M3 Gesture Menu: Design and Experimental Analyses of Marking Menus for Touchscreen Mobile Interaction

Jingjie Zheng[1,4], Xiaojun Bi[2], Kun Li[3], Yang Li[4], Shumin Zhai[4]
[1]University of Waterloo, [2]Stony Brook University, [3]University of Colorado Boulder, [4]Google Inc.
jingjie@acm.org, xiaojun@cs.stonybrook.edu, kkli@colorado.edu, {yangli,zhai}@acm.org

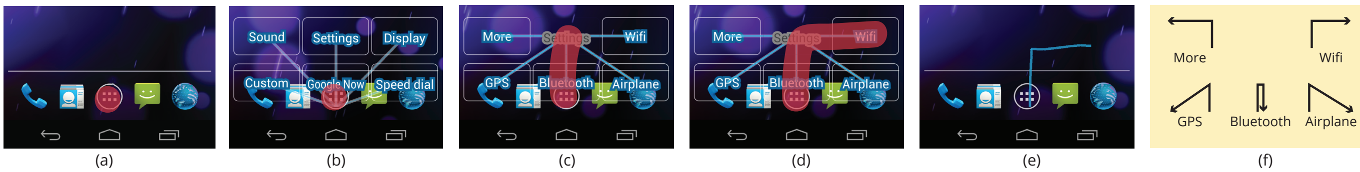

Figure 1: M3 Gesture Menu prototyped on Android with two levels of 2×3 tiles. (a) M3 is initially hidden with a designated area (e.g. the white circle) presented as its activation button. (b) Pressing and holding the button for 0.5 seconds pops up the first-level menu. (c) Sliding the finger to an item (e.g. *Settings*) reveals its submenu which replaces the the higher level content in the same space. (d) Further sliding to a terminal node (e.g. *Wifi*) activates its command. (e) Alternatively, experienced users may directly draw a gesture from the activation button approximately to *Settings*, then to *Wifi* to trigger the same command. (f) Different gestures trigger different commands, e.g. the figure illustrates all the gestures in the *Settings* submenu.

ABSTRACT
Despite their learning advantages in theory, marking menus have faced adoption challenges in practice, even on today's touchscreen-based mobile devices. We address these challenges by designing, implementing, and evaluating multiple versions of *M3 Gesture Menu* (M3), a reimagination of marking menus targeted at mobile interfaces. M3 is defined on a grid rather than in a radial space, relies on gestural shapes rather than directional marks, and has constant and stationary space use. Our first controlled experiment on expert performance showed M3 was faster and less error-prone by a factor of two than traditional marking menus. A second experiment on learning demonstrated for the first time that users could successfully transition to recall-based execution of a dozen commands after three ten-minute practice sessions with both M3 and Multi-Stroke Marking Menu. Together, M3, with its demonstrated resolution, learning, and space use benefits, contributes to the design and understanding of menu selection in the mobile-first era of end-user computing.

Author Keywords
marking menu; smartphone; novice to expert behavior transition; gesture interface

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation: User Interfaces

CHI 2018, April 21–26, 2018, Montréal, QC, Canada.
ACM ISBN 978-1-4503-5620-6/18/04.
http://dx.doi.org/10.1145/3173574.3173823

INTRODUCTION
Marking Menu [28, 32, 33] is among the most insightful contributions of interaction design research. Prior to that, designing interface features was either aimed at ease of use or at efficiency, but not both. Ease of use is typically achieved by the visual and textual affordances of graphical input – users can look, scroll, and click through graphical widgets to activate commands without memorizing the exact movement. Efficiency, on the other hand, is achieved through executing learned motor procedures like hotkeys on desktop computers. The challenge, however, is that their movement patterns are drastically different from making menu selections – users have to make a conscious and effortful switch from GUI to hotkeys [31]. Research showed such transition is very difficult to facilitate in interface design [14, 21, 37]. Even if hotkeys are effective, they are not applicable to touchscreen devices – the mainstay of today's mobile computing.

Marking Menu combines hotkeys' efficiency with GUI's ease of use. Unlike hotkeys, users apply the same movement pattern from the first time onwards. Before having memorized the movement pattern for a particular command, users visually trace the radial menu slices from one level of a pie menu (e.g. a pie menu with its up slice as *Edit*) to another (e.g. a pie menu with its right slice as *Copy*). This is called the "novice mode". Later, users may recall and draw the same directional patterns (e.g. up-then-right) directly without relying on the menu display (the "expert mode"). The movement consistency between the two modes is expected to enable a seamless transition from the former to the latter.

Marking Menu is defined on radial (or pie) menus for good reasons. Pie menus were shown to be faster than linear menus for up to 8 slices/items [10]. Further, it is believed that human memory of directions, particularly anchored ones such as those along the horizontal and vertical axes, and the directions between them (i.e. NW, NE, SW, and SE by map conventions), are easy to differentiate and remember [33].

Another important factor of the original Marking Menu design is the intentional 1/3 second delay of the radial menu display. We are not aware of strong and direct empirical studies of this feature, but abundant cognitive psychology research shows that mental elaboration and active retrieval of past behavior with depth of processing [16] is critical to memory performance. The cost of waiting for the pop-up incentivizes active memory retrieval. Research shows that an appropriate amount of cost indeed facilitates user interface learning, but too much cost could be discouraging and detrimental [17, 15].

Marking Menu has attracted a large amount of research over the last twenty years [4, 5, 18, 25, 39, 54, 55], but they still have not entered the mainstream practice of user interface design. This is particularly surprising given the fact that in the last ten years, user computing had another revolution since the PC. Touchscreen mobile devices have shaken and shaped the entire computing industry. Given their cognitive and motor control advantages, we would expect marking menus to be on major mobile operating systems or applications used by billions of users every day. With exceptions (e.g. [30]), users hardly ever see a marking menu in their everyday experience. While it is hard to ascertain all the reasons that stand in the way of user interface innovation and user behavior change, we summarize the following factors that may impede the adoption of marking menus, particularly on touchscreen mobile devices:

Inefficient space use. The original Marking Menu is not mobile-friendly [28] – each level of a pie menu is centered on an edge node of the previous level, so the interaction space may move beyond the small screen of a mobile device. On the other hand, Multi-Stroke Marking Menu [55] and its variants [4, 5, 18, 25, 54] do not move in space – they detect a series of simple straight marks for sub-selections in a limited space rather than a whole unistroke gesture. The problem is that in practice, they require a dedicated, occlusive space to disambiguate the interaction with the menu from other visible graphical elements occupying the same space. In contrast, unistroke gestures used by the original Marking Menu and M3 avoid this ambiguity because the end of a gesture automatically concludes the user's engagement with the menu.

Limited menu resolution. The original Marking Menu suffers from poor accuracy when the menu item is deep in the menu hierarchy, so the number of practically usable commands, or menu resolution, is limited. Kurtenbach et al. [33] showed that, for example, to keep the error rate under 10%, a menu with eight targets on a level can not go beyond the level of two. Even though Multi-Stroke Marking Menu was proposed to relieve this issue, again, applying it in practice is hampered by the requirement of a dedicated, occlusive space we have discussed above.

Unknown learning curves. The learning effort associated with marking menus' novice to expert transition curves is under-investigated, making it difficult for product designers to make informed decisions. Most research focused on the resolution, or expert performance of marking menus. User learning by nature is hard to study and quantify, so the field has not offered enough empirical research to support the conceptual and theoretical advantage of marking menu learning.

Under-researched subjective experience. Subjective experience is as important a criterion as performance metrics. Although some research (e.g. [5, 18]) touched on the preference and feedback of marking menus, our understanding towards this critical issue is clearly not enough.

In this paper, we begin addressing the four problems by proposing, implementing, and studying a new marking menu variant we call *M3 Gesture Menu* (M3). A controlled experiment on expert performance showed M3 was faster and less error-prone compared to those reported in the literature. A second experiment on learning demonstrated for the first time at a menu capacity beyond a few commands, users successfully transitioned to recall-based execution for a dozen gestures after 30 minutes of practice over three days in M3 and Multi-Stroke Marking Menu.

