The Performance of Touch Screen Soft Buttons

Seungyon "Claire" Lee[§*] **Shumin Zhai**[§]

[§] IBM Almaden Research Center
650 Harry Road, San Jose, California
zhai@almaden.ibm.com

[*] Georgia Institute of Technology
Atlanta, Georgia
sylee@gatech.edu

ABSTRACT
The introduction of a new generation of attractive touch screen-based devices raises many basic usability questions whose answers may influence future design and market direction. With a set of current mobile devices, we conducted three experiments focusing on one of the most basic interaction actions on touch screens: the operation of soft buttons. Issues investigated in this set of experiments include: a comparison of soft button and hard button performance; the impact of audio and vibrato-tactile feedback; the impact of different types of touch sensors on use, behavior, and performance; a quantitative comparison of finger and stylus operation; and an assessment of the impact of soft button sizes below the traditional 22 mm recommendation as well as below finger width.

Author Keywords

Touch screen, mobile, input, feedback, stylus, finger, buttons, keyboard, tangible interface.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI).

INTRODUCTION
Touch screens are currently gaining popularity in the information technology industry. In the second half of 2007, LG, HTC and Apple Inc all released new touch screen mobile phones, while Microsoft announced its surface computing initiative. Later many more mobile phone companies followed suit. Touch screens can also be found on newer desktop consumer PCs such as HP's TouchSmart line of products. This rise in popularity is a continual development of decades of touch screen applications in early PDAs, tablet PCs, ATMs, auto-checkout machines, and even casino gambling machines.

Although it is more feasible to employ fluid finger/pen gestures on touch screens than on mouse based computers, the use of soft buttons (which are synonymous with keys and active icons in this paper) remains the most basic interaction action on touch screens. There are obvious advantages to soft buttons compared to their hard counterparts. First and foremost, because soft buttons are rendered graphically on the touch screen, they can appear or disappear according to the interaction context. They can be bigger or smaller depending on the screen-space availability. They can also be stylized or updated in appearance according to fashion and taste more easily. It is also easier to design clean looking sleek all soft button based devices than arrays of hard button based devices.

However, soft buttons also have disadvantages. Chief among them is the lack of tactile feedback. Though the ideal hard button mechanical properties such as the force-displacement profile are not completely clear in the literature (and there appears a quite wide acceptable range of hard button force and displacement), there is no doubt that tactile feedback is important [16]. For example, typing on flat piezo-electric buttons that had no tactile feedback significantly reduced expert typists' performance [2]. Indeed, the initial concerns and criticisms by the press, technology reviewers, and bloggers on the iPhone after its announcement were largely focused on its all touch screen design. With the rise of interest in applying modern touch screen technologies in mobile forms and beyond, many user interface questions arise: Do soft buttons perform worse than hard buttons? If so by how much (so rational design decisions can be made). If the answer is "it depends," then what factors contribute to the performance of soft buttons? Can synthetic audio or the commonly available vibrato-tactile feedback compensate for the lack of mechanical properties? Do different types of touch sensors affect soft button performance? What is the performance difference between stylus and finger operated soft buttons?

Given the renewed interest in touch screen interfaces and the general importance of understanding input properties to HCI, we conducted a series of three basic empirical experiments on the performance of soft buttons. In the rest of the paper, we first briefly review related work. We then outline the basic questions to be studied in the current work and a few general methodology decisions. We then report the empirical experiments addressing these questions. Last we draw a set of conclusions and discuss design implications and directions for future research.

RELATED WORK

The past literature on touch screen soft buttons is limited and mostly based on older touch screen technologies in large desktop settings. Lewis and colleagues [16] give a comprehensive survey on all types of keys and keyboards. Lewis [14] and Greenstein [8] both reviewed touch screen viewing angle, soft button size-accuracy tradeoff and target selection strategies. Buxton and colleagues [3, 5] provide a high level behavioral analysis and design insights on using touch tablet primarily as an analogue input device in comparison to other input devices such as the mouse. An influential series of studies at the University of Maryland (e.g. [17, 21]) focused on target selection strategies on touch screens. They found that a "take-off" strategy in which a cursor placed ½ inch above the finger tip was less error prone (but slower) than the more common "land-on" strategy. More recent studies along this line (e.g. [1]) explored other more complex techniques (e.g. using a graphic "precision handle" that amplify the finger tip movement's precision) which further trade off speed for higher precision. This line of studies is more aimed at target selection than soft buttons which are a special case of screen "targets" where speed and single step activation, rather than multiple feedback based adjustment steps, is more desirable.

Sears [19] studied pointing biases due to thick touch screen parallax and other factors when the screen was mounted 30 degrees from horizontal. He found that typing with two hands and multiple fingers on a soft keyboard in this setting after bias correction was still far worse than on a standard physical keyboard, but faster than using a mouse on the same soft keyboard, achieving on average 25.4 WPM (or 0.47 s/character, or 2.12 CPS — character per second) speed. Later Sears and colleagues [20] studied soft keyboard performance in varying sizes: 24.6, 13.2, 9.0 and 6.8 mm per side (22.7, 11.4, 7.6, 5.7 mm not counting the gap between keys). They found that typing speed was slower with smaller keys, but still possible with the smallest keys tested with a take-off cursor (above the finger tip). Land-on cursor was used for the larger sizes in the study. Today's mobile devices usually do not use cursors.

More recently, Popyrev and Maruyama explored tactile feedback on a small touch screen [18]. Although not the focus of this paper, the understanding of soft buttons in comparison to hard buttons also bears relevance to the foundation of another area of user interface research – tangible user interfaces (TUI). TUI offers a compelling alternative to graphical user interfaces (GUI), due to its physical and "graspable" interface representation [7, 11, 24]. One often can feel that using a simulated calculator on a computer screen with a mouse is less convenient than using a physical calculator.

