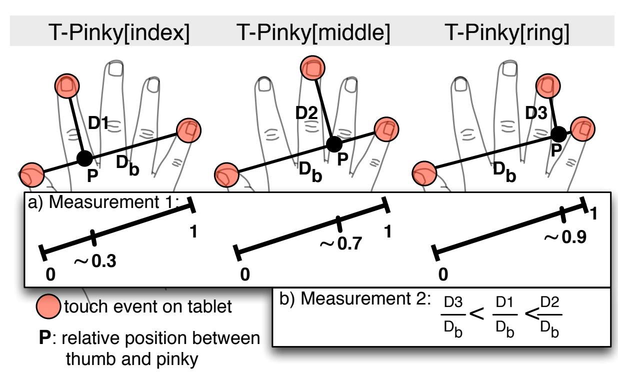

Figure 4. Thumb-Pinky basis family (T-Pinky): a) index, middle and ring finger are at a common relative position between thumb and pinky. b) relative distances between basis fingers and third finger are different.

O3: Order of fingers

The *thumb-pinky base (T-Pinky)* family is based on the observation that index, middle and ring finger are attached to the palm in a consistent order. We suggest to measure (1) the relative position (P) of each finger on the basis line between thumb and pinky (see Fig. 4a) and (2) the distance of 'P' to the basis line in relation to the length of the basis line (see Fig. 4b). The relative position 'P' is calculated by an orthogonal projection of the third touch to the basis line.

To distinguish T-PINKY[INDEX] from T-INDEX[PINKY], the T-Pinky family is designed as a two-step posture as well: users first hold simultaneously down thumb and pinky as a basis and then add the third finger touch in a second step.

Design Rationale

The proposed postures and families have the following contributions, limitations and characteristics:

Structure:
Each family follows a simple logic that is easy to understand and to communicate.

Technical Limitations:
Multi-touch sensors are capable of providing information about touch shape. However, many commercial available tablet APIs, e.g., Apple's IOS 7[2], or tablet providers, e.g., Samsung Galaxy tab 10.1[3], do not provide such information to developers. We present an approach that works on very limited data – the position of touch – to distinguish various multi-finger chords on all off-the-shelf hand-held tablets.

Minimal Requirement:
Each posture requires pressing at least three fingers if the hand-held tablet provides only the location of touch contacts. The positional information of two touch events is not sufficient to infer the hand posture (note example in Fig. 1).

Limitations of our Approach:
Our relative measurements are invariant to a limited range of rotation of the users hand relative to the interactive surface (between 0°and 90°counter clockwise). This is an acceptable *rotational interaction range* for hand-held tablets since it would be hard to work from other angles. Further investigation could evaluate the addition of using a built-in camera for finger orientation recognition [28] in combination with our approach for use on interactive tabletops.

The approach is limited to a single hand since some postures are symmetric: e.g. holding down the right hands index, middle, and ring finger at the same time is – from a system's perspective – like using the left hand. This is also an acceptable condition for hand-held tablets, since one hand is involved in the device support and not available for such input.

Extensibility:
We present postures that meet the minimum requirement of three touches to identify a hand posture and identify the touching fingers. However, since our algorithm computes an approximate position of the remaining fingers, we can extend our simple set of postures in two ways: (1) holding more fingers down; postures with 4-5 fingers are a super set of the above-introduced postures: all possible 4-finger postures include one of the three-finger chords already detected; or (2) designing tapping sequences with the remaining fingers which extends the chord set with *multi-step chords*. Using (1), we can extend our 9 three-finger postures by 1 five-finger and 5 four-finger postures (sum of 15). In order to investigate a final posture-set size when using (2), further studies are required to determine a suitable tapping sequence depth.

In summary, we introduced relative measurements invariant of users' hand size and contribute a posture set that requires the user to hold down a minimum of three fingers to identify the involved fingers and approximate the position of the remaining hovering fingers. We introduced ways of extending our set. The classifier is available for download[4].

STUDY ONE: RECOGNITION ACCURACY

We collected data from 20 participants with varying hand size performing all postures. We created a *KNN* classifier [12] and analyzed its accuracy. We conducted several tests on our data in order to address various real-world settings with tablets: (1) *private*, (2) *shared*, (3) *public* device setup.

In a *private* device setup, only one user interacts with the device; a prior calibration process can collect user-specific data. In a *shared* device setup, e.g., when used in multi-surface environments as controlling input device [8], each user can perform a prior calibration; it is, however, unknown at a given time which user is interacting on the device. Lastly, in a (3) *public* device setup tablets are used in public places and for short period of time, e.g., in museum installations. Users walk up and use the tablet without prior calibration or personalization process.