DESIGNING M3 GESTURE MENU

The essential design of M3 Gesture Menu is quite simple. The clearest way of describing it is through the example illustrated in Figure 1, although the specifics in practice can vary. In M3, a rectangle menu, a grid of tiles each filled with a menu item label, pops up upon the user pressing and holding on a predesignated location we refer to as *activation button*. Upon selecting an item, its submenu will replace the same space the current menu is occupied. M3 provides three types of interaction for users with different levels of proficiency:[1]

A beginning user can tap one of the menu items which, if not a terminal node, expands to and replaces the top-level menu with the lower-level menus spawn from the menu node being tapped. This process can be nested till a terminal node at the N^{th} level is selected. Optionally, upon tapping a terminal node an animation is played showing the trajectory starting from the M3 activation button to the intermediate tile centers ending at the terminal tile center, therefore revealing the next two types of interaction. **An intermediate user** can visually trace the finger from the activation button to the category on the top level of M3, to lower levels, and all the way to the terminal tile. This trajectory once again can be animated to reinforce the gesture shape associated with this command. **An experienced user** can directly gesture the memorized shapes without relying on the menu to display.

We should note that the inclusion of tapping for beginning users is optional. This option may ease the initial adoption of M3, for it is just an ordinary menu. It may not necessarily facilitate the learning of gestures per se, but may help users become familiar with the menu layout. It is possible some users find the tapping mode good enough and never make the transition to gesturing ("satisficing" [44, 42]). At this point, we view tapping also as a design choice. If "forcing" users to gesture is a paramount objective, the option can be disabled.

M3 recognizes gestures using the proportional shape matching algorithm adapted from SHARK[2] [26]. The algorithm

[1] See the accompanying video for a demonstration.

searches for the minimal proportional shape matching distance from the user input gesture to a set of canonical gestures that correspond to the menu commands.

As we have discussed earlier, marking menus in general are expected to support seamless novice to expert behavior transition, facilitate memorization with directional marks, and incentivize active memory retrieval with a cost added to revealing the novice mode menu interface. In addition to these theoretical advantages, we also envision M3 to provide the following benefits to fit better to mobile form factors:

Expressive gesture shapes. Shapes as defined by tile locations from one level to another define, encode, and decode the final commands. Shapes are made of both angular and relative distance features, not angles alone. This enables more expressive gestural marks for human memory to take advantage of. Using unistroke gestures also reduces the need for a dedicated, occlusive interaction space as in Multi-Stroke Marking Menus. Matching shapes as a whole instead of aggregating stroke direction recognition results may also bring better recognition accuracy.

Efficient space use. Applying a rectangle layout allows for more efficient real estate use. All used space is fully packed and time multiplexed between different layers of menus.

Stable interaction space. Since the lower-level menus occupy the same space as the upper level, the menu does not move in space, offering better spatial stability.

IMPLEMENTATION

The prototype in Figure 1 was built on a version of the Android System UI codebase. Our initial and informal daily "dogfood" use showed that a system implementation of M3 was feasible and its experience was positive, although such tests were inherently anecdotal and subjective.

For more controlled and systematic studies to be presented next, we implemented a 3×3 tile M3 Gesture Menu as an Android application, shown in Figure 4a, on a Google Nexus 5X smartphone running Android 6.0. Each tile of M3 was a square with a side length of 19 mm. The device has a screen size of 147×72.6 mm and a resolution of $1,920 \times 1,080$ pixels. The menu was aligned slightly to the right for right-handed participants and vice versa.

EXPERIMENT 1: M3 GESTURE MENU RESOLUTION

The focus of this experiment was to test M3 Gesture Menu's resolution, or performance limit. Previous research showed the original Marking Menu tends to suffer from low resolution particularly at depth level three and beyond [28]. We define resolution as the time (T) and error (E) of differentiating one command from all other commands when the user reaches expert performance in a menu system. A hierarchical menu's resolution depends on its depth (D, number of levels) and capacity (C, total number of permissible commands) – comparison of resolution is only meaningful when the depth and capacity of two menus are the same. Together, we denote resolution as $\{T,E\}@\{D,C\}$. For example, a menu with two levels and 64 commands having resolution of 2.1 seconds in time and 15% in error can be expressed as $\{T: 2.1s, E: 15\%\}@\{\text{D: } 2, \text{C: } 64\}$. Of course, resolution also depends on the form factor and input device used for testing. This experiment tested M3's resolution on a small screen smartphone (rather than a tablet) with a finger (rather than a stylus), which were more likely to be biased against M3's resolution results [28, 55, 45].

Participants and Apparatus

We recruited 12 right-handed participants (3 female, age ranging from 18 to 40 with the majority between 24 to 30). All reported extensive experience with touchscreen smartphones, self-reporting average weekly usage ranged from 7 to 100 hours (M=26.9, SD=26.4). Participants were seated with one or two hands holding the phone depending on their preference.

Task and Stimuli

We measured M3's resolution at an expert performance level by suggesting the exact actions participants need to take, as is typically done in marking menu research [28, 55]. The menu tiles were labeled with digits from 1 to 9 like a dial pad on the top level (Figure 2c). The submenu tiles were labeled with its parent item as prefix with the digits again from 1 to 9 (e.g. Figure 2d for the submenu of tile 3).

The experiment enforced seven consecutive repetitions of each gesture to quickly obtain expert performance metrics – seven since it was shown to be sufficient to learn and reach the peak performance of a gesture [51, 8]. The first three repetitions were *guided execution* (Figure 2), where participants visually traced the suggested gesture on the menu display expanded immediately after touching the activation button. The next four repetitions were *recalled execution*, where the menu display was disabled and participants were required to execute the same gesture through forced recall. All repetitions started with a button press (Figure 2a), after which an instruction to the current gesture would appear on top of the screen (Figure 2b). Our focus was to measure the eventual resolution of M3. The experimental manipulation, therefore, was to reach the expert performance as quickly as possible. The menu behavior of guided and recalled execution does not map directly to the novice and expert modes of M3.

When a wrong gesture was executed, participants were asked to immediately repeat the trial. For each trial, we measured reaction time, execution time, and total time. Reaction time was measured from pressing the start button to pressing the activation button of M3. Execution time was measured from pressing the activation button to eventually triggering a command. Total time was the sum of reaction and execution time.

Design

Our experiment design was mixed-factorial and repeated measures, with the between-subject factor being FINGER. Half of the participants were asked to consistently use their right-hand thumb and the other half, the index finger, to perform the gestures [3, 20]. Prior to the formal experiment, six warm-up gestures were demonstrated and participants practiced the same gestures. We tested M3 to up to three levels of eight items, since this setting is a high-capacity yet usable design [24, 32]. We refer to the gestures activating commands

Measurement	Depth 1	Depth 2	Depth 3	Depth 1, 2, 3
Reaction Time (ms)	303.92 ± 84.00	265.38 ± 108.50	266.88 ± 110.58	268.81 ± 107.88
Execution Time (ms)	256.27 ± 96.19	479.90 ± 174.25	723.67 ± 256.91	537.71 ± 241.78
Total Time (ms)	564.08 ± 161.74	746.96 ± 248.45	998.54 ± 333.38	810.31 ± 303.55
Error Rate	0.35% ± 1.20%	2.91% ± 2.33%	8.85% ± 4.47%	4.54% ± 2.69%

Table 1: Mean and SD of reaction time, execution time, total time, and error rate of executing gestures in depth 1, 2, 3 of M3.

Technique	Literature	Stroke Type	Input Device	Total Time (s)		Error Rate	
				Depth 2	Depth 3	Depth 2	Depth 3
M3	This Paper	Compound-Stroke	Finger	0.7	1.0	3%	9%
Original Marking Menu	Kurtenbach and Buxton [28]	Compound-Stroke	Stylus	1.4	2.3	7%	17%
Original Marking Menu	Zhao and Balakrishnan [55]	Compound-Stroke	Stylus	2.3	3.6	10%	17%
Multi-Stroke Marking Menu	Zhao and Balakrishnan [55]	Multi-Stroke	Stylus	2.3	3.4	4%	7%

Table 2: M3's resolution compared to other marking menu variants with a breadth of eight reported in the literature. If other factors such as device size were tested, we report the one with the best performance.

