No-look Flick: Single-handed and Eyes-free Japanese Text Input System on Touch Screens of Mobile Devices

Yoshitomo Fukatsu, Buntarou Shizuki, and Jiro Tanaka
University of Tsukuba, Japan
{fukatsu, shizuki, jiro}@iplab.cs.tsukuba.ac.jp

ABSTRACT

We present a single-handed and eyes-free Japanese *kana* text input system on touch screens of mobile devices. We firs conducted preliminary experiments to investigate the accuracy with which subjects could single-handedly point to and flic without using their eyes. We found from the results that users can point at a screen that was divided into 2 × 2 with 100% accuracy and that users can flic at a 2 × 2 grid without using their eyes with 96.1% accuracy using our algorithm for flic recognition. The system used kana letter input based on two-stroke input with three keys to enable accurate eyes-free typing. First, users flic for consonant input, and then similarly flic for vowel input. We conducted a long-term user study to measure basic text entry speed and error rate performance under eyes-free conditions, and readability of transcribed phrases. As a result, the mean text entry speed was 51.2 characters per minute (cpm) in the 10th session of the user study and the mean error rate was 0.6% of all characters. The mean text entry speed was 33.9 cpm in the 11th session, which was conducted under totally eyes-free conditions and the mean error rate was 4.8% of all characters. We not only measured cpm and error rate, but also measured error rate of reading, which we devised as a novel metric to measure how accurately users can read transcribed phrases. The mean error rate of reading in the 11th session was 5.7% of all phrases.

Author Keywords

Text entry; touch typing; multi-tap; heads-up writing; touchscreen phones; one-handed interaction; pointing; shoulder surfing

ACM Classificatio Keywords

H.5.2. Information Interfaces and Presentation: User Interfaces - Evaluation/methodology - Input devices and strategies

INTRODUCTION

Characters are input on touch screens of mobile devices by using software keyboards. Eyes-free input with touch screens is difficul because of two issues resulting from touch screen properties. First, the lack of haptic feedback requires users to be visually attentive [27, 31]. Second, it is difficul for them to tap keys accurately because of the "fat finge problem" [26], which causes false input.

Karlson et al. [14] stated that the vast majority of users want single-handed interaction with mobile devices, e.g., when the other hand is preoccupied. Moreover, Yi et al. [32] demonstrated that there are some situations where users do want to use their mobile devices while continuing to talk with others, whereas such overt use of mobile devices is socially inappropriate (e.g., in replying to incoming messages).

We built a single-handed and eyes-free Japanese *kana* text input system for touch screens on mobile devices to explore the above issues. Our system, called No-look Flick, is specifi cally designed to accomplish two main purposes:

Taking personal notes in social situations
Our system allows users to take notes without interrupting anyone talking at meetings or in classes by enabling users to input texts under desks because the system only requires a single hand for text input and can be used eyes-free. In addition, users can take notes by directing their visual attention forward in situations where users need to look forward (e.g., when walking or waiting at stoplights).

Protection from eavesdropping
As users of our system can input texts with the display turned off, they can prevent input texts from eavesdropping by others in crowded environments (e.g., on terribly overcrowded trains in Japan). In addition, they can type single handedly even if the other hand is occupied (e.g., hanging onto a strap on trains).

Our main purpose involved two unique challenges in designing the text input system:

- The design needed to avoid visual or audible feedback, since these would have hindered use in some social situations. However, some subtle tactile feedback could be used.
- The design needed to enable users to input texts in social situations in which protection from eavesdropping and later readability take precedence over text entry speed.

This paper presents a single-handed and eyes-free Japanese *kana* text input system on touch screens of mobile devices. We conducted two preliminary experiments as the firs step in designing such system to investigate the accuracy with which

Table 1. Japanese syllabary. (‘K*’ can be transcribed into combination of consonant ‘K’ and symbol ‘*’, and can be transcribed as ‘G’ in a phonetic alphabet. The same applies to other voiced letters and p-sound letters.)

Vowels	Consonants																		
	Basic letter											Voiced letter				P-sound	Small letter		
	A	K	S	T	N	H	M	Y	R	W	-	K*(G)	S*(Z)	T*(D)	H*(B)	H**(P)	A˜	Y˜	T˜
a	あ	か	さ	た	な	は	ま	や	ら	わ		が	ざ	だ	ば	ぱ	ぁ	ゃ	
i	い	き	し	ち	に	ひ	み		り			ぎ	じ	ぢ	び	ぴ	ぃ		
u	う	く	す	つ	ぬ	ふ	む	ゆ	る			ぐ	ず	づ	ぶ	ぷ	ぅ	ゅ	っ
e	え	け	せ	て	ね	へ	め		れ			げ	ぜ	で	べ	ぺ	ぇ		
o	お	こ	そ	と	の	ほ	も	よ	ろ	を		ご	ぞ	ど	ぼ	ぽ	ぉ	ょ	
-											ん								
-											ー								

users pointed at and flic ed eyes-free on touch screens of mobile devices. We designed No-look Flick based on the results of these experiments. We then implemented this system as an iOS application that worked on iPhone 4. We conducted a long-term user study to measure basic performance as a text input system and performance under totally eyes-free conditions.