Participants
20 right-handed volunteers participated (13 males, average age 29 years). Their mean (± SD) anthropometric hand characteristics were: hand length 19.19 (± 1.57 cm) and hand width 9.22 (± 0.76 cm). The hand length was measured between the middle fingertip and the distal crease of the wrist with the hand extended, and the hand width was measured between the lateral aspect of the index and pinky finger at the joint where the finger is attached to the palm [19]. 11/20 participants practiced activities that train finger dexterity, e.g., playing instruments.

Apparatus
Samsung Galaxy tab, 10.1 inch × 6.9 inch × 0.3 inch depth, 565g, 149 ppi, Android version 4.0.4. The *JavaML* library [1] was used to create a *KNN* classifier (K=3) to distinguish hand postures.

Procedure:
The experiment lasted approximately ten minutes. When participants arrived, they filled out a questionnaire inquiring about habits in interacting with capacitive-touch technology and activities that could train hand dexterity, e.g., playing instruments. We measured the width and height of the hand of participants and captured the outline of their flat hand on millimeter paper. They were instructed to keep their hands in

[2]http://www.apple.com/ios/
[3]http://www.samsung.com/global/microsite/galaxytab
[4]`http://www.medien.ifi.lmu.de/multifingerchord`

vocabulary	private tablet	shared tablet	public tablet
Neighboring finger (NF)	96.67%	96%	93.67%
Thumb-Index (T-Index)	100%	92%	83%
Thumb-Pinky (T-Pinky)	99%	98%	91%
NF + T-Index	98.33%	94%	88.33%
NF + T-Pinky	97.66%	96.83%	91.5%
T-Pinky + T-Index	99.5%	95%	88.67%
NF + T-Pinky + T-Index	98.44%	95.22%	91.5%

Table 1. The mean accuracy of input vocabularies in private, shared and public device setup.

a relaxed posture with neither pinched nor out-stretched fingers. We conducted a 3 POSTURE FAMILY × 3 POSTURE ID within subjects experiment. We collected a total of:

(NEIGHBORING FINGERS (NF) × [INDEX, MIDDLE, RING] +
THUMB-INDEX BASIS (T-INDEX) × [MIDDLE, RING, PINKY] +
THUMB-PINKY (T-PINKY) × [INDEX, MIDDLE, RING])
× 5 replications = 45 trials/participant.

Trials were blocked by POSTURE FAMILY and the order was altered using a Latin square design on the POSTURE FAMILY factor. The order of POSTURE ID was randomized and no two equal postures appeared successively, to avoid the repetitive input of the same posture by remaining in a stiff hand position. In each trial, users were instructed to perform the appropriate posture indicated by an instruction image on the upper left screen corner. The experimenter and a video camera verified that the correct posture was performed.

Training:
Before each block of POSTURE FAMILY, users performed each posture once in a training block.

Data collection:
We collected position and size of all registered touch events and camera video showing the subject performing the tasks.

Data post-processing:
The three finger-touches form a triangle: we extracted the four relative measures illustrated in Figure 2-4 for each of the three triangle-sides or corners. A total of 12 measurements (3 corners × 4 measurements) were used for the KNN classifier.

Results and Discussion: Recognition Accuracy

We used a *m-fold cross validation* procedure with the appropriate size of 'm' for each tablet setup: a cross validation partitions a sample of data into a *training* set and *test* set and validates the test set against the training set; each part of the data becomes a test set once and the accuracy result of all validations are averaged. Table 1 present the average accuracy value by device setup and vocabulary. Figure 5 illustrates the distribution of accuracy by vocabulary.

Private tablet setup - 5-fold cross validation by participant:
We determined one mean accuracy value by participant and took the average across participants. As expected, the private tablet setup has the most robust detection; most precise are small vocabulary sizes, NF (Mean = 96.67%), T-INDEX (Mean = 100%), and T-PINKY (Mean = 99%); however, the accuracy rate of all 9 postures is acceptable (98.44% which is > 95%).

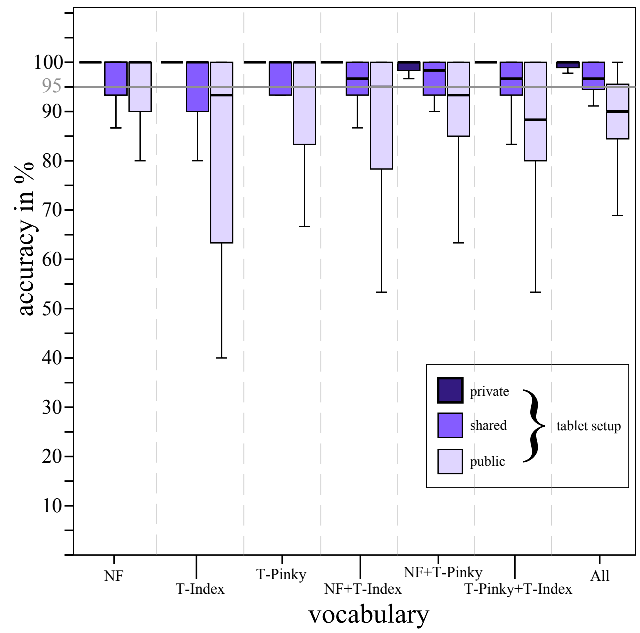

Figure 5. The distribution of classifier accuracy (in %) by vocabulary across participants in private, shared and public tablet setup.