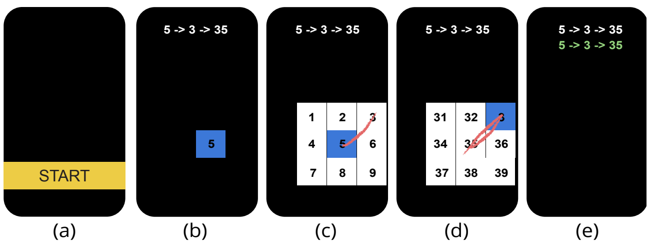

Figure 2: Guided execution: (a) a start button appears at the bottom of the screen; (b) after pressing the button, a gesture instruction, $5 \rightarrow 3 \rightarrow 35$, is displayed, and the activation button of the menu 5 is shown; (c) pressing the activation button 5 expands the menu, then sliding the finger to 3 displays its submenu; (d) further sliding to 35, then lifting the finger activates the item; (e) feedback is provided after the trial.

on the N^{th} level as *depth-N* gestures. The experiment breaks into three parts: Part 1 tested the complete set of eight depth-1 gestures; Part 2 tested the complete set of 64 depth-2 gestures; and Part 3 tested 32 depth-3 gestures randomly sampled from the total 512 gestures. The same gesture set was used across all the participants with the order randomized. Participants were asked to take a break between two parts.

In summary: 12 PARTICIPANTS × 1 of 2 FINGERS × 104 GESTURES × 7 REPETITIONS = 8,736 data points in total.

Results

Mixed-factorial analysis of variance and pairwise Tukey t-tests with Bonferroni correction were used for all measures. Trials were aggregated by participant and the factors being analyzed. Time data were aggregated using median.

Repetition and Practice Effect

Strong repetition and practice effects on total time were found for guided execution ($F_{1.17,12.9} = 60.78$, $p < .0001$, $\eta^2 = .682$) and recalled execution ($F_{1.81,13.00} = 52.47$, $p < .0001$, $\eta^2 = .235$), but no effect on error was found. Figure 3 shows that the time of completion at all three levels of depth could quicken to a stable level after a few repetitions. The last three trials of recall production appear to have reached a performance limit, so we use them to estimate menu resolution.

Time and Error

No effect of FINGER on time or error was found, so we combined the data for analysis. As shown in Table 1, the resolution for depth-1 and depth-2 gestures were $\{T: 0.6s, E: 0.35\%\}@\{\text{D: 1, C: 8}\}$ and $\{T: 0.7s, E: 2.91\%\}@\{\text{D: 2, C: 64}\}$, respectively, suggesting users could accurately articulate as many as 64 commands very quickly. This is a fairly good menu capacity for smartphones, sufficient for accommodating common features such as launching frequent applications, toggling setting options, or providing context menus. With depth-3 gestures, the menu resolution was $\{T: 1.0s, E: 8.85\%\}@\{\text{D: 3, C: 512}\}$, providing a much larger number of permissible commands, but with only slightly longer time and lower accuracy.

To better interpret M3's resolution, we compared against the original Marking Menu and Multi-Stroke Marking Menu, both tested with a tablet and a stylus. We should be cautious of the fact that the previous studies were conducted in different form factors and with different input devices, so these comparisons should be interpreted accordingly. The mobile display size and finger rather than stylus operation in this study were more likely to be biased against the M3's results [28, 55, 45].

As shown in Table 2, at the same menu depth (2 or 3) and capacity (64 or 512 commands), M3 produced less or much less than half the error in less or much less than half of the total completion time of the original Marking Menu according to the data reported by two separate studies. In comparison to the revised Multi-Stroke Marking Menu, M3 took less than one third of time at somewhat lower (at Depth 2) or somewhat higher (at Depth 3) error rate. Taking Depth 3 as an example, the original Marking Menu had resolution of $\{T: 2.3s, E: 17\%\}@\{\text{D: 3, C: 512}\}$ by Kurtenbach and Buxton [28] and $\{T: 3.6s, E: 17\%\}@\{\text{D: 3, C: 512}\}$ by Zhao and Balakrishnan [55]. Multi-Stroke Marking Menu had resolution of $\{T: 3.4s, E: 7\%\}@\{\text{D: 3, C: 512}\}$ according to Zhao and Balakrishnan. At the same depth, the resolution of M3 was $\{T: 1.0s, E: 8.85\%\}@\{\text{D: 3, C: 512}\}$,

The difference in accuracy is likely due to the stroke type and the way a stroke is recognized. The original Marking Menu uses compound strokes (unistrokes) and recognizes each level of the menu only by the stroke orientation. When the user makes an error on one level, that error will inevitably propagate to the subsequent levels. The deeper the commands are, the more accumulative the error may become. The Multi-

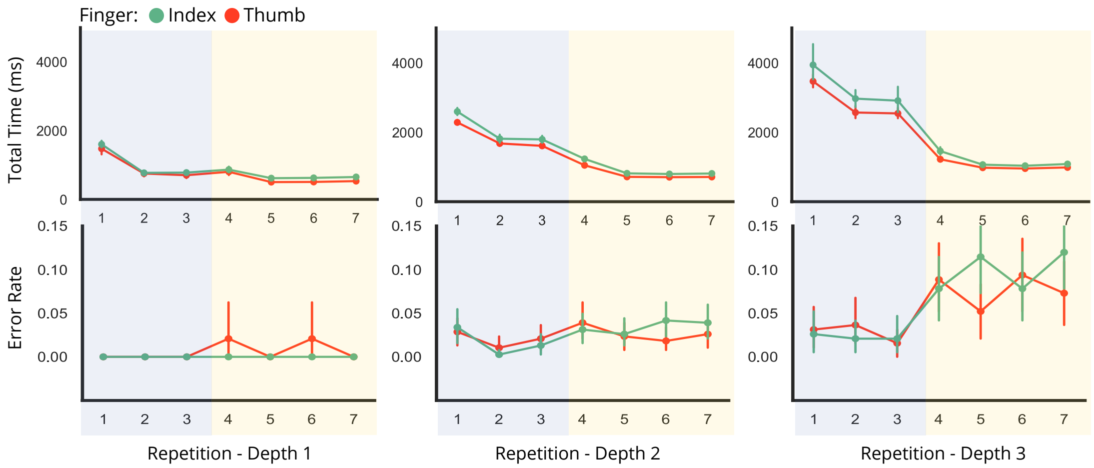

Figure 3: Total time and error rates for gestures in different depths: blue shade indicates guided execution; yellow shade indicates recalled execution.

Stroke Marking Menu, on the other hand, separates a marking gesture into several simple ones, resulting in a higher accuracy [11]. Although also using compound strokes, M3 attained much higher resolution than the original Marking Menu and comparable accuracy (but at faster speed) as the Multi-Stroke Marking Menus, probably because M3 relies on the entire shape of the gesture in recognition and uses a stationary layout.

In sum, M3 departs from other marking menus in significant ways, most notably its stationary layout and the use of total shapes. The findings of experiments suggest the eventual speed and accuracy of M3 at all three depth levels of resolution are comparable to or better than those of the conventional marking menus reported in the literature.

EXPERIMENT 2: NOVICE TO EXPERT TRANSITION

This experiment switches focus to another important aspect of M3 Gesture Menu, and marking menus in general – learning and specifically the transition from novice to expert behavior through practice. Although the expectation of novice to expert progression is a critical marking menu design rationale, its empirical support is very limited in the literature. The research cited earlier were all focused on expert performance, as we did in Experiment 1. Other research tended to focus on user interface features that provide better navigation and exploration to encourage learning [4, 5, 18].

The only research we are aware of that investigated the transition effect of marking menus was by Kurtenbach et al. [29]. In a longitudinal case study tested with an audio editing program, they showed that although users switched back and forth from Marking Menus to linear menus, they eventually graduated to using Marking Menu that enabled better performance. However, their evaluation was only limited to one level of six commands.

There are worth noting important differences between M3 and traditional marking menus, which may affect their respective learning. Traditional marking menus use consecutive movement directions as the main memory cue, whereas M3 uses shapes determined by both distances and directions of the consecutive gesture segments. Although shape writing / SHARK experiments [51, 26] did show users' ability to memorize shapes (or "sokgraphs"), they were defined on a static single layer layout rather than a multi-layered dynamic layout as in M3, so it is not a sufficient indication of M3's recall performance.

To further strengthen the theoretical and empirical foundation of marking menu design in general, and M3 in particular, we investigated how fast this transition may take place on a two-level M3 along with Multi-Stroke Marking Menu that also fits the mobile form factor. We compared their performance metrics against a linear menu baseline. The transition from visually-guided novice to recall-based expert performance has not been previously studied beyond one level of a few commands to our knowledge. In the end, we also estimated the cost and benefit of learning M3 to help designers make informed decisions.