BASIC ISSUES AND RESEARCH METHODS

In order to study the efficacy of soft buttons on touch screens in comparison to that of hard buttons, we first classify soft buttons along a set of dimensions or attributes. Touch sensing can be achieved through various physical mechanisms, including the change of electrical resistance, capacitance, inductance, acoustic wave, and optical occlusion or reflection. More relevant to user interface design is the different types of behavior the touch sensors may afford. We identify a few attributes that are particularly important in this regard. Note that we take a broad definition of touch screens, including those that are sensitive to touch with an intermediary implement (a stylus).

Basic issues

Operational mode (stylus vs. finger)

Some sensors, such as the inductive touch screens commonly found in today's tablet PCs (e.g. the Lenovo Thinkpad X61), can only be operated with a stylus. Other sensors, such as the capacitive sensor found in the Apple iPhone can only be operated with fingers. In contrast, either a stylus or fingers can operate resistive sensors or optical sensors, such as those found in Palm Pilot PDAs or the HP TouchSmart PCs. The choice of the operational mode, whether stylus or finger, causes a number of obvious behavior differences, including absolute precision, dexterity, and ease of access (pulling out a stylus vs. using a finger directly). The speed and accuracy performance difference between using stylus and using finger to operate a soft button is one of the dimensions we investigate in the current work.

Activation mechanism (contact vs. force)

A capacitive or an optical sensor can be activated by a bare finger contact whereas a resistive sensor is activated by a perceptible level of force, however slight. This difference in technology may afford different behavior. For example a resistive screen is more responsive when tapped with the fingernails than with finger tips (pads). A capacitive sensor, on the other hand, is only responsive to finger skin but not to the fingernail. The difference is particularly pronounced when drawing a stroke on a touch screen. On a resistive screen, because one has to pass a certain level of force threshold to keep the sensor engaged, one can only reliably and easily draw a stroke with a stylus or finger nail, but not finger tip (see [3] for a very detailed analysis on "inking" on a pressure-sensitive touch tablet). On a capacitive touch screen, on the other hand, one can draw a stroke with light contact with a finger tip. Whether activation by force or by contact makes a substantial difference on soft button tapping performance is another dimension we study in this paper.

Feedback enhancements

In addition to visual feedback (changing size or color), the two common means to provide synthetic feedback to compensate for the lack of intrinsic tactile feedback in soft buttons are audio and vibrato-tactile feedback. Both are available in current mobile phones. Their effectiveness and how they interact with each other will be also explored in the current work.

Button size

Previous literature suggests that for soft buttons to work well with fingers, the button size need to be larger than 22 mm in width [8, 14]. The anthropomorphic average width of the index finger and the thumb for adult men are 18.2 mm and 22.9 mm respectively and women 15.5 mm and 19.1 mm respectively [9]. Sears and colleagues demonstrated that it was still possible to tap on soft buttons 6.8 mm per side with a take-off cursor. Although 22 mm wide soft buttons are possible to implement on, for example, supermarket auto-checkout machines, they are bigger than what mobile devices can accommodate. We are interested in whether the range of button sizes offered on today's mobile devices can still effectively work without a cursor. We are also interested in identifying where the lower limit is below which soft button performance dramatically deteriorates.

There are other dimensions to touch screen properties which are impossible to explore all at once. We focus on the above four properties since they are fundamental but their impact on performance is not obvious or cannot be quantitatively known without systematic empirical studies. Once we establish a basic understanding of these dimensions, further research issues will be exposed and can be pursued in future studies.

Methodology

As in all fields, HCI research can be conducted at different levels of analysis. Lower level analyses tend to manipulate one or two general and "pure" conceptual attributes of user interfaces (e.g. [4, 6, 26]) while holding all other behavioral variables constant. While work at this level is fundamental, its results often require further work to be applied to practical questions if possible at all due to the fact that dimensions in actual interfaces are often coupled. The current investigation aims at a higher and practical level of analysis but still targets at concepts that are more general than specific models of products. We chose a set of representative commercial products for this series of experiments, so that the conclusions we draw from the results of the study are likely to reflect what an actual user would experience with each type of product in today's market place. In total five off-the-shelf state-of-the-art products were used in our study. They were P1) an Apple iPhone (contact activated capacitive sensor) operated by finger, P2) an HTC Touch phone (force activated resistive sensor) operated by finger or stylus, P3) a Canon DK10i calculator (with hard buttons), P4) an iPAQ 6315 that had the same type of touch screen as in P2, but more similar in size to P3, and P5) a Nokia 6260 phone representing mass market hard button based mobile devices. During the experiments, these devices were hand held by the participants in any way that they felt was the most natural.

The level of analysis decision also concerns the amount of experimental control that needs to be exerted in this type of investigation. To be at a practical level we gave the participants certain amount of flexibility in the way they operate the devices (with their preferred finger or thumb) in the first two experiments. In a similar spirit, we choose two simple, practical and representative tasks in the study one involving calculator operation and the other phone dialing, both using numeric and operator buttons. The label (or function) on the buttons were not intrinsic to button performance. We chose not to involve a text entry task (alphabetic keys) because it is more complex and more dependent on factors such as layout familiarity, although the results here will be relevant to text entry tasks as well.

A total of 13 volunteers (9 male 4 female, age from 20's to 40's) participated in the study. Seven of them reported daily experience of touch screen devices. Each study condition followed a three-phase procedure: practice, test, and questionnaire plus interview session. During the practice session, participants explored the testing device, the task procedure, and their operational preference for as long as they needed to. In the test session, fast and accurate performance with error correction was encouraged.

When appropriate, study data were tested through repeated measurement variance analysis for statistical significance. Post hoc tests following main and interaction effect report were done by Fisher's PLSD. Standard deviations are reported in parentheses following means. Results graphs show mean values with 95% confidence interval error bars. For brevity and readability, these explanations will be repeated to the minimum extent.

EXPERIMENT 1: OPERATING MODE AND FEEDBACK

Goals and Task

This experiment simultaneously investigates two variables: operating modes (finger, stylus, and hard buttons) and various feedback conditions (audio, vibrato-tactile, both, and none) See Figure 1 for details.