Shared tablet setup - 20-fold cross validation:
This corresponds to the case where users share the tablet with others but first performed an initial calibration. Although less accurate than the private tablet setup, the global accuracy rate remains reasonably high (95.22% which is >95%). Interestingly, the T-INDEX family is the most accurate in the private case but the least accurate in the shared case (see Tab. 1) compared to both other families; T-INDEX has a *low within-subject variability* and a *high between-subject variability*. This may be because some participants tended to perform pinched positions of the thumb and index finger instead of keeping their hand relaxed.

Public tablet setup - 20-fold cross validation:
The NF family obtains the best result in this case (see Tab. 1). It is interesting to notice that its performance is almost the same for the three setups, showing a low between-subject variability. However, the global accuracy rate (91.5%) is probably not sufficient for real usage.

In summary, our classifier is accurate enough to apply the tested nine posture vocabulary in private (98.44% accurate) and shared tablet setups (95.22% accurate). The touch-positional data is, however, not sufficient for a public tablet setup (91.5%). Future work can include, e.g. camera halo or increased sophisticated hand models, to enhance the accuracy of the public tablet setup. In addition, increased data on touch orientation might help as well if made available on a wide range of off-the-shelf hand-held tablets [28].

STUDY TWO: MEMORIZING INPUT VOCABULARY

Structure improves memorability [16]. Previous findings in psychology have demonstrated that people can recall more items, e.g., 'dog' and 'table', if those items are presented

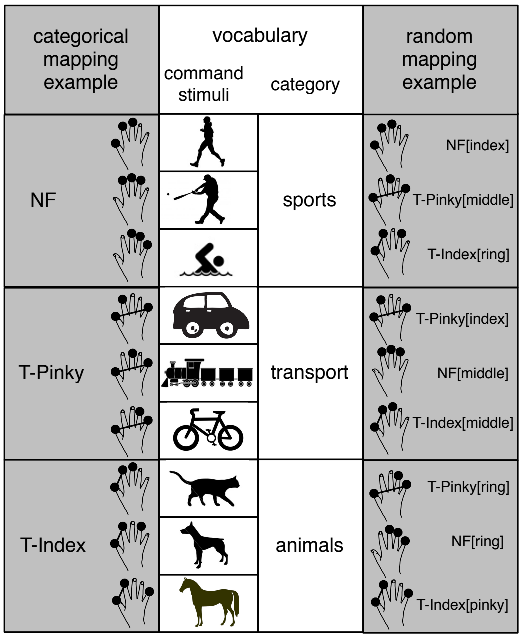

Figure 6. Participants learned 9 commands divided into 3 command-categories using either a (1) categorical mapping (left) (chord family to category mapping) or (2) random chord-command mapping (right).

grouped by category, e.g., ‘animal’ and ‘furniture’ [10]. Inspired by this finding, we investigate whether such a categorical structure can facilitate the learning and long-term retention of chord-command mappings.

Hypothesis

If a multi-finger chord language is structured in such a way that it reflects the menu-structure of commands, users (H1) learn the chord-command mappings faster and (H2) keep those mappings more accurately in mind over a long period of time. We test our hypothesis in the concrete case of mapping gesture families to command categories.

Participants:

18 right-handed participants, 13 males (avg. age 27 years, SD=8 years), paid with a 10€ gift card. None of these subjects had participated in the first study.

Apparatus:

Samsung Galaxy tab (as in study one) running our posture recognizer.

Method

We performed a between-subjects design. Subjects were randomly assigned to two groups: (1) subjects taught CATEGORICAL associations, the other (2) subjects taught RANDOM associations. All participants were instructed to learn nine commands organised in three command-categories: *transportation*, *animals*, and *sports*. Each category has three commands, e.g. transportation has *car*, *train*, and *bike*.

Stimulus:

We chose black-and-white illustrations for the nine commands (see Fig. 6, *’command stimuli’*).

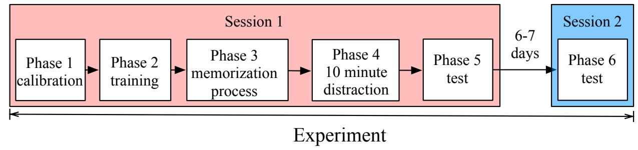

Figure 7. The different phases of the experiment in two sessions with 6-7 days in between two sessions.