Participants and Apparatus

We recruited 15 right-handed participants (5 female, age ranging from 18 to 60 with the majority between 24 to 30). All reported extensive experience with touchscreen smartphones, self-reporting average weekly use ranged from 5 to 98 hours (M=21.9, SD=23.0). Participants were all familiar with linear menus, but none had prior experience with any marking menu, nor were they aware of which technique was proposed by us. Apparatus were the same as Experiment 1.

Task and Stimuli

After pressing a start button, participants were prompted with a command stimulus to be activated in a two-level menu. The command was presented as its full menu path like "*furniture* → *chair*". To minimize the impact of personal experience, we followed similar experiments [2, 21], using everyday objects organized into eight categories, each containing eight objects, As shown in Figure 4, participants were tested on three menu conditions: M3 Gesture Menu (M3), Multi-Stroke Marking Menu (MM), and Linear Menu (LM):

M3 / MM initially only displayed an activation button. The novice mode interface could be revealed by holding the button for 0.5 seconds. Participants could then either tap on an item or slide to an item at their choice. The marking menu interface could also be dismissed by tapping outside the menu area. If participants activated a command *correctly* but using the novice mode, an animation of its canonical gesture would be played to prompt them to gesture directly. The expert mode could be triggered by directly gesturing out of the activation button. The color of the activation button would change slightly to indicate the menu was in a different mode.

LM always displayed the items on a level, without the need to reveal the interface or scrolling. LM also provided a back button to navigate to the previous level.

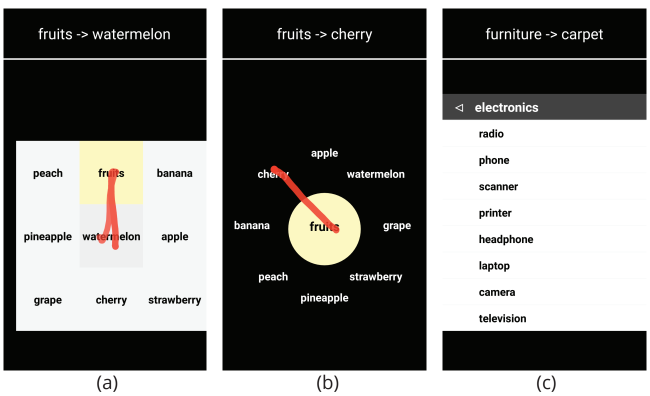

Figure 4: Menu techniques: (a) M3 Gesture Menu, (b) Multi-Stroke Marking Menu, and (c) Linear Menu. Screenshots were taken when selecting terminal targets on the second-level menus with visual feedback.

Practice Trial and Successful Memory Recall

In the experiment, stimuli were presented in the unit of practice trials. A *practice trial* consists of one or more attempts at activating the prompted command correctly, which we call *sub-trials*. For M3 or MM, passing a practice trial required participants to activate that command using the correct gesture without revealing the menu display. In other words, participants could execute any number of times in the novice mode until they felt confident to try the expert mode, but the practice trial would only end when the gesture was executed correctly in the expert mode. Otherwise, the experiment would immediately repeat the same sub-trial. If a practice trial takes exactly one sub-trial to get to executing the command in the expert mode correctly, we call it a *successful memory recall*.

Design

Our experiment consisted of three sessions. Each session was conducted at least one day after the previous one. A session had three blocks. In each block, we tested one menu technique for 10 minutes. The order of menus was counterbalanced. Menu content was randomized for each participant.

Expanding Rehearsal Interval

For greater practice efficiency, we used expanding rehearsal interval (ERI) to schedule the presence of stimuli [36, 51]. Each participant was assigned with one ERI scheduler for each menu technique. Subsequent sessions resumed previous sessions' scheduler states. The scheduler maintained two lists: a *candidate list* containing the stimuli to be learned and a *rehearsal list* containing the stimuli being actively rehearsed at various rehearsal intervals. The *rehearsal interval* associated with a stimulus defines the time to its next scheduled practice trial after the current practice trial was completed. The scheduling algorithm worked as follows:

1. Initially, the candidate list contained all the stimuli and the rehearsal list was empty. The rehearsal interval for all the stimuli was set to zero, meaning that once a stimulus was put into the rehearsal list, it would need to be rehearsed immediately.

2. If a stimulus in the rehearsal list had reached the scheduled rehearsal time, poll that stimulus for a practice trial. Otherwise, randomly poll a new stimulus from the candidate list. If the candidate list was empty, terminate the session.

3. If the practice trial for this stimulus was a successful memory recall for M3 or MM (or consisted of a single successful sub-trial for LM), increase the rehearsal interval. If it was originally zero, set to 30 seconds. Otherwise, double the rehearsal interval.

4. If the practice trial for this stimulus was not a successful memory recall for M3 or MM (or had an error for LM), the rehearsal interval stayed unchanged.

5. After the practice trial was completed, put the stimulus back to the rehearsal list.

The dynamic nature of ERI and the limit in time also means there would be varied numbers of stimuli presented per participant and menu. To summarise the design: 15 PARTICIPANTS $\times$ 3 MENUS $\times$ varied number of STIMULI.

Results

Repeated measures analysis of variance and pairwise Tukey t-tests with Bonferroni correction were used for all measures. Trials were aggregated by the participant and factors being analyzed. Time data were aggregated using median. Subjective scores were first transformed using the Aligned Rank Transform [50]. Due to the time limit in the experiment, different stimuli had different numbers of practice trials. To align the data and preserve enough number of stimuli for each menu technique, we chose the set of stimuli that had at least seven practice trials. This gave us 878 stimuli in total, with more than 250 stimuli for each menu technique. These practice trials came from different sessions for different stimuli.

Time

As shown in Figure 5, we separated all the sub-trials into novice and expert modes and analyzed their total time (measured the same as Experiment 1) for M3 and MM, respectively. We then compared them to the total time of LM, all as a function of the number of practice trials.

In the novice mode, we found a main effect of MENU on total time ($F_{2.15,30.09} = 533.53$, $p < .0001$, $\eta^2 = .876$). LM (M=2834.63ms, SD=383.16ms) was faster than MM (M=5509.03ms, SD=656.44ms), which was faster than M3 (M=6044.00ms, SD=564.35ms, all $p < .005$). In the expert mode, we also found a main effect of MENU on total time ($F_{2.15,30.09} = 533.53$, $p < .0001$, $\eta^2 = .876$). M3 (M=1107.33ms, SD=274.82ms) and MM (M=1270.20ms, SD=495.81ms) were both significantly faster than LM (M=2834.63ms, SD=383.16ms, both $p < .0001$).

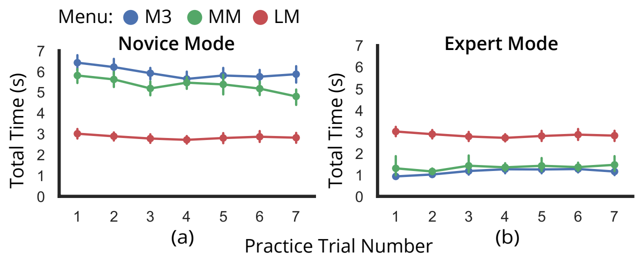

Figure 5: Time of each practice trial by menu. The novice mode of marking menus was slower than LM, but the expert mode was significantly faster.

This suggests that for beginners, marking menus (M3, MM) take longer time than linear menus to trigger a command in their novice mode (guided by the menu display). Once the users have transitioned to the expert mode, executing them directly through memory recall will lead to a large performance gain in speed.

Recall Rate

To measure how fast the participants made the transition from relying on the menu display to drawing gestures directly through memory recall, we analyzed successful recall rates for M3 and MM as the gestures received more practice (Figure 6a). The increase of rehearsal interval could also reflect the number of gestures being memorized, although less direct (Figure 6b). The recall rate was measured by the percentage of stimuli that were successfully executed in the expert mode without popping up the menu display in M3 or MM.

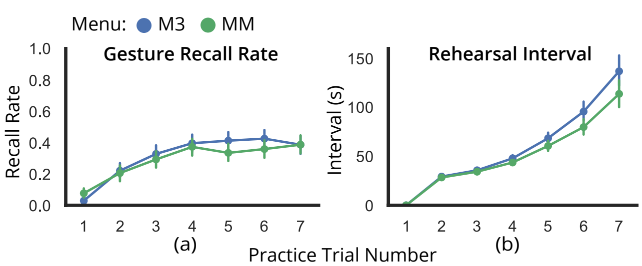

Figure 6: Increasing gesture recall rate and rehearsal interval indicate more gestures were being memorized with more practice.