The task used in this experiment was a simple multiplication operation involving numeric (0~9) and operator (x and =) buttons on calculators. Each trial of the task involved entering 8 digits and 2 operators (e.g. *1450*X*9276=)*. To measure learning effect and its interaction with other independent variables (IVs), 5 fixed sets of target characters (digits and operators) were repeated in 3 blocks of trials. The order of the 5 sets was randomly shuffled in each block of trials. In this within subject experiment, the order of the feedback conditions was randomized while the order of the three operating modes (hard, soft-finger, soft-stylus) was balanced across 12 participants in a Latin square pattern. The 13th participant had the same order as the 1st participant.

Apparatus and Feedback Conditions

Hard Button conditions

For the hard button conditions, we used P3 (Canon calculator), measured 71 mm (W) x 122 mm (H) x 13 mm (T) in size, with each button measured in 12.8 mm x 10.5 mm[1] in size. The calculator was connected to the test application running in another computer through a USB port. The application written in Adobe Flash displayed the target and entered characters, provided audio feedback and captured the time stamped data for speed and accuracy analysis. The calculator was held close to the bottom of the test computer's LCD monitor (Figure 1).

Soft Button Conditions

The soft button conditions were implemented on P4 - an HP iPAQ 6315 smart phone measuring 76 mm(W) x 122 mm(H) x 18 mm(T) in size. The iPAQ touch screen could be activated by either finger or stylus.

Hard button		Soft button - stylus				Soft button - finger			
									
A	N	A	T	A+T	N	A	T	A+T	N

Figure 1. Apparatus and conditions in Experiment 1 (A: audio, T: vibrato-tactile, N: None)

A calculator application that displayed target characters, entered characters and soft buttons was written in C#. The soft button layout approximately matched that of the hard button calculator. Each button was 10.4 mm x 10.0 mm in size (Figure 1). The audio feedback was a 130ms long system beep sound and the 50 ms long vibrato-tactile feedback were implemented through the built-in actuator in the iPAQ which could be both heard and felt in the holding hand of the device.

Overall 10 (conditions) x 3 (blocks) x 5 (trials) x 13 (participants) = 1950 trials of tests were collected in this experiment. Each trial involved striking 10 characters (plus correction strokes).

[1] Here and throughout the paper, button sizes are measured from the center of one button to the center of the next button, including the gap between the buttons because 1. This is the screen space a key effectively take. 2. The gap contributes to error tolerance of the keys.

Results

The accuracy in all conditions and blocks of trials were close to 100%. The lowest accuracy was 98. (7.9) percent in Block 3 of the hard button without audio feedback condition whereas the highest accuracy was 99.4 (2.4) percent in Block 1 of finger soft button with audio feedback condition. The consistently high accuracy results enabled us to focus on speed and the number of correction button presses to discriminate the different conditions in the experiment. For ease of understanding, in what follows we first focus on performance within the hard button conditions, followed by performance within the soft button conditions, and finally comparison between hard and soft buttons.

Hard Button Speed

The mean character entry speed, calculated from the ratio between the number of characters entered in each trial (always 10 in Experiment 1) and the trial completion time, improved from 2.64 characters per second (CPS) in Block 1 to 2.89 CPS in Block 3. This improvement was statistically significant ($F_{2,24}$ = 11.36, p = 0.0003). Although the participants had prior experience in using hard buttons such as mobile phones, they might have gained skills for the specific device and task by, for example, memorizing some of the numbers used in the experiment. The practice effect, however, had no significant interaction with the feedback conditions (with or without audio): $F_{2,24} = 0.4$, $p = 0.68$.

The addition of audio feedback brought no significant improvement to the hard button performance: The mean speed was 2.75 (0.71) CPS with audio and 2.73 (0.74) CPS without audio: $F_{1,12}$ = 0.22, p = 0.65. The intrinsic kinesthetic feedback in the hard buttons provided sufficient feedback so the audio feedback provided no additional functional value. Interestingly, participants' mean subjective rating on a -3 to + 3 scale was 1.15 (1.86) with audio but 0.15 (1.57) without audio. However this difference was hardly significant given the large individual variances (p=0.08).

As one would expect from Fitts' law (the impact of distance and target size on target acquisition time), the different sets of characters afforded significantly different mean speeds (from 2.58 to 2.81 CPS, $F_{4,48}$ = 3.89, p = 0.008) since the inter-character distances varied between the sets. There was no significant interaction between the character sets and other IVs. Since Fitts' law effect (which may concern the layout of soft letter keys [15, 23, 27] is not the focus of the current study, we will not report further data on this dimension. See the attempts in standardizing the use of Fitts' law for input device evaluation [22] and associated complications [10, 25, 28].

Amount of Corrections in Hard Button Conditions

The mean number of key presses per character entered (KPC) was 1.01 (.025) and 1.022 (.055) with and without audio feedback respectively. In other words, there were 1 to 2 additional key presses made for every 100 characters entered. Note that since the calculation of speed performance in CPS was based on the number of effective characters entered in the end, the additional correction presses panelizes the speed performance in CPS.

Soft Button Speed

The overall mean speed with the soft buttons also improved significantly from 2.52 CPS in Block 1 to 2.73 CPS in Block 3 ($F_{2,24}$ = 14.33, p < 0.0001). However this practice effect did not interact significantly with other IVs.