CATEGORICAL

Each chord family (Neighboring Fingers, Thumb-Pinky, Thumb-Index) is mapped to one command-category (see Fig. 6, grey right column as one example).

RANDOM

Items were randomly assigned to one of the nine postures such that none of the command-categories matches one chord family (see Fig. 6, grey right column as one example).

Procedure

The experiment was divided into two sessions and lasted in total approximately one hour. Figure 7 shows that session 1 has 5 phases; session 2 has only one test phase and takes place 6-7 days later. All participants were identically introduced to the chord vocabulary and the set of commands through an oral introduction of the experimenter.

Phase 1: calibration

Users performed the same trials as in the first study for calibration purpose. A per-user classifier (private tablet setup) was created per participant.

Phase 2+3: training and memorization process

The goal of Session 1 was to make participants learn nine chord-command mappings by heart. In Phase 2 and 3, each *block* contained 9 trials, each eliciting one element of the vocabulary in random order. Each block is one repetition of all items in the vocabulary. A *trial* had the following procedure: participants hold down a button until a stimulus image appeared showing one of the command stimuli in Figure 6. Participants then performed the appropriate chord that maps to the particular stimulus. They received feedback about errors: the screen flashed red and an instruction image of the correct chord appeared. Chord-command mappings were different for all participants of both groups. Participants learned a novel posture set, new command symbols and the mappings between them; to reduce the learning effort to learning mappings, we showed two sheets of paper to the user: one illustrated all hand postures, another one showed all commands. The chord-command mapping was not shown.

Training (phase 2):

Participants performed two blocks. To introduce the mapping, each trial displayed two instruction images: the stimulus showing the command and the corresponding chord. Participants had to perform the correct chord to finish the trial.

Memorization (phase 3):

Trials only displayed the stimulus and asked to recall the appropriate chord. Participants continued to receive feedback about errors and the correct chord instruction image. The memorization ended when participants both reached an *objective* and a *subjective* end-criterion.

Participants reached the *objective end criterion* of the memorization phase when they successfully reproduced all chords of the vocabulary twice in sequence (measured as two blocks without errors). Participants reached the *subjective end criterion* when they decided that they trained enough to be able to reproduce the command-gesture mapping 6-7 days later in session 2.

Phase 4: Distraction
All participants watched a 10-minute cartoon.

Phase 5+6: Test
Participants performed one block as in *phase 2+3* all items of the vocabulary were asked once; participants had to perform the corresponding chord of their mapping for a given stimulus. Phase 5 in session 1 tested the *short-term retention* and phase 6, six or seven days later, tested the *long-term retention*. No feedback was provided. Sessions were video recorded to eliminate recognition errors. Chord-command mappings were validated post-hoc by analyzing the video.

Data Collection
Reported misclassification: participants received error feedback during phase 2+3 and were instructed to report recognition errors.

Objective end-criterion: number of required blocks until participants recalled correctly all items twice in sequence. Objective end-criteria are always reached before subjective end-criteria.

Subjective end-criterion: number of required blocks until user subjectively feels 'trained enough to successfully recall the vocabulary one week later'.

Number of errors: the number of chord-command mappings participants did not recall correctly.

Trial time: time from the moment a stimulus appeared to the performance of the corresponding chord.

Results and Discussion
We analyzed our data using the IBM SPSS statistics package version 21. We performed one-way ANOVA tests on the normally distributed subjective and objective end-criterium data, and the non-parametric Mann-Whitney test on number of errors due to a significant Levene's test.

Classifier Precision
During the training and memorization process (phase 2+3), we collected a total of 2421 trials, in average 134 trials by participant (Median=126 trials, SD=53.15 trials). Participants reported on average 7 *misclassifications*; one user reported 22 misclassifications with an accuracy of 79.63%. However, this was a particular case: this participant had a long thumb that triggered several touch events. This detection issue can easily be addressed by updating our hand model to ignore such touches.

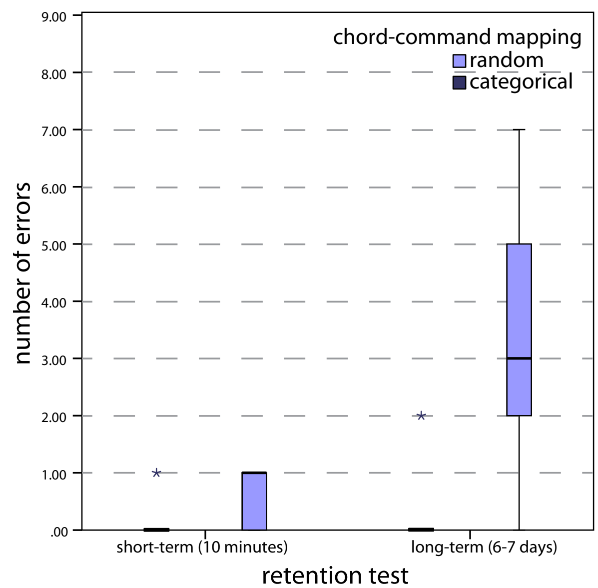

Figure 8. The distribution of number of errors during a short-term (after 10 minutes) and long-term (after 1 week) test. Participants in the CATEGORICAL group made no errors except two outliers (green stars) in both retention tests, and remembered chord-command mappings better than the RANDOM group.