The three main finding of this research were:

- Users could accurately point to screens of mobile devices, which are generally used, in eyes-free under conditions where the screens were divided into a 2 × 2 grid.
- They could flic with 96.1% accuracy on the 2 × 2 grid by using our flic recognition algorithm.
- They could input texts single-handed and eyes-free on touch screens of mobile devices to take reasonably accurate personal notes.

JAPANESE WRITING SYSTEM

This section briefl describes the Japanese writing system and the basis for Japanese text input as prior knowledge for the following discussion.

Formal Japanese text consists of *kana* and *kanji* where kana are phonograms and kanji are ideograms derived from Chinese characters. Each kanji character’s phonetic value can be written with one or more kana letters. First, users input words in kana by utilizing this fact in most Japanese text input systems. Then, the input system displays possible candidates that consists of kana and kanji, each of which has the phonetic value of the input kana letters. The main reason there are *candidates* is that some kanji share the same phonetic value. Users choose a candidate to input the required words that consist of kana and kanji characters (*kana-kanji* conversion).

Users of our system only input kana because they cannot choose kanji candidates in eyes-free. In addition, taking personal notes with phrases written only in kana is acceptable and in fact common in daily life because such phrases make sufficien sense and writing kanji characters is too time-consuming; one kanji character requires 12.2 strokes on average while one kana letter requires 2.8 strokes on average.

Kana letters can be transcribed into one to three alphabetical characters (*mean* = *1.8*) based on the Japanese syllabary in Table 1, which is taught in elementary schools throughout Japan. Most basic kana letters (“Basic letters” in Table 1) can be transcribed into a consonant and a vowel (e.g., “く” can be transcribed into ‘K’ and ‘u’). Some kana letters (“Voiced letters”, “P-sounds”, and “Small letters” in Table 1) have corresponding voiced letters, p-sound, and small letters (called “special letters” after this) by adding a symbol (e.g., ‘*’ to voiced letters). That is, a special letter can be transcribed into a consonant, a vowel, and a symbol (e.g., “ぐ” can be transcribed into ‘K’, ‘u’, and ‘*’).

RELATED WORK

Our No-look Flick is based on the following three areas of prior work.

Japanese Kana Input

Since No-look Flick is a Japanese kana text input system, we will describe three conventional methods of inputting Japanese kana into mobile devices.

QWERTY

Japanese speakers usually use QWERTY keyboards to input kana into desktop computers. However, few people use QWERTY keyboards on mobile devices because it is too difficul for them to touch such tiny keys. Keyboards that have a small number of keys have been proposed [28] to resolve this diffi culty, which are described below.

Multi-tap Input

Mobile phones with no touch screens use tactile keyboards and adopt a method of multi-tap input. Figure 1 outlines a typical key layout for such keyboards. This key layout consists of 12 keys that are arranged in a 3 × 4 grid. Users input a kana letter by pressing a key one to f ve times (multi-tap input). For example, they input “せ” by pressing the ‘3’ key four times. They inputs a special letter by using two keys (a numeric key and a ‘*’ key). For example, users input “ぜ” by pressing the ‘3’ key four times, and then pressing the ‘*’ key once.

A key corresponds to a consonant, and the number of presses corresponds to a vowel. Therefore, this method of input is easy to learn. However, it suffers in operation, since it requires one to six taps to input a kana letter (*mean* = *3.3*).

Pocket Bell Input

Pagers called *pocket bells* spread as personal communications systems before mobile phones became widely used in Japan. The method of input for pocket bells is called *pocket bell input*, which needs numeric keys to be pressed twice to input a letter (Table 2). For example, users input “せ” by pressing the ‘3’ and ‘4’ keys. They input a special letter by inputting keys twice. For example, users input “ぜ” by firs pressing the ‘3’ and ‘4’ keys, and then by pressing the ‘0’ and ‘4’ keys.

Figure 1. Key layout of tactile keyboard for multi-tap input.

First input	Second input									
	1	2	3	4	5	6	7	8	9	0
1	あ	い	う	え	お	A	B	C	D	E
2	か	き	く	け	こ	F	G	H	I	J
3	さ	し	す	せ	そ	K	L	M	N	O
4	た	ち	つ	て	と	P	Q	R	S	T
5	な	に	ぬ	ね	の	U	V	W	X	Y
6	は	ひ	ふ	へ	ほ	Z	?	!	-	/
7	ま	み	む	め	も	¥	&			
8	や	(	ゆ	)	よ	*	#			
9	ら	り	る	れ	ろ	1	2	3	4	5
0	わ	を	ん	゛	゜	6	7	8	9	0

Table 2. Character correspondence table for pocket bell input.