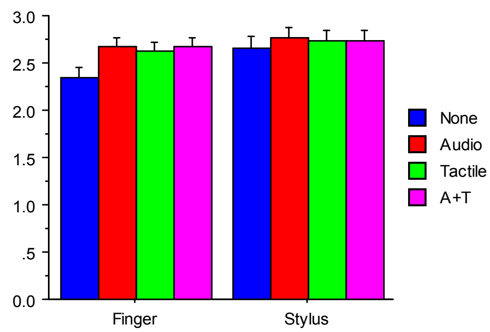

Figure 2. Input speed as measured by characters per second (CPS) in the soft button conditions in Experiment 1

Different feedback conditions significantly influenced speed performance ($F_{3,36}$= 12.61, p < 0.0001, see Figure 2). In contrast, there was no significant difference between the mean speeds of the two operating modes: Finger was 2.58 CPS and Stylus was 2.72 CPS, $F_{1,12}$ = 1.45, p = 0.26. There was a significant interaction between operating mode and feedback enhancement conditions ($F_{3,36}$= 3.63, p = 0.022). As shown in Figure 2, feedback (of various types) improved the speed of finger operated soft buttons more than stylus operated soft buttons. It appears that the sharper stylus impact on the screen felt in the hand, and possibly the sound generated, both coincided with the button actuation in time, provided good confirmatory feedback. In contrast the bare fingertip touch on the resistive sensor screen might not give accurate confirmation on whether the force threshold required was passed to activate the soft button. The fingernail may have a similar feel of impact as the stylus. However only 4 of the 13 participants used their fingernails in this experiment.

The mean soft button operating speed in each of the feedback enhancement conditions (audio, vibrato-tactile, and both) was significantly higher than the mean speed with no feedback enhancement (p < 0.0001), but none of the pair-wise comparisons between different types of feedback enhancement was significant (p > 0.05). Particularly noteworthy was the fact that while audio and vibrato-tactile feedback both improved soft button's operating speed, they made no further improvement together.

Amount of corrections in soft button conditions

Both feedback conditions ($F_{3,36}$= 12.8, p < .0001) and operating mode (finger vs. stylus, $F_{1,12}$ = 12.5, p = .004) as well as their interaction ($F_{3,36}$= 7.8, p = .0004) had significant impact on the Key Presses per Character measure. As shown in Figure 3 and Table 1, the number of additional correction strokes was higher with finger operation than with stylus operation, particularly when there was no feedback. The mean KPC in the finger and stylus conditions were 1.011 (.026) and 1.036 (.067) respectively. Quite plausibly, more errors were made (but corrected) with the bare figure than with the stylus since the former is wider and less precise.

	Count	Mean	Std. Dev.
Stylus, None	39	1.016	.035
Stylus, Audio	39	1.009	.019
Stylus, Vibrato-tactile	39	1.009	.023
Stylus, A+T	39	1.010	.024
Finger, None	39	1.089	.095
Finger, Audio	39	1.019	.049
Finger, Vibrato-tactile	39	1.014	.028
Finger, A+T	39	1.019	.048

Table 1. KPC in the soft button conditions

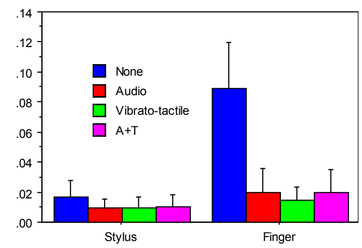

Fig 3. Additional number of key presses per character in Experiment 1

Post hoc tests show that the KPC values were lower in all feedback enhancement conditions than in the no feedback enhancement condition (p < .0001) but no significant difference was measured between any pair of feedback enhancements (audio, vibrato-tactile, and both).

Effect: Conditions	Count	Mean	Std. Dev.
Hard Key Audio	13	1.154	1.864
Hard Key no Audio	13	.154	1.573
Soft Key Stylus Audio	13	1.154	1.951
Soft Key Stylus Vibrato-tactile	13	.462	1.664
Soft Key Stylus Audio + Vibrato-tactile	13	1.000	1.871
Soft Key Stylus No Feedback	13	-.385	1.938
Soft Key Finger Audio	13	.692	1.750
Soft Key Finger Vibrato-tactile	13	.769	1.691
Soft Key Finger Audio + Vibrato-tactile	13	.769	1.878
Soft Key Finger No Feedback	13	-1.308	1.750

Table 2. Subjective rating on a -3 to +3 scale in Experiment 1

Subjective ratings and comments

Subjectively, among all soft button conditions participants were most positive on the stylus operated soft buttons with audio feedback and most negative on finger operated soft buttons with no feedback (Table 2). Both operating mode (finger vs. stylus, $F_{1,12}$ = 5.47, p = 0.038) and feedback conditions ($F_{2,24}$ = 4.7, p = .007), as well as the interaction between the two variables ($F_{2,24}$ = 3.3, p = .031) were significant. On average, the participants rated stylus operation more positively than finger operation. All feedback enhanced conditions were rated significantly higher than no feedback enhancement but the differences between the different types of feedback enhancements were not significant.

Comparison between hard and soft buttons

We now focus on comparing the hard and the soft button conditions for the calculator task. Since there was no significant difference between the feedback types, we use audio feedback as a representative soft button feedback condition so the comparison to the hard button operation is balanced. With this reduced and balanced data set, we see the following significant factors to the mean operating speed in the experiment: Trial block ($F_{2,24}$ = 17.2, p <0.0001), Feedback (with and without audio, $F_{1,12}$ = 23.8, p = 0.0004), and Feedback X Button Type (hard, soft with finger, soft with stylus) interaction ($F_{2,24}$ = 6.4, p = 0.006). As illustrated in Figure 4, audio feedback improves the speed of the soft buttons, particularly the speed of finger operated soft buttons. With audio feedback, the operating speeds of the three types of buttons were quite comparable. None of the other factors or interactions, including Button Type (hard, soft with finger, soft with stylus, $F_{2,24}$ = 1.84, p = 0.18) was significant.