Other participants had between 1 and 16 misclassifications; the classifiers accuracy was between 81.48% and 99.35% (Mean=94.06%, SD=5.07%) for a private tablet setup. Our classifiers accuracy decreased from 98.44% in the first study to 94.06% in the second study; the cognitively demanding caused probably that participants tensed up their hands.

H1: Do people learn chord-command mappings faster?
A one-way ANOVA of *subjective end-criterion* showed no significant difference between groups ($F_{1,16} = 1.4, p = 0.254$): participants in the CATEGORICAL group required in average 10.33 blocks (SD=1.11 blocks) and participants in the RANDOM group required in average 12.67 blocks (SD=1.63 blocks).

A one-way ANOVA of *objective end-criterium*, however, showed a significant difference between groups ($F_{1,16} = 4.498$, $p = 0.05$): participants in the CATEGORICAL group required less repetition to reach the objective end-criterium (Mean = 7 blocks, SD = 3.81 blocks) compared to participants in the RANDOM group (Mean = 11.44 blocks, SD = 5 blocks). The CATEGORICAL mapping accelerated the learning and successful repetition of the vocabulary. Both groups, however, continued the training. To make our findings stronger, it was particularly important that participants end the training with the confidence to recall the items in later tests.

H2: How well can participants retain mappings?
The *number of errors* data is not normally distributed: a Levene's test shows that the variance of *number of errors* was significantly different between groups for short-term retention ($F_{1,16} = 11.18$, $p = 0.048$) and long-term retention ($F_{1,16} = 46.72, p < 0.0001$).

We performed a non-parametric Mann-Whitney test on *number of errors* in both retention tests. We found no significant difference between groups in the short-term retention test, 'p' being slightly greater than 5% ($U = 22.5$, $z = -1.94$, $p = 0.052$, $r = -0.46$): 1 participant in the CATEGORICAL group did one error; and 5 participants in the RANDOM group did one error; the remaining had no errors. There might be a significant different result with a larger sample size.

We found a significant difference between groups in the long-term retention test: participants in the CATEGORICAL group (see Figure. 8) made significantly less errors when tested six to seven days after the training than participants in the RANDOM group ($U = 6.5$, $z = -3.21$, $p < 0.001$, $r = -0.76$). We were surprised how well participants in both conditions retained nine mappings for approximately one week without further training or practice; especially, since previous work indicate difficulties in learning more than two abstract gestures [14].

Figure 8 shows the distribution of *number of errors* during the short- and long-term retention test. Participants in the CATEGORICAL group made in average. 0.11 errors (Median = 0, SD = 0.33) in the short-term retention test and 0.22 errors (Median = 0, SD = 0.67) in the long-term retention test 6-7 days later. Participants in the RANDOM group made in average 0.56 (Median = 1, SD = 0.53) errors in the short-term retention test and 3.4 (Median = 3, SD = 2.3) errors in the long-term retention test. Note that not only the mean values are different between groups in the long-term retention test, but also the variance (0.44 vs. 5.28).

We analyzed the errors and categorized them into three types of errors: (1) *right family wrong finger* (RFWF), (2) *wrong family right finger* (WFRF), and (2) *completely wrong* (CW). Participants in the CATEGORICAL group had few errors: P2 had 1 RFWF in the short-term retention test and P6 had 2 RFWF in the long-term retention test; he swapped two postures of the same family. Participants in the RANDOM group had in comparison lots of errors, a total of 5 in the short-term test and 36 in the long-term retention test. We found 2 WFRF, 2 RFWF, and 1 CW in the short-term retention test and 8 RFWF, 9 WFRF, and 19 CW errors in the long-term retention test. Our results indicate that a structured mapping leads to less error-prone long-term memorization, which is highlighted by the types of errors participants made: participants in the CATEGORICAL group did not mix up mappings between families and menu category, and did not perform a completely wrong posture.

Performance

We ran a full-factorial ANOVA on *trial time* using the factors *group*, *retention test*, and *chord family*. We found a significant effect of group ($F_{1,312} = 10.95$, $p < 0.0001$), chord family ($F_{2,312} = 6.05$, $p = 0.003$), and retention test ($F_{1,312} = 17.69$, $p < 0.0001$).