As shown in Figure 6a, the recall rates for both menus quickly increased. The main effect of PRACTICE on recall rate ($F_{3.81,53.31} = 35.83$, $p < .0001$, $\eta^2 = .288$) was significant.

Note that although the relative recall rate did not further increase after fourth practice trial, the absolute number of gestures remembered increased as more commands were moved from the candidate list to the rehearsal list since the interval of some of the successfully recalled commands increased (Figure 6b). We will return to the absolute number of gestures remembered in the follow-up experiment.

We should also note that the learning performance scheduled with ERI was optimized for memorizing a maximum number of gestures within a limited time. For example, participants practiced on average 20 different gestures for M3 in the first session within 10 minutes. This could be different from the user behavior in practice. In practice, users might not likely attempt to memorize this many gestures in such a short time.

Perceived Workload

To measure the perceived task load across all three techniques, we used NASA-TLX and found main effects of all subscales (Figure 7): mental demand ($F_{1.99,27.79} = 53.63$, $p < .0001$, $\eta^2 = .580$), physical demand ($F_{1.95,27.24} = 7.38$, $p < .005$, $\eta^2 = .170$), temporal demand ($F_{1.85,25.85} = 8.75$, $p < .005$, $\eta^2 = .093$), effort ($F_{2.12,29.67} = 29.59$, $p < .0001$, $\eta^2 = .421$), performance ($F_{2.11,29.58} = 36.44$, $p < .0001$, $\eta^2 = .455$), and frustration ($F_{1.71,23.96} = 17.68$, $p < .0001$, $\eta^2 = .348$). As expected, LM had the lowest task load for all subscales (all $p < .005$), but M3 had significantly lower task load than MM on physical demand and frustration. This signifies a better learning experience for M3 than MM.

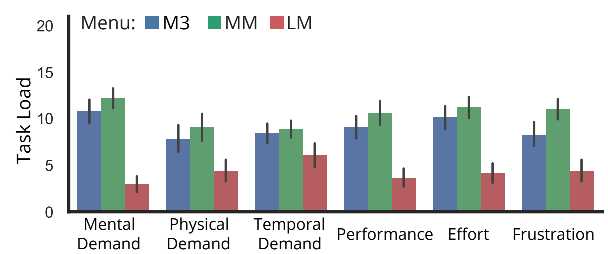

Figure 7: NASA-TLX for three menu techniques.

Subjective Responses

In terms of overall preference, nine participants preferred M3, five participants preferred LM, and one equally liked M3 and MM. A semi-structured interview at the end of the experiment helped us identify what had driven the participants to prefer one menu technique over the other at a more nuanced level.

Movement. Comments from the participants suggested drawing a stroke as a complete shape resulted in smoother finger movement (10 participants), less physical effort (9 participants), and better memorability (11 participants). Participants described the interaction with M3 to be “continuous”, “fluid”, and “smooth”, which facilitates the development of muscle memory (7 participants). Interestingly, two participants compared the open-loop gesture production in M3 to word-gesture keyboards [53], as one of them said: “this is like the gesture keyboard, once you trust it, it just works”. In contrast, separating one continuous gesture into two strokes as in MM could break this “flow”. Four participants attributed this to the difficulty of retrieving the relative order of the strokes. Eight participants considered drawing two strokes requires more physical movement and effort. Nevertheless, two participants suggested drawing strokes on MM was easier because they could take a break in between.

Memorization. Eleven participants found gestures on M3 to be easier to remember, one found MM to be easier, and the other three indicated memorizing M3 and MM gestures requires a similar effort. Few comment explained how participants memorized M3 gestures, but many talked about the strategies of memorizing MM gestures. Two participants related the MM strokes to the hour and minute hands of a clock. Three participants memorized the strokes as static angles or shapes, but acknowledged that this made it difficult to retrieve the stroke order. Ten participants found stroke orientations harder to memorize than shapes. Two participants suggested MM be only used for one level because memorizing items on the second level was too difficult.

Layout. A few participants suggested defining the menu on a grid helps provide better user experience by making efficient space use (1 participant), enabling compact item placement (1 participant), and offering better perceived control ability than the radial arrangement (1 participant).

Follow-up Experiment on Memory Recall

We conducted a follow-up experiment to explicitly measure the number of gestures memorized after each day's practice. The experiment used the same format as the main experiment, but with four sessions and an explicit recall test at the beginning of the last three sessions, similar to the sokgraph experiment in shape writing [51]. Ten participants were recruited (two left-handed, three female, age ranging from 18 to 30 with the majority between 24 to 30). The recall tests evaluated all the stimuli in the current rehearsal list with their order randomized. For each stimulus, participants could make up to two attempts to execute the correct gesture. Only the expert mode was enabled during the test.

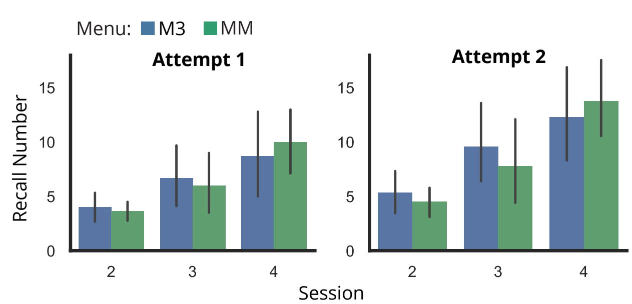

Figure 8: Number of recalled gestures after the first and second attempts in the recall tests.

Memory Recall

The results are shown in Figure 8. For both menus, the numbers of recalled gestures were increasing after each session. There were main effects of SESSION on recall number after the first attempt ($F_{1.42,12.74} = 12.89$, $p < 0.0005$, $\eta^2 = 0.21$) and the second attempt ($F_{1.45,13.02} = 21.78$, $p < 0.0001$, $\eta^2 = 0.244$). Post hoc tests showed the last test had significantly more recalled gestures than the first one for both attempts ($p < .005$).

After three sessions of practice (30 minutes in total, 10 minutes in each session), the two menus attained a similar level of recall numbers ($p > .05$). With M3 and MM, participants recalled 12.30 ± 7.56 and 13.78 ± 6.16 gestures after the second attempt, respectively. In addition, the newly recalled gestures in each test (that also retained in the next test) on average received 6.5 and 6.3 times of practice in the training for M3 and MM, respectively. This provides us with an estimation of the times of practice needed before transitioning to recall-based expert behavior.

Discussion

The two multi-session lab studies in this section showed that through practice users could indeed transition to the recall-based expert mode for an increasing number of commands with both M3 and MM. The follow-up experiment revealed that after 30 minutes of practice over three days, participants on average could use the expert mode for over a dozen commands in M3 and MM (Figure 8).

The results also showed that once transitioned to the expert mode through practice, the total time of M3 and MM were less than half of the total time of a comparable linear menu (Figure 5b), clearly demonstrating the invaluable long-term benefits of M3 and MM. In comparison to the speed limits shown in Experiment 1, the expert speed measured here was still just at the beginning of the expert mode, hence lower than the limit speed measured in Experiment 1. The results also showed the costs of learning M3 or MM. Before transitioning to the recall-driven expert mode, the time of visually tracing M3 or MM was about twice of the time of using the familiar linear menu (Figure 5a). The workload of learning M3 or MM was clearly higher than using a linear menu by all NASA-TLX measures (Figure 7).

Between the more traditional multi directional-stroke Marking Menu (MM) and the novel shape-based M3, more participants preferred M3 to LM (and MM), despite (and after) the burden of practices during the learning sessions totaling 30 minutes for each menu system. It is important to note one-third of the participants still preferred the linear menu after the practice sessions.

As lab experiments, the measurements made in the study could differ from the everyday experience in the practical use of M3 or MM in many ways. For example, due to the ERI schedule, the learning efficiency in these experiments could be greater than in the wild on the basis of each command (Figure 5), but less efficient as a measure of the total number of gestures memorized (Figure 8) since many more commands in the rehearsal list were only practiced a small number of times. However, we believe the experiments still captured the essence and relative trends and patterns between techniques.

The biases or incentives of the novice to expert transition were probably stronger if it were in practical use when users are intrinsically motivated to improve their performance driven by need. In our experiment where the task was artificial, the same setups provided by M3 to incentivize the transition, including the delay for showing the visual guidance and the animation reminding users of the expert mode gestures, were probably not enough. The experiment manipulation that blocked users until they performed the expert technique was to further increase the cost of relying on visual display and making an error.