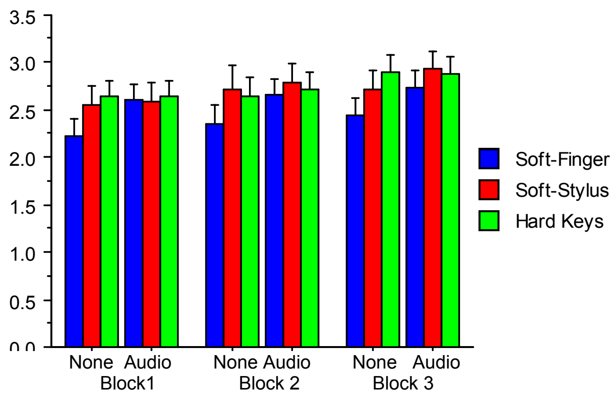

Figure 4. Speed (CPS) in Experiment 1: soft vs. hard buttons and audio vs. none feedback conditions

Summary of Experiment 1

The most informative findings of this experiment in our view were as follows: In comparison to the hard buttons, soft buttons were quite usable, either operated by finger or with stylus. Their performance (effective speed) is comparable to that of hard buttons, particularly when synthetic feedback enhancement was provided. Finger operation of soft buttons tended to be more error prone than stylus operation. Either audio or vibrato-tactile feedback helps the performance of soft buttons, but together they made no further improvement.

EXPERIMENT 2. CONTACT VS FORCE ACTIVATION

Goal and Motivation

Unlike Experiment 1 that involved a resistive touch sensor only, this experiment includes a capacitive sensor. As discussed earlier, the behavior difference afforded by these two types of sensors lies in the activation mechanisms: a resistive sensor is activated by a force threshold and a capacitive sensor by skin contact. Since there is no force threshold required, it is conceivable that contact activated touch screen can be perceived more responsive. The main goal of this experiment was to explore the performance difference between the two activation mechanisms and their comparison to hard buttons.

Apparatus, Design and Procedure

Three mobile phones (P5-Nokia 6260, P2-HTC Touch, P1-Apple iPhone, see Figure 5) were used representing hard button, force-activated soft button (resistive sensor), and contact-activated soft button (capacitive sensor) devices. Given the results in Experiment 1 on feedback, audio feedback was used throughout this experiment. The button sizes were 13.3 mm x 8.0 mm (hard button), 13.8 mm x 9.8 mm (force soft button), and 19.0 mm x 10.3 mm (contact soft button) respectively. The size of the contact activated soft buttons was bigger than the other two types of buttons. In this regard the set-up favored the contact activation condition.

The task used in this experiment was 7 digit phone number dialing. In the Nokia and the HTC phones, the target phone number to be dialed and number entered were displayed on the phone's LCD screen by test programs written in Python and C# respectively. On the iPhone, due to the restriction of software installation at the time of the experiment, the numbers were displayed on an external monitor next to the top of the phone. In each block of trials, 5 unique phone numbers, randomly shuffled in order, were presented and entered. This procedure was repeated in three blocks of trials. The participants, order balancing, and procedure in this (and the next) experiment were the same as Experiment 1. A total of 3 (conditions) x 3 blocks x 5 trials x 13 participants = 585 trials were recorded in the test sessions.

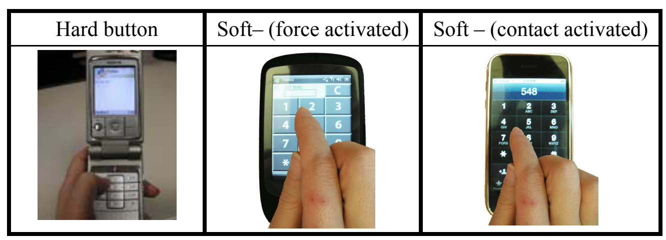

Figure 5. Apparatus and conditions in Experiment 2

Due to the programming restriction at the time, we used iPhone's built-in dial pad in this experiment and digital audio video recording to capture the test sessions for data analysis. Data were recorded in the test software installed on the Nokia and the HTC phone. The soft buttons on the HTC phone displayed by the test program matched those of the iPhone dial pad as closely as possible.

Results

Similar to Experiment 1, the participants consistently maintained a high level of end results accuracy in all conditions. The accuracy mean in the contact activated soft button, the force activated button, and the hard button conditions were 99.9 (1.02), 99.2 (4.2) and 99.6 (2.47) percent respectively.

Speed

Both button type (force activated soft button, contact activated soft button, and hard button: $F_{2,24}$ = 4.76, p = 0.018) and trial block ($F_{2,24}$ = 8.36, p = 0.0018) had significant main effect on speed. There was no significant interaction between the two. The means speed for the contact activated soft buttons, the force activated soft buttons, and the hard buttons were 2.56 (0.54), 2.65 (0.81), and 2.37 (0.60) CPS respectively (Figure 6). Post hoc analyses show that the difference between the two types of soft buttons were not significant (p = .34) but the difference between the soft and the hard button conditions was either significant (force activated soft button vs. hard button: p = 0.006) or borderline significant (contact activated soft button vs. hard button: p = 0.0515). Despite the difference in task, the mean CPS speeds in the soft buttons were very similar to those in Experiment 1 with audio feedback, indicating consistent performance across devices and tasks.

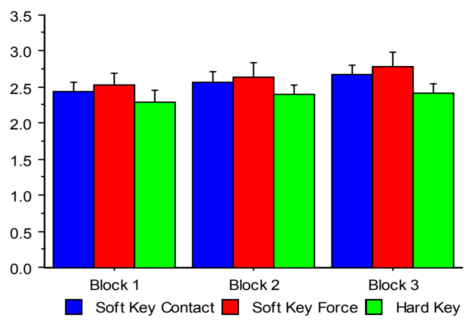

Figure 6. Speed (CPS) in Experiment 2

Amount of corrections

Very few error correction strokes were made in any conditions, as reflected by the low mean values of KPC: 1.009 (.031), 1.019 (.044), and 1.009 (.021) for hard buttons, forced activated soft buttons and contact activated soft buttons respectively, with no significant difference ($F_{2,24}$ = .98, p = 0.39).

Subjective ratings and comments

Subjectively, the participants' rating on a -3 to +3 scale across the conditions had no significant difference ($F_{2,24}$ =.058, p = 0.94, Table 3).