Participants in the CATEGORICAL group performed trials significantly faster (Mean = 2572ms, SD = 1940ms) than those in the RANDOM group (Mean = 3330ms, SD = 2327ms). We analyzed the difference between posture families: a post-hoc Tukey test revealed that postures in the *neighboring fingers (NF)* family were performed faster (Mean = 2416ms, SD = 1834ms) than *thumb-pinky basis (T-Pinky)* (Mean = 3062ms, SD = 1770ms). *Thumb-index basis (T-Index)* was not significantly different from both families (Mean = 3373ms, SD = 2696ms).

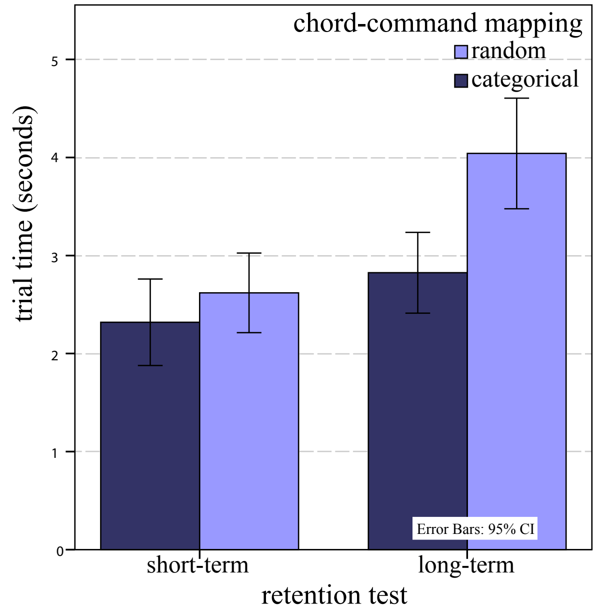

Figure 9. Interaction effect: trial time was not significantly different in the short-term test, but significantly different in the long-term retention test.

All participants required more time to perform postures in the long-term retention test (Mean = 3433ms, SD = 2308ms) than in the short-term retention test (Mean = 2469ms, SD = 1917ms). However, we found a significant interaction effect between *group* and *retention test*: participants in the CATEGORICAL group performed not significantly faster in the short-term retention tests than the RANDOM group (Mean = 2319ms, SD = 1994ms vs. Mean = 2619ms, SD = 1839ms); however, in the long-term retention test, participants in the CATEGORICAL group performed significantly faster (Mean = 2824ms, SD = 1863ms) than participants in the RANDOM group (Mean = 4041ms, SD = 2549ms). This supports ongoing discussions that some effects on memorization might first show up after some time has passed [21, 1].

Qualitative Evaluation

We found that most participants preferred the NF family (12/18) compared to the T-PINKY (4/18) and T-INDEX(2/18) family. This result is consistent with participants rating of postures with respect to both perceived *comfort* and *ease-of-use*. Figure 10 illustrates participants' rating of *ease-of-use* on a 5-point Likert scale (1-very difficult, 5-very easy).

We found an interesting effect between groups on the qualitative rating of participants. All participants performed the same multi-finger chords, only the mapping changed. We found no effect between groups on participants ratings of *perceived comfort* of postures ($U = 2870$, $z = -1.476$, $p = 0.140$, $r = -1.12$). However, when asked about perceived *ease-of-use* – directly below the question about comfort – participants answered more positive in the CATEGORICAL condition than in the RANDOM condition ($U = 2609$, $z = -2.47$, $p = 0.014$, $r = -0.19$). Figure 10 illustrates the difference by *chord fam-*

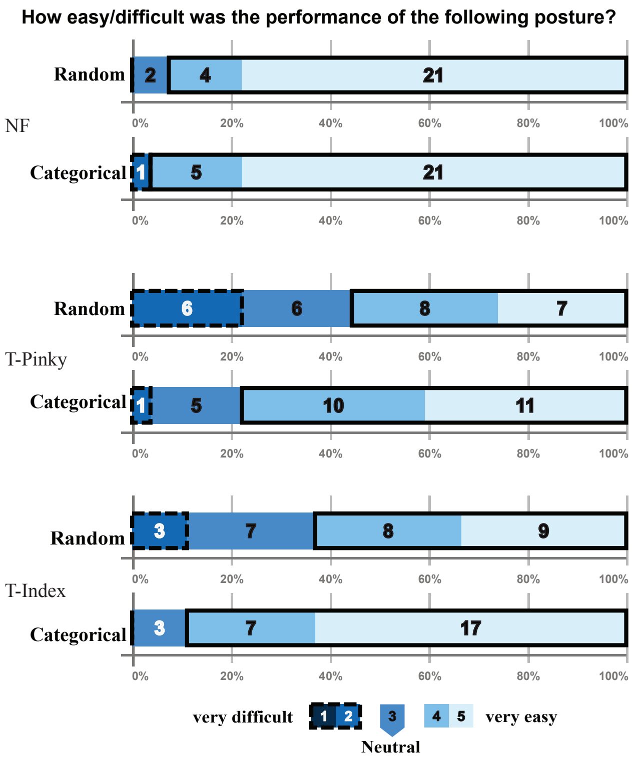

Figure 10. Participants ranking of ease-of-use on a 5-point Likert scale for all three chord families by condition: neighboring fingers (NF), thumb-pinky basis (T-Pinky), and thumb-index (T-Index). Ratings were more positive in the categorized group.

ily: we see that the solid black box around positive answers is larger for CATEGORICAL and the dashed black box around negative answers is larger for RANDOM (except for the NF chord family where answers are equally positive).