Based on the data from the experiments we may draw a first-order "back-of-the-envelop" estimate of the learning cost versus time-saving benefit once learned. Such estimates may help guide future designers and researchers to judge when and where methods like M3 Gesture Menu should replace traditional methods such as linear menus.

Let X and Y be the average of completion time in the novice mode or learning stage of M3 and a linear menu baseline, respectively. Let U and V be the average of completion time in the expert mode or proficient stage of M3 and a linear menu baseline, respectively. Let N be the average number of trials to reach the expert mode, and M be the average number of trials in the expert mode beyond which the time savings in the expert mode begin to outweigh the time invested in learning M3. In other words, the cost-benefit equilibrium point is at

$$(X-Y)\cdot N \leq (V-U)\cdot M \tag{1}$$

From Experiment 2, we estimate, the "waste" of using M3 in novice mode instead of a linear menu is

$$(X-Y) = 6044 - 2834 = 3210 \text{ ms} \tag{2}$$

Also from Experiment 2, we estimate the saving of using M3 in expert mode instead of a linear menu, quite conservatively since the expert time in Experiment 2 was just the beginning and well above what was measured in Experiment 1, at

$$(V-U) = 2834 - 1107 = 1727 \text{ ms} \tag{3}$$

From Equation (1)

$$3210\cdot N \leq 1727\cdot M \tag{4}$$

$$M \geq 1.85\cdot N \tag{5}$$

From Experiment 2, the average number of trials to reach recall was 6.5. Putting it at N = 10 to be conservative, we have

$$M \geq 18.5 \tag{6}$$

In other words, quite conservatively, a command used

$$N+M \approx 29 \tag{7}$$

times will begin to pay off in time savings. After that, each new use of M3 in the expert mode would be another 1727 ms saving.

OTHER RELATED WORK

Our work can be placed in a broader research area that uses gestures to invoke commands on touchscreen mobile devices (see Zhai et al. [53] for a review of gesture interfaces and [2, 9, 38, 45] for some well-known work) – discussing this broader area is beyond our scope. In this section, we focus on the most relevant research, including marking menu variants proposed to offer better menu resolution, techniques that facilitate gesture learning, techniques in a wider context that also aim to support smooth transition to the expert method, FastTap that is an expert menu technique also designed for mobile form factors, and word-gesture keyboards [52] that inspired many of M3's design considerations.

Marking Menu Variants for Better Resolution

Early marking menu variants mostly sought to improve menu resolution. The original Marking Menu [28, 32, 33] was shown to have a very limited resolution, especially after reaching the depth of three (eight targets on each level). Multi-Stroke Marking Menu [55] separates the single gesture into several simple straight marks, which effectively increased the depth limit to exceed three.

Also using multiple strokes, a number of research projects further increased the breadth limit, i.e., the number of targets on the same level. Zone Menu and Polygon Menu [54] both increased the breadth to be over eight by considering not only the stroke orientation but also the starting position of that stroke about a user-defined origin. Flower Menu [5] supports seven different curved gestures towards each orientation, expanding the breadth to be even more than twenty items.

These menu designs are effective in accommodating many items on the same level than Multi-Stroke Marking Menu, but they also require larger interaction space to resolve the larger number of commands. When designing for mobile devices, however, improving menu resolution by increasing space use becomes impractical. Our work showed that M3 not only provides the space efficiency and menu resolution needed for mobile, but also avoids the interaction ambiguity caused from using multiple strokes.

Gesture Learning

Very little research in the literature directly examined the novice to expert transition behavior of marking menus, even though this expected transition is critical to marking menus' main premise. Past work has mostly focused on designing user interface features to improve the learning experience. For example, Bailly et al. [4] provided previsualization for novices to better explore the menu; Bailly et al. [5] proposed to use curved gestures instead of straight marks to improve memorability.

Related to gesture learning in general is a line of research that provides animated gesture trails as guidance. Kurtenbach et al. [31] presented a crib sheet where users can play animated demonstrations of gestures. Bau and Mackay [6] proposed OctoPocus which displays the most likely gestures as the user is completing gesture input. It has been shown that an appropriate amount of guidance promotes learning, whereas excessive guidance hinders it [41, 1]. In our work, we animated the gesture to reveal and reinforce the gesturing option. We believe this should provide sufficient guidance in a lab setting without becoming a hindrance, but certainly the best form and amount of guidance need to be further studied.

Another way to aid gesture learning is to tap into spatial memory [23, 35, 47, 48, 49, 19, 39, 7]. Spatial constancy allows users to rapidly locate target items when well practiced [13, 12, 43]. To take its advantage, we kept M3's menu item locations static, not fast adapting (e.g., to recency). More generally on spatial information processing, M3, unlike other marking menus, recognizes not only angular feature, but also relative distances in gestures. We believe the richer total shape spatial features could be a benefit to human memory.

Novice to Expert Behavior Transition

In a wider context than marking menus, supporting novice to expert behavior transition has been more extensively explored and studied (see [14, 34, 56] for reviews). Many interaction techniques have been proposed. For example, word-gesture keyboards [51, 26, 52] enables better typing efficiency by connecting individual key taps into a single stroke gesture; ExposeHK [40] prompts hotkey information when the modifier key is being held; FastTap [23, 35, 22, 34] displays a grid of buttons upon a finger pausing on a corner activation button; and similarly, HandMark Menus [49, 46] reveal the menu interface when placing a hand onto the touchscreen after a delay.

This class of interaction techniques, sometimes referred to as "rehearsal-based interfaces" [22, 56, 34], is similar to marking menus in four aspects regardless of their interaction modalities. First, they all follow the Kurtenbach et al.'s rehearsal hypothesis [31], i.e., the novice method is the physical rehearsal of the expert technique. Second, a cost or delay is imposed on the novice method to nudge users towards learning or memorizing the expert commands. Third, the novice method serves the purpose of guiding the expert, often recall-based, method with complete visual information. Fourth, they all provide spatial constancy to leverage the power of spatial memory. Our work incorporated all the four common characteristics of rehearsal-based interfaces.

FastTap

Among these rehearsal-based interfaces, M3 Gesture Menu is most closely related to FastTap in terms of end-to-end objectives. FastTap and M3 were both proposed to improve the command selection efficiency on mobile form factors. They also split space into static hierarchical grids to provide space efficiency and constancy. The key difference lies in their input modalities. FastTap chunks two taps into a two-finger chord, whereas M3 chunks a sequence of items into a single gesture.

When more than two layers are involved, FastTap needs to temporally and sequentially compose multiple taps, whereas M3 remains a single unistroke. The use of multiple taps for hierarchical menus in the expert mode may introduce the same ambiguity with Multi-Stroke Marking Menu, but less confusing since one finger is consistently holding the activation button. FastTap may suit better when using tablets, when the number of commands does not require multiple layers of menus, or when one hand is steadily holding the phone, whereas M3 works well with only one hand and potentially scales up to multi-layer menus.

One important takeaway from FastTap research is that following the rehearsal hypothesis alone does not guarantee the transition to the expert technique [22, 34]. The transition may involve a global decision change by users driven by their intrinsic or extrinsic motivation such as the need for improving performance, perceived cost of error, and guidance.

Word-Gesture Keyboards

Word-gesture keyboards [52], also variably known as shape writing or gesture typing, is closely related to this work. Although conceived a decade later than marking menus, one of the key design objectives of word-gesture keyboards, namely seamlessly shifting the closed-loop visual tracing behavior to the open-loop memory recall gesture behavior, is the same as the vision of marking menus [51].

There are also major differences between shape writing and marking menus. Shape writing uses word shapes defined by the letter key positions on a keyboard layout to represent, encode, and decode information (words). In this regard, M3 is closer to shape writing than to marking menus which use radial directions to represent information. Unlike marking menus which "chunk" hierarchical selections into a gestural mark, shape writing chunks a sequence of movements from letter to letter into a word gesture.

Another difference is that word-gesture keyboards are stationary, so it is the user's choice to look, less intensively as they gain proficiency, at the graphical display of the keyboard. In contrast, hierarchical menus including M3 are dynamic. Each submenu is only displayed when an item on the top menu is selected. This raises the question if the empirical research on shape writing such as [51] that showed users could indeed memorize the word shapes also applies to M3.

Another relevant off-shoot of shape writing is the concept of "command strokes" [27] – issuing commands by gesture typing their names from a command key on the graphical keyboard. The obvious drawback of this approach in comparison to M3 is the space needed to display the graphical keyboard.