	Count	Mean	Std. Dev.	Std. Err.
Hard Button	13	.538	1.713	.475
Soft Button Force	13	.538	1.713	.475
Soft Button Contact	13	.346	1.819	.504

Table 3. Participants' subjective ratings in Experiment 2

The soft buttons were on average faster than the hard buttons. It is also somewhat surprising that the contact activated soft buttons were not on average faster than force activated soft buttons, although they felt more responsive and in this experiment were bigger in size. In the post test interview, several participants reported that although they felt force activated soft buttons were less responsive than contact activated buttons, the former allowed them to put fingers on the surface without accidentally trigger a button whereas the latter did not.

Summary

This experiment shows that soft buttons could afford even better performance than some of the hard buttons embedded in common mobile phones. With different pros and cons, the performance difference between the two soft button activation mechanisms (force vs. contact) is quite subtle.

EXPERIMENT 3. BUTTON SIZE AND MODE OF ACTIVATION

Motivation

Experiment 1 and 2 have shown that the soft buttons performed rather well in the tasks and settings tested. In both experiments although below the 22 mm recommendation [8, 14], the button sizes tested were still large by mobile standard. While it is possible to have a phone number dial pad in such size, many other applications, such as text and numeric input into email or a web browser demand much less space devoted to soft buttons for each character. We hypothesized that the soft button performance will begin to deteriorate if the size of the button is further reduced. In this experiment we tested two smaller sizes of buttons, both drawn from a commercial product (iPhone).

It is possible that the more demanding smaller buttons will differentiate the activation mechanism (contact vs. force) further than what was found in Experiment 2 so in this experiment we included activation mechanism as another IV, resulting in a 2 (sizes) x 2 (activation mechanisms) design.

Apparatus, Design and Procedure

The apparatus, design, procedure and participants were all the same as Experiment 2's soft buttons conditions, except the layout, the button size, and use of finger. In each trial, the target and entered number were displayed adjacently. The one row numeric buttons were part of the alphanumeric soft keyboard of the iPhone (Figure 7). The wide condition was when the iPhone was held laterally (landscape mode) and the narrow condition vertically (portrait mode). The button size was 7.2 mm (W) x 6.5 mm (H) in the wide and 4.9 mm x 8.3 mm in the narrow conditions. These are less than half of average adult women finger width [9]. In pilot tests, we observed as the button size gets smaller, participants tended to use the index finger instead of the thumb. In order to explore the interrelationship between the button size and activation mechanism as clearly as possible, we decided to ask the participants to use the index finger only for this experiment.

On the force activated touch surface of HTC phone (P2), a test application written in Adobe Flash was run to simulate the iPhone soft buttons in the same size, layout and visual feedback (letter pop-up on button press). To maintain the same button size as the iPhone on a smaller screen of the

HTC Touch, only a subset of buttons (4, 5, 6, 7, 8, 9, 0) were displayed and the numbers in the task did not include digit 1, 2, and 3. For iPhone (P1), an application written in JavaScript through a web browser displayed the target and entered numbers on the iPhone. User actions on the iPhone were captured through digital audio video recording. A total of 4 (conditions) x 3 blocks x 5 trials x 13 participants = 780 trials were recorded in the tests.

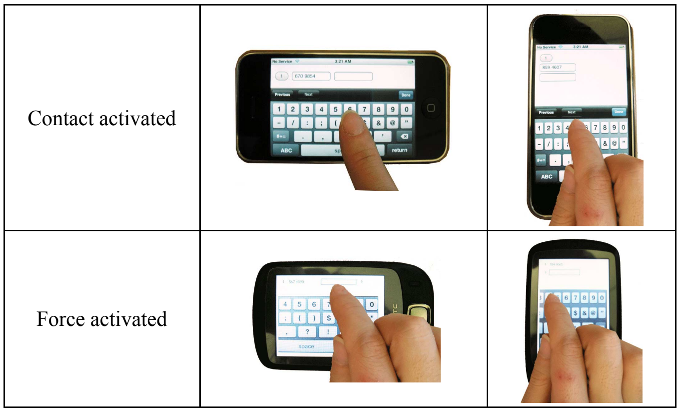

Figure 7. Apparatus and conditions in Experiment 3

Results

Participants maintained high level end result accuracy in this experiment as well. The mean accuracy in the force and contact activation conditions were 98.6 (5.57) and 97.77 (7.00) percent respectively. The mean accuracy in the wide and narrow button conditions were 98.6 (6.00) and 97.7 (6.64) percent respectively.

Speed

The practice effect was less strong in this experiment. The mean speed in the three blocks of trials was 1.77 (0.62), 1.82 (0.62) and 1.83 (0.60) CPS respectively. The difference among them was not significant ($F_{2,24} = 0.77$, $p = 0.47$) .

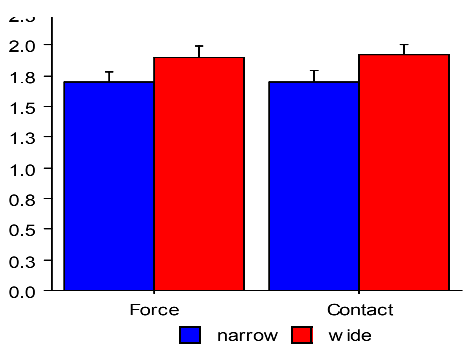

Figure 8. Speed in CPS in Experiment 3

The mean characters entered per second in the force activated and contact activated soft button conditions was 1.80 (0.59) and 1.81 (0.63) CPS respectively, with no significant difference ($F_{1,12} = 0.42$, $p = 0.84$, Figure 8,).

The difference in mean speed in the wide and narrow button conditions, 1.91 (0.60) and 1.70 (0.61) CPS respectively, was significant ($F_{1,12} = 11.4$, $p = 0.005$). In comparison to the mean of the soft button conditions in Experiment 2 (2.61 CPS) with wider buttons, the means speeds in this experiment were 26.7% and 34.7% slower respectively.

Amount of corrections

Quite informatively, the KPC values, which reflect the number of corrective strokes made in order to enter the right digit, were much higher in this experiment (from 7% to 25% additional stroke, Table 4, Figure 9) than the previous two experiments (less than 3% with feedback).