Participants equally rated comfort of postures independent of group. This is not surprising, since they performed the same postures; it is, however, interesting to note that participants seem to rate to above demonstrated quality of the mapping as *easier to use*. This might be explained through the *integrality of posture and mapping*: the demonstrated advantages of the CATEGORICAL mapping might have led to more positive rating of perceived ease-of-use.

Memorization Strategies

Participants reported on their strategies to remember mappings. We classified answers into five types of memorization strategies summarized in table 2. 'Metaphors' were the only strategy we found represented in both groups. 'Exclusion', 'order', and 'category' were used in the CATEGORICAL group; 'practice' in the RANDOM group. The introduction of structure leads to logical and ordered memorization strategies in the CATEGORICAL group. 5/9 participants mentioned 'exclusion' as strategy to minimize the amount of information to retain.

CONCLUSION

This paper presents a novel approach using relative measurements of touch-positions between three fingers to determine the hands' posture over off-the-shelf hand-held tablets. We

strategy	group	participants	example
Metaphors	RAND.	3/9	*[I remembered all vehicles] from left to right [Finger] by what I can afford to use: bike, train, [own] car.* (P8)
	CATEG.	7/9	*I found similarities between the pictures and the chord []: the dog stands pointing to the left, the chord is NF[index].* (P15)
Exclusion	RAND.	0/9	–
	CATEG.	5/9	*I remembered two chords and deducted the third mapping.* (P7)
Order	RAND.	0/9	–
	CATEG.	5/9	*I remembered the finger-order for a given menu category.* (P3)
Category	RAND.	0/9	–
	CATEG.	7/9	*"I first memorized which family is in which category []"* (P1)
Practice	RAND.	4/9	*"[I] learned by repeating it."* (P14)
	CATEG.	0/9	–

Table 2. Five memorization strategies of participants: metaphors were used by both groups; participants in the categorical group used logical strategies (exclusion, order, category); participants in the random group relied on practice.

present nine postures that meet the minimum requirements for calculating a hand model and discuss further possibilities for extending the vocabulary. We demonstrate that the accuracy of our gesture recognizer is acceptable for *private* (98.44%) and *shared* tablet setups (95.22%). It remains future work to improve the detection accuracy in *public* tablet setups (91.5%).

Our posture vocabulary introduces a categorical structure of physical input movements to access items in a menu; similar input movements access similar commands. This introduces an organization on two levels: (1) the posture vocabulary leverages a hierarchical order and (2) the corresponding commands are hierarchical structured. We found that a homomorph structure between the input posture set and corresponding command set leads to increased long-term retention of chord-command mappings.

We compared *categorical* and *random* mappings in a between-subjects experiment. We found that participants in the categorical group learned quicker; they also retained a 9-item chord-command vocabulary with fewer errors for the time period of one week. Moreover, people in the categorical group performed trials faster and rated 'ease-of-use' of the identical posture set more positively than people in the random group. The logical input structure we introduced to facilitate memorization has a consequence on participants' strategies to cope with the memorization task: participants adapted their strategies correspondingly to *order*, *category* and *exclusion*, minimizing the amount of information to retain.

This research demonstrates that large posture vocabularies can be learned and memorized. Multi-finger gestures can be recognized by standard multi-touch technology and such gesture-languages can be larger than previously thought. This opens new ways to introduce lanuguages for multi-touch tablets.