CONCLUSION

In this paper, we proposed M3 Gesture Menu, a variant of the marking menu that is defined on grids rather than radial space, relies on gestural shapes rather than directional marks, and has constant and stationary space use so as to fit the mobile form factor. In addition, gestures in M3 are more robustly recognized using a shape matching algorithm than aggregating stroke direction results as in conventional marking menus.

Two experiments were conducted to understand M3's menu resolution and novice to expert transition behavior. Compared to the literature, M3 exhibited much higher resolution than traditional Marking Menu and the revised Multi-Stroke Marking Menu. We estimated that M3 was at least twice as fast and twice more accurate than traditional Marking Menu and at least twice as fast as Multi-Stroke Marking Menu at a comparable accuracy. Furthermore, M3 was designed for mobile phone's relatively small screen and tested with the finger rather than the more precise stylus. Because gestures are unistrokes, M3 does not need a dedicated, occlusive space to disambiguate the menu interaction from other overlapping graphical elements as Multi-Stroke Marking Menu does.

We also for the first time, to our knowledge, systematically examined and demonstrated empirically successful user behavior transition from novice to expert mode. Users of M3 and Multi-Stroke Marking Menu were able to switch from visually-guided use to recall-based gesture articulation for a dozen commands after a total of 30-minute practice over three days. Subjectively, most but not all users preferred M3 over other alternatives. They found M3 to provide more fluid interaction flow, facilitate memorization of commands with expressive shape gestures, and enable more compact space use.

We also found on average it took only seven trials of use before users could switch from novice to expert behavior with M3 and Multi-Stroke Marking Menu. There is a cost to learning these marking menus because in the novice mode, they took a longer time to execute than a linear menu did. We estimated the time cost and saving equilibrium point for M3 versus linear menus at about 30 trials of use, after which each trial of use translates to estimated time savings of 1.7 seconds. Future designer and researchers may use these first-order approximations in their design context to estimate the cost and benefit of adopting menu innovations like M3.

In sum, by incorporating innovative features such as shape recognition, M3 Gesture Menu offers a more practical working solution for applying marking menu concepts on to touchscreen mobile devices. Our work also makes foundational contributions to future menu system design and research with empirical findings of M3 and marking menus' resolution and learning characteristics.

ACKNOWLEDGEMENT
We thank Daniel Vogel, Philip Quinn, Hemant Surale, and Alix Goguey for their valuable input.

REFERENCES
1. Fraser Anderson and Walter F. Bischof. 2013. Learning and performance with gesture guides. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1109–1118. http://dl.acm.org/citation.cfm?id=2466143
2. Caroline Appert and Shumin Zhai. 2009. Using Strokes as Command Shortcuts: Cognitive Benefits and Toolkit Support. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, 2289–2298. DOI: http://dx.doi.org/10.1145/1518701.1519052
3. Shiri Azenkot and Shumin Zhai. 2012. Touch Behavior with Different Postures on Soft Smartphone Keyboards. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*. ACM, 251–260. http://dl.acm.org/citation.cfm?id=2371612
4. Gilles Bailly, Eric Lecolinet, and Laurence Nigay. 2007. Wave Menus: Improving the Novice Mode of Hierarchical Marking Menus. In *Human-Computer Interaction – INTERACT 2007*. Number 4662 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, 475–488. http://link.springer.com/chapter/10.1007/978-3-540-74796-3_45
5. Gilles Bailly, Eric Lecolinet, and Laurence Nigay. 2008. Flower Menus: A New Type of Marking Menu with Large Menu Breadth, Within Groups and Efficient Expert Mode Memorization. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '08)*. ACM, 15–22. DOI: http://dx.doi.org/10.1145/1385569.1385575
6. Olivier Bau and Wendy E. Mackay. 2008. OctoPocus: A Dynamic Guide for Learning Gesture-based Command Sets. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology (UIST '08)*. ACM, 37–46. DOI: http://dx.doi.org/10.1145/1449715.1449724
7. Joanna Bergstrom-Lehtovirta, Sebastian Boring, and Kasper Hornb\a ek. 2017. Placing and Recalling Virtual Items on the Skin. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1497–1507. DOI: http://dx.doi.org/10.1145/3025453.3026030
8. Xiaojun Bi and Shumin Zhai. 2016. IJQwerty: What Difference Does One Key Change Make? Gesture Typing Keyboard Optimization Bounded by One Key Position Change from Qwerty. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 49–58. DOI: http://dx.doi.org/10.1145/2858036.2858421
9. Andrew Bragdon, Eugene Nelson, Yang Li, and Ken Hinckley. 2011. Experimental Analysis of Touch-screen Gesture Designs in Mobile Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, 403–412. DOI: http://dx.doi.org/10.1145/1978942.1979000
10. J. Callahan, D. Hopkins, M. Weiser, and B. Shneiderman. 1988. An Empirical Comparison of Pie vs. Linear Menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '88)*. ACM, 95–100. DOI: http://dx.doi.org/10.1145/57167.57182
11. Xiang Cao and Shumin Zhai. 2007. Modeling Human Performance of Pen Stroke Gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1495–1504.
12. Andy Cockburn and Carl Gutwin. 2009. A predictive model of human performance with scrolling and hierarchical lists. *Human–Computer Interaction* 24, 3 (2009), 273–314. http://www.tandfonline.com/doi/abs/10.1080/07370020902990402
13. Andy Cockburn, Carl Gutwin, and Saul Greenberg. 2007. A Predictive Model of Menu Performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 627–636. DOI: http://dx.doi.org/10.1145/1240624.1240723
14. Andy Cockburn, Carl Gutwin, Joey Scarr, and Sylvain Malacria. 2014. Supporting Novice to Expert Transitions in User Interfaces. *ACM Comput. Surv.* 47, 2 (Nov. 2014), 31:1–31:36. DOI: http://dx.doi.org/10.1145/2659796

15. Andy Cockburn, Per Ola Kristensson, Jason Alexander, and Shumin Zhai. 2007. Hard Lessons: Effort-inducing Interfaces Benefit Spatial Learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, 1571–1580. DOI: http://dx.doi.org/10.1145/1240624.1240863

16. Fergus IM Craik and Robert S Lockhart. 1972. Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior* 11, 6 (1972), 671–684.

17. Brian D Ehret. 2002. Learning where to look: Location learning in graphical user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 211–218.

18. Jérémie Francone, Gilles Bailly, Eric Lecolinet, Nadine Mandran, and Laurence Nigay. 2010. Wavelet Menus on Handheld Devices: Stacking Metaphor for Novice Mode and Eyes-free Selection for Expert Mode. In *Proceedings of the International Conference on Advanced Visual Interfaces (AVI '10)*. ACM, 173–180. DOI:http://dx.doi.org/10.1145/1842993.1843025

19. Bruno Fruchard, Eric Lecolinet, and Olivier Chapuis. 2017. MarkPad: Augmenting Touchpads for Command Selection. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 5630–5642. http://dl.acm.org/citation.cfm?id=3025486

20. Mayank Goel, Alex Jansen, Travis Mandel, Shwetak N. Patel, and Jacob O. Wobbrock. 2013. ContextType: Using Hand Posture Information to Improve Mobile Touch Screen Text Entry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2795–2798. DOI:http://dx.doi.org/10.1145/2470654.2481386

21. Tovi Grossman, Pierre Dragicevic, and Ravin Balakrishnan. 2007. Strategies for Accelerating On-line Learning of Hotkeys. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 1591–1600. DOI:http://dx.doi.org/10.1145/1240624.1240865

22. Carl Gutwin, Andy Cockburn, and Benjamin Lafreniere. 2015. Testing the Rehearsal Hypothesis with Two FastTap Interfaces. In *Proceedings of the 41st Graphics Interface Conference (GI '15)*. Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 223–231. http://dl.acm.org/citation.cfm?id=2788890.2788930

23. Carl Gutwin, Andy Cockburn, Joey Scarr, Sylvain Malacria, and Scott C. Olson. 2014. Faster command selection on tablets with FastTap. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2617–2626. http://dl.acm.org/citation.cfm?id=2557136

24. John I Kiger. 1984. The depth/breadth trade-off in the design of menu-driven user interfaces. *International journal of man-machine studies* 20, 2 (1984), 201–213.