	Count	Mean	Std. Dev.
Contact, wide	39	1.079	.129
Contact, narrow	39	1.245	.263
Force, wide	39	1.075	.111
Force, narrow	39	1.070	.088

Table 4, KPC in Experiment 3

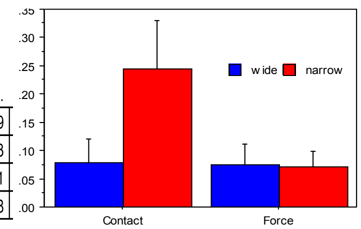

Figure 9. Additional no. of key strokes made per digit entered in Experiment 3

As shown in Figure 9, the experimental conditions had a significant impact on KPC ($F_{3,36} = 5.57$, $p = 0.003$). In particular, contact activated narrow buttons were more error prone than all other conditions ($p \leq 0.0025$) The rest of the conditions had no significant difference from each other. Practice (Block number or its interaction with conditions) had no significant effect.

The fact that contact activation buttons had higher amount of error corrections than force activation buttons and yet the two activations mechanisms had very similar CPS speed reflect a type of speed accuracy trade off between them. Participants post test interviews shed some lights on this:

> *"I like capacitive because it's more responsive…but often it's so responsive that I pressed the wrong button."*

> *"I had to use my fingernails on 4 (Force-narrow)."*

> *"I often tried to use fingernail in narrow button."*

These self reports illustrate the inability to take advantage of the sharp edge of the fingernail with the contact activated narrow buttons. No participants used the advanced feature of iPhone that allowed them to roll the finger on buttons for correct selection. Due to the nature of information processing, feedback based adjustment is relatively slow in comparison to one step tapping.

Subjective ratings

Subjectively, the participants rated the conditions in this experiment lower than those in Experiment 2. The mean ratings changed from moderately positive in Experiment 2 (0.35 and .0.53) to neutral or near neutral (~0) in the wide conditions to negative (~ -1) in the narrow conditions.

	Count	Mean	Std. Dev.	Std. Err.
Contact, Wide	13	.154	1.725	.478
Contact, Narrow	13	-.846	1.519	.421
Force, Wide	13	0.000	1.414	.392
Force, Narrow	13	-1.077	1.656	.459

Table 5. Subjectively ratings in Experiment

Summary

As we hypothesized, the performance of finger operated touch screen soft buttons deteriorated when the size of the button fells below a fraction of the finger width. The width of the wide buttons in this experiment was 7.2mm, which is about 40% of average finger width of men and 47% of women [9]. The mean speed was around 1.9 CPS. This was much lower than the mean speed measured in Experiment 2, where the size of the buttons was 16.5mm x 10.5mm and speed was around 2.6 CPS. Even with the wide buttons in this experiment the participants had more error correction strokes (7% to 9%) than in Experiment 2 (around 1 to 2%) and moved their average ratings from positive to neutral.

Furthermore, in the narrow (but taller) button condition with each button measured 4.9 mm x 8.3 mm, the mean speed further decreased to about 1.7 CPS and mean subjective rating became negative. In the contact-narrow condition the mean number of correction strokes went up to 25%.

CONCLUSIONS, DISCUSSION, AND FUTURE WORK

The advent of a new generation of touch screen devices presents users, researchers, and future product designers many basic usability questions such as the performance of soft buttons in comparison to hard buttons. The three experiments presented in this paper contribute to a body of empirical knowledge and understanding of soft button performance in this era. The results suggest that for tasks such as dialing phone numbers and using calculator functions, soft buttons augmented with synthetic feedback can offer a level of performance that is similar, and in *some* cases even superior, to the level of performance offered by typical hard buttons embedded in today's handheld devices (Experiment 1 and 2). Either audio or vibrato-tactile feedback improves soft button performance, but no further improvement is made when both are combined (Experiment 1). They therefore can be alternatively offered depending on the context of use. Button size affects performance, particularly when buttons are smaller than 10 mm in width (Experiment 3). On the other hand the current study shows that soft buttons still worked quite well when the button size was less than the average finger width or the 22 mm recommendation in the previous literature (Experiment 1 and 2). Styli can more accurately handle smaller buttons and they depend less on synthetic feedback than fingers do (Experiment 1), but they can be lost easily and require an acquisition step that bare fingers do not. The two types of touch sensors explored, capacitive and resistive, afford very different behavior but only subtle performance difference (Experiment 2 and 3). The former, activated by contact, can be operated by fingers (only) with very sensitive response, but is more error prone when the soft buttons are below 5 mm in width whereas the latter, activated by a force threshold, may be less responsive but can afford fingernail or stylus operation when the buttons are very small (Experiment 3). Automatic error correction for text entry, found in for example the iPhone, might be necessary when the keys in a soft keyboard are small and operated with finger tip, although there are both pros and cons to automatic error correction [12].

Clearly the current study is limited in both scope and depth, but it may provide a baseline for many future studies. One direction is to expand the task and context scope. For example, the current study was conducted with undivided attention of the user which can be true in many real world situations but not in an extremely mobile context such as dialing a phone number or sending a short text message while walking or driving (which should not be recommended in any case). The results presented here may change when the user's attention is divided. More advanced location sensitive feedback in either the audio or the tactile modality is also a direction for exploration. Another expansion in scope to the current study is to increase the amount of experience of the participants with each condition. Undoubtedly their performance would increase with any of the soft button conditions as they acquire more skills specific to each type of soft buttons, but it is unknown how the improvement rate would differ between conditions. Finally, yet another promising direction for tackling small touch screen size is to use stroke gestures and pattern recognition such as those in [13, 29] which can be much more error tolerant than relying on precise operation of small soft buttons.

ACKNOWLEDGMENT

This work was conducted when Seungyon Claire Lee was a graduate intern at the IBM Almaden Research Center in the summer of 2007. We thank Bill Buxton for a discussion that partly instigated this work. His suggestion of a phone number dialing task and other comments are also gratefully acknowledged. We thank John Barton and other colleagues at IBM Almaden for their invaluable advice and support.