REFERENCES

1. Anderson, F., and Bischof, W. F. Learning and performance with gesture guides. In *CHI '13*, ACM (2013), 1109–1118.
2. Annett, M., Grossman, T., Wigdor, D., and Fitzmaurice, G. Medusa: a proximity-aware multi-touch tabletop. In *UIST '11*, ACM (2011), 337–346.
3. Au, O. K.-C., and Tai, C.-L. Multitouch finger registration and its applications. In *OZCHI '10*, ACM (2010), 41–48.
4. Bailly, G., Demeure, A., Lecolinet, E., and Nigay, L. MultiTouch menu (MTM). In *IHM '08*, ACM (2008), 165–168.
5. Bailly, G., Lecolinet, E., and Guiard, Y. Finger-count & radial-stroke shortcuts: 2 techniques for augmenting linear menus on multi-touch surfaces. In *CHI '10*, ACM (2010), 591–594.
6. Bailly, G., Lecolinet, E., and Nigay, L. Flower menus: a new type of marking menu with large menu breadth, within groups and efficient expert mode memorization. In *AVI '08*, ACM (2008), 15–22.
7. Bau, O., and Mackay, W. E. OctoPocus: a dynamic guide for learning gesture-based command sets. In *UIST '08*, ACM (2008), 37–46.
8. Beaudouin-Lafon, M., Huot, S., Nancel, M., Mackay, W., Pietriga, E., Primet, R., Wagner, J., Chapuis, O., Pillias, C., Eagan, J., Gjerlufsen, T., and Klokmose, C. Multisurface interaction in the WILD room. *Computer 45*, 4 (2012), 48–56.
9. Benko, H., Wilson, A. D., and Baudisch, P. Precise selection techniques for multi-touch screens. In *CHI '06*, ACM (2006), 1263–1272.
10. Bower, G. H., Clark, M. C., Lesgold, A. M., and Winzenz, D. Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior 8*, 3 (June 1969), 323–343.
11. Buxton, W. An informal study of selection-positioning tasks. ACM (1982), 323–328.
12. Fix, E., and Hodges, J. L. Discriminatory analysis. nonparametric discrimination: Consistency properties. *International Statistical Review / Revue Internationale de Statistique 57*, 3 (Dec. 1989), 238–247.
13. Ghomi, E., Huot, S., Bau, O., Beaudouin-Lafon, M., and Mackay, W. E. Arpège: Learning multitouch chord gestures vocabularies. In *ITS '13*, ACM (2013), 209–218.
14. Jego, J.-F., Paljic, A., and Fuchs, P. User-defined gestural interaction: A study on gesture memorization. In *2013 IEEE Symposium on 3D User Interfaces (3DUI)* (2013), 7–10.
15. Lepinski, G. J., Grossman, T., and Fitzmaurice, G. The design and evaluation of multitouch marking menus. In *CHI '10*, ACM (2010), 2233–2242.
16. Mandler, G. Organization and memory. *The psychology of learning and motivation: I* (1967), 381.
17. McDonald, J., Stone, J., and Liebelt, L. Searching for items in menus: The effects of organization and type of target. *Proc. of the Human Factors and Ergonomics Society Annual Meeting 27*, 9 (Oct. 1983), 834–837.
18. Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review 63*, 2 (Mar. 1956), 81–97.
19. Missile, U. S. A. *Human Engineering Design Data Digest*. Human Engineering Laboratory Detachment, 1984.
20. Morris, M. R., Wobbrock, J. O., and Wilson, A. D. Understanding users' preferences for surface gestures. In *GI '10*, Canadian Information Processing Society (2010), 261–268.
21. Nacenta, M. A., Kamber, Y., Qiang, Y., and Kristensson, P. O. Memorability of pre-designed and user-defined gesture sets. In *CHI '13*, ACM (2013), 1099–1108.
22. Payne, S. J., and Green, T. R. G. Task-action grammars: A model of the mental representation of task languages. *SIGCHI Bull. 19*, 1 (July 1987), 93–133.
23. Peter, A. L., Foltz, P. W., and Polson, P. G. Memory for task-action mappings: Mnemonics, regularity, and consistency. *International Journal of Human-Computer Studies 40* (1994), 771–794.
24. Ramakers, R., Vanacken, D., Luyten, K., Coninx, K., and Schöning, J. Carpus: a non-intrusive user identification technique for interactive surfaces. In *UIST '12*, ACM (2012), 35–44.
25. Richter, S., Holz, C., and Baudisch, P. Bootstrapper: recognizing tabletop users by their shoes. In *CHI '12*, ACM (2012), 1249–1252.
26. Roudaut, A., Lecolinet, E., and Guiard, Y. MicroRolls: expanding touch-screen input vocabulary by distinguishing rolls vs. slides of the thumb. In *CHI '09*, ACM (2009), 927–936.
27. Schmidt, D., Chong, M. K., and Gellersen, H. HandsDown: hand-contour-based user identification for interactive surfaces. In *NordiCHI '10*, ACM (2010), 432–441.
28. Wang, F., Cao, X., Ren, X., and Irani, P. Detecting and leveraging finger orientation for interaction with direct-touch surfaces. In *UIST '09*, ACM (2009), 23–32.
29. Westerman, W. *Hand Tracking, Finger Identification, and Chordic Manipulation on a Multi-Touch Surface*. PhD thesis, University of Delaware, 1999.
30. Wobbrock, J. O., Morris, M. R., and Wilson, A. D. User-defined gestures for surface computing. In *CHI '09*, CHI '09, ACM (2009), 1083–1092.
31. Wu, M., Shen, C., Ryall, K., Forlines, C., and Balakrishnan, R. Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. In *TABLETOP '06*, IEEE Computer Society (2006), 185–192.