25. Kenrick Kin, Björn Hartmann, and Maneesh Agrawala. 2011. Two-handed Marking Menus for Multitouch Devices. *ACM Trans. Comput.-Hum. Interact.* 18, 3 (Aug. 2011), 16:1–16:23. DOI: http://dx.doi.org/10.1145/1993060.1993066

26. Per Ola Kristensson and Shumin Zhai. 2004. SHARK2: A Large Vocabulary Shorthand Writing System for Pen-based Computers. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST '04)*. ACM, 43–52. DOI: http://dx.doi.org/10.1145/1029632.1029640

27. Per Ola Kristensson and Shumin Zhai. 2007. Command Strokes with and Without Preview: Using Pen Gestures on Keyboard for Command Selection. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, 1137–1146. DOI: http://dx.doi.org/10.1145/1240624.1240797

28. Gordon Kurtenbach and William Buxton. 1993. The Limits of Expert Performance Using Hierarchic Marking Menus. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI '93)*. ACM, 482–487. DOI: http://dx.doi.org/10.1145/169059.169426

29. Gordon Kurtenbach and William Buxton. 1994. User Learning and Performance with Marking Menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94)*. ACM, 258–264. DOI:http://dx.doi.org/10.1145/191666.191759

30. Gordon Kurtenbach, George W. Fitzmaurice, Russell N. Owen, and Thomas Baudel. 1999. The Hotbox: Efficient Access to a Large Number of Menu-items. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, 231–237. DOI:http://dx.doi.org/10.1145/302979.303047

31. Gordon Kurtenbach, Thomas P. Moran, and William Buxton. 1994. Contextual animation of gestural commands. In *Computer Graphics Forum*, Vol. 13. Wiley Online Library, 305–314. http://onlinelibrary.wiley.com/doi/10.1111/1467-8659.1350305/full

32. Gordon Paul Kurtenbach. 1993. *The Design and Evaluation of Marking Menus*. Ph.D. Dissertation. University of Toronto. https://autodeskresearch.com/pdf/theses/kurtenbach-phd.pdf

33. Gordon P. Kurtenbach, Abigail J. Sellen, and William A. S. Buxton. 1993. An Empirical Evaluation of Some Articulatory and Cognitive Aspects of Marking Menus. *Hum.-Comput. Interact.* 8, 1 (March 1993), 1–23. DOI: http://dx.doi.org/10.1207/s15327051hci0801_1

34. Benjamin Lafreniere, Carl Gutwin, and Andy Cockburn. 2017. Investigating the Post-Training Persistence of Expert Interaction Techniques. *ACM Trans. Comput.-Hum. Interact.* 24, 4 (Aug. 2017), 29:1–29:46. DOI:http://dx.doi.org/10.1145/3119928

35. Benjamin Lafreniere, Carl Gutwin, Andy Cockburn, and Tovi Grossman. 2016. Faster command selection on touchscreen watches. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 4663–4674. http://dl.acm.org/citation.cfm?id=2858166

36. Thomas K Landauer and Robert A Bjork. 1978. Optimum rehearsal patterns and name learning. *Practical aspects of memory* 1 (1978), 625–632.

37. David M. Lane, H. Albert Napier, S. Camille Peres, and Aniko Sandor. 2005. Hidden Costs of Graphical User Interfaces: Failure to Make the Transition from Menus and Icon Toolbars to Keyboard Shortcuts. *International Journal of Human-Computer Interaction* 18, 2 (May 2005), 133–144. DOI: http://dx.doi.org/10.1207/s15327590ijhc1802_1

38. Yang Li. 2010. Gesture search: a tool for fast mobile data access. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. ACM, 87–96. http://dl.acm.org/citation.cfm?id=1866044

39. Yuexing Luo and Daniel Vogel. 2015. Pin-and-Cross: A Unimanual Multitouch Technique Combining Static Touches with Crossing Selection. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, 323–332. DOI: http://dx.doi.org/10.1145/2807442.2807444

40. Sylvain Malacria, Gilles Bailly, Joel Harrison, Andy Cockburn, and Carl Gutwin. 2013. Promoting hotkey use through rehearsal with exposehk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 573–582. http://dl.acm.org/citation.cfm?id=2470735

41. Richard A. Schmidt, Douglas E. Young, Stephan Swinnen, and Diane C. Shapiro. 1989. Summary knowledge of results for skill acquisition: Support for the guidance hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15, 2 (1989), 352–359. DOI: http://dx.doi.org/10.1037/0278-7393.15.2.352

42. Barry Schwartz, Andrew Ward, John Monterosso, Sonja Lyubomirsky, Katherine White, and Darrin R. Lehman. 2002. Maximizing versus satisficing: happiness is a matter of choice. *Journal of personality and social psychology* 83, 5 (2002), 1178. http://psycnet.apa.org/psycinfo/2002-18731-012

43. Susanne Tak, Andy Cockburn, Keith Humm, David Ahlström, Carl Gutwin, and Joey Scarr. 2009. Improving Window Switching Interfaces. In *Human-Computer Interaction – INTERACT 2009 (Lecture Notes in Computer Science)*. Springer, Berlin, Heidelberg, 187–200. DOI: http://dx.doi.org/10.1007/978-3-642-03658-3_25

44. Susanne Tak, Piet Westendorp, and Iris van Rooij. 2013. Satisficing and the Use of Keyboard Shortcuts: Being Good Enough Is Enough? *Interacting with Computers* 25, 5 (Sept. 2013), 404–416. DOI: http://dx.doi.org/10.1093/iwc/iwt016

45. Huawei Tu, Xiangshi Ren, and Shumin Zhai. 2015. Differences and Similarities Between Finger and Pen Stroke Gestures on Stationary and Mobile Devices. *ACM Trans. Comput.-Hum. Interact.* 22, 5 (Aug. 2015), 22:1–22:39. DOI: http://dx.doi.org/10.1145/2797138

46. Md Sami Uddin and Carl Gutwin. 2016. Rapid Command Selection on Multi-Touch Tablets with Single-Handed HandMark Menus. In *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces*. ACM, 205–214. http://dl.acm.org/citation.cfm?id=2992172

47. Md Sami Uddin, Carl Gutwin, and Andy Cockburn. 2017a. The Effects of Artificial Landmarks on Learning and Performance in Spatial-Memory Interfaces. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3843–3855. http://dl.acm.org/citation.cfm?id=3025497

48. Md. Sami Uddin, Carl Gutwin, and Alix Goguey. 2017b. Artificial Landmarks Augmented Linear Control Widgets to Improve Spatial Learning and Revisitation Performance. In *Proceedings of the 5th Symposium on Spatial User Interaction (SUI '17)*. ACM, New York, NY, USA, 153–153. DOI: http://dx.doi.org/10.1145/3131277.3134367

49. Md Sami Uddin, Carl Gutwin, and Benjamin Lafreniere. 2016. HandMark Menus: Rapid command selection and large command sets on multi-touch displays. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5836–5848. http://dl.acm.org/citation.cfm?id=2858211

50. Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 143–146.

51. Shumin Zhai and Per-Ola Kristensson. 2003. Shorthand Writing on Stylus Keyboard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, 97–104. DOI: http://dx.doi.org/10.1145/642611.642630

52. Shumin Zhai and Per Ola Kristensson. 2012. The Word-gesture Keyboard: Reimagining Keyboard Interaction. *Commun. ACM* 55, 9 (Sept. 2012), 91–101. DOI: http://dx.doi.org/10.1145/2330667.2330689

53. Shumin Zhai, Per Ola Kristensson, Caroline Appert, Tue Haste Andersen, and Xiang Cao. 2012. Foundational Issues in Touch-screen Stroke Gesture Design – An Integrative Review. *Foundations and Trends in Human-Computer Interaction* 5, 2 (2012), 97–205. https://hal.inria.fr/hal-00765046/

54. Shengdong Zhao, Maneesh Agrawala, and Ken Hinckley. 2006. Zone and Polygon Menus: Using Relative Position to Increase the Breadth of Multi-stroke Marking Menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, 1077–1086. DOI: http://dx.doi.org/10.1145/1124772.1124933

55. Shengdong Zhao and Ravin Balakrishnan. 2004. Simple vs. Compound Mark Hierarchical Marking Menus. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST '04)*. ACM, 33–42. DOI: http://dx.doi.org/10.1145/1029632.1029639

56. Jingjie Zheng. 2017. Enabling Expressive Keyboard Interaction with Finger, Hand, and Hand Posture Identification. (Sept. 2017). https://uwspace.uwaterloo.ca/handle/10012/12371