REFERENCES

1. Albinsson, P.-A. and Zhai, S., High Precision Touch Screen Interaction. *Proc. CHI 2003, ACM Conference on Human Factors in Computing Systems, CHI Letters 5(1)* (*2003*), ACM, 105-112.
2. Barrett, J. and Krueger, H. Performance effects of reduced proprioceptive feedback on touch typists and casual users in a typing task. *Behavior & Information technology*, *13*, (1994). 373-381.
3. Buxton, W. Chapter 5, Case Study 1:Touch Tablets. in W.Buxton (ed). *Human Input to Computer Systems: Theories, Techniques and Technology. Unfinished book manuscript.*,

http://www.billbuxton.com/input05.TouchTablets.pdf (in progress, accessed Sept 2007).

4. Buxton, W. Lexical and Pragmatic considerations of input structures. *Computer Graphics*, *17*, 1 (1983). 31-37.
5. Buxton, W., Hill, R. and Rowley, P. Issues and Techniques in Touch-Sensitive Tablet Input. *Computer graphics, Proceedings of SIGGRAPH '85*, *19*, (1985). 215-224.
6. Card, S., Mackinlay, J.D. and Robertson, G.G., The design space of input devices. *Proc. CHI'90: ACM Conference on Human Factors in Computing Systems* (*1990*).
7. Fitzmaurice, G.W., Ishii, H. and Buxton, W., Bricks: Laying the foundation for graspable user interfaces. *Proc. CHI'95: ACM Conference on Human Factors in Computing Systems* (*1995*), 442-449.
8. Greenstein, J.S. Pointing Devices. in M. Helander, T.L., P. Prabhu (ed). *Handbook of Human-Computer Interaction*, North-Holland, Amsterdam. (1997), 1317-1348.
9. HenryDreyfusAssociates *The Measure of Man and Woman*. Whitney Library of Design, New York, NY, 1993. (1993).
10. Hoffmann, E.R. and Sheikh, I.H. Finger width corrections in Fitts' law: implications for speed-accuracy research. *Journal of Motor Behavior*, *23*, 4 (1991). 258-262.
11. Ishii, H. and Ullmer, B., Tangible bits: towards seamless interfaces between people, bits and atoms. *Proc. CHI'97* (*1997*).
12. Kristensson, P.O. and Zhai, S., Relaxing stylus typing precision by geometric pattern matching. *Proc. ACM International Conference on Intelligent User Interfaces (IUI '05)* (*2005*), ACM, 151-158.
13. Kristensson, P.-O. and Zhai, S., SHARK2: A Large Vocabulary Shorthand Writing System for Pen-based Computers. *Proc. ACM Symposium on User Interface Software and Technology* (*2004*), 43 - 52.
14. Lewis, J.R. Literature Review of Touch-Screen Research from1980 to 1992, IBM, Boca Raton, FL (1993).
15. Lewis, J.R., Kennedy, P.J. and LaLomia, M.J. Improved typing-key layouts for single-finger or stylus input, IBM Technical Report TR 54.692 (1992).
16. Lewis, J.R., Potosnak, K.M. and Magyar, R.L. Keys and Keyboards. in Helander, M.G., Landauer, T.K. and Prabhu, P.V. (ed). *Handbook of human-computer interaction*, Elsevier Science, Amsterdam (1997), 1285-1315.
17. Potter, R.L., Weldon, L. J., and Shneiderman, B. (1988). Improving the accuracy of touch screens: An experimental evaluation of three strategies. *Proc. ACM CHI Conference on Human Factors in Computing Systems* (*1988*), 27-32.
18. Poupyrev, I. and Maruyama, S., Tactile interfaces for small touch screens. *Proc. ACM Symposium on User Interface Software and Technology* (*2003*), 217–220.
19. Sears, A. Improving touchscreen keyboards: Design issues and a comparison with other devices. *Interacting with Computers*, *3*, (1991). 253-269.
20. Sears, A., Revis, D., Swatski, J., Crittenden, R. and Shneiderman, B. Investigating touchscreen typing: the effect of keyboard size on typing speed. *Behaviour & Information Technology*, *2*, 1 (1993). 17-22.
21. Sears, A. and Shneiderman, B. High Precision Touchscreens: Design Strategies and Comparison with a Mouse. *International Journal of Man-Machine Studies*, *43*, 4 (1991). 593-613.
22. Soukoreff, R.W. and MacKenzie, I.S. Towards a standard for pointing device evaluation: Perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies*, *61*, (2004). 751-789.
23. Soukoreff, W. and MacKenzie, I.S. Theoretical upper and lower bounds on typing speeds using a stylus and keyboard,. *Behaviour & Information Technology*, *14*, (1995). 379-379.
24. Ullmer, B., Ishii, H. and Jacob, R.J.K. Token+Constraint Systems for Tangible Interaction with Digital Information. *ACM Transactions on Computer-Human Interaction*, *12*, 1. 81-118.
25. Zhai, S. Characterizing computer input with Fitts' law parameters -- The information and non-information aspects of pointing. *International Journal of Human-Computer Studies*Special Issue of Fitts (1954) 50th Anniversary (2004).
26. Zhai, S. User Performance in Relation to 3D Input Device Design, Computer Graphics. *Computer Graphics*, *32*, 4 (1998). 50-54.
27. Zhai, S., Hunter, M. and Smith, B.A. Performance optimization of virtual keyboards. *Human-Computer Interaction*, *17*, 2,3 (2002). 89-129.
28. Zhai, S., Kong, J. and Ren, X. Speed-accuracy trade-off in Fitts' law tasks - on the equivalency of actual and nominal pointing precision. *International Journal of Human-Computer Studies*, *61*, 6 - Special Issue of Fitts (1954) 50th Anniversary (2004). 823-856.
29. Zhai, S. and Kristensson, P.-O., Shorthand Writing on Stylus Keyboard. *Proc. CHI 2003, CHI Letters 5(1)* (*2003*), ACM, 97-104.