Pocket bell input is also available in many old mobile phones. However, due to the difficult of memorizing the character correspondence table (Table 2), few people use this method of input [28]. In contrast, users can easily learn our method of input because No-look Flick is similar to the Grid Flick described below.

Grid Flick

Grid Flick, whose original form was available on Apple Newton and that was adopted into Apple iPhone in 2008, is one of the de facto methods of standard input into mobile devices with touch screens [9], which has been installed by default on Apple iPhone as well as Android for input of Japanese text. Grid Flick has a 3 × 4 key layout, which is similar to the 3 × 4 key layout of old mobile phones (Figure 2). Grid Flick adopts four-directional flic input, taking advantage of touch screen properties in that gestural inputs are possible. Users can make f ve kinds of input per key by tapping it or by flickin from its center to one of four directions (left, upward, right, and downward). Users basically input a kana letter by tapping or flickin a key once. For example, they input "ふ" by flickin the 'は' key upward. They input a special letter by two flic or tap operations. For example, they input "ぶ" by flickin the 'は' key upward, and then tapping the '*' key once.

A key corresponds to a consonant and the tapping or flickin direction corresponds to a vowel. Therefore, this method of input is also suitable for kana. The Grid Flick's main advantage over multi-tap and pocket bell inputs is that users can input kana letters with fewer strokes. Therefore, if they get used to the method of input, they can type faster with flic input than with the other two methods.

We adopted Grid Flick's method of input (i.e., flic input) in No-look Flick. In addition, we redesigned the key layout and method of input for eyes-free interaction.

Eyes-free Japanese Text Input System

Some researchers have proposed eyes-free Japanese text input system. Ikawa et al. [11] proposed a method of "Direction Only" Flick Input, which had only one key (it was the same as the screen of a mobile device) and adopted eight-directional flic input. They reported typing speeds in cpm and error rates for two existing methods (Grid Flick and Google Handwrite [8]) and two proposed methods (TT-Flick and MS-Flick). Participants input sentences using Grid Flick at 45.8 cpm with an error rate of 48.0%, Google Handwrite at 26.2 cpm with an error rate of 23.6%, TT-Flick at 21.7 cpm with an error rate of 13.1%, and MS-Flick at 19.5 cpm with an error rate of 12.6% after 15 minutes of practice. Drag & Flick [1] is an eyes-free Japanese text input system for people with visual impairments, which has only one key (just like Ikawa et al.'s [11]). Users input one kana letter through a series of drags and flick with audio feedback. Three participants with visual impairments input letters (50 kinds of letters excluding special letters) at approximately 40 cpm in a user study. No error rates were reported.

Figure 2. Key layout of Grid Flick. (a) When pressing 'は' key, letters that can be input by tapping or flickin 'は' key are provided as feedback, and then, (b) by flickin 'は' key upward, (c) 'ふ ' is input.

Accessibility Technology

Many researchers have proposed eyes-free text input systems for people with visual impairments. Mobile Messenger for the Blind [21] is a messaging system on mobile devices that has nine software keys and each key has three to four letters. Users input letters through multi-tapping with text-to-speech provided as feedback. No-look Notes [4] is a text input system on mobile devices that uses multi-touch input and audio feedback. No-look Notes significantl outperformed VoiceOver [29], which Apple offers to enable keyboard access to the visually impaired, in terms of speed, accuracy, and user preferences. BrailleTouch [6] is a text entry system with six software keys that represent braille with audio feedback. Similarly, TypeInBraille [18] and Perkinput [2] have adopted braille techniques. These braille based system are efficien for users who are familiar with braille because their key layout is easy to memorize. However, these systems for the visually impaired rely on audio feedback. In contrast, our system only uses subtle tactile feedback to support people with situational impairments [25], e.g., those who want to take personal notes in social situations.

Some researchers have proposed text input systems for people with situational impairments. PocketTouch [22] enables eyes-free multi-touch input with a capacitive touchscreen on the back of a smartphone that detects finge -strokes through fabric, allowing users to input alphanumerics without taking the device out of their pocket. However, PocketTouch requires auxiliary hardware. In contrast, our system applies existing mobile devices without the need for any additional equipment, enabling users to use our system on their own devices by only installing an application. Jain et al. [13] proposed a bezel-based text input system that could be used ambidextrously without looking at a screen. Participants in their study achieved a text entry speed of 9.2 words per minute in situations that required minimal visual attention focused on the screen. While this system was evaluated with two-handed use, in contrast, we evaluated our system with single-handed use. Banovic et al. [3] proposed Escape-Keyboard where users typed letters eyes-free and single-handed by pressing their thumb on different areas of the screen and making flic gestures. They reported a user study that include 16 sessions under eyes-free and non-eyes-free conditions.

PRELIMINARY EXPERIMENT 1: POINTING ACCURACY UNDER EYES-FREE CONDITIONS

We conducted a preliminary experiment to investigate pointing accuracy on the touch screens of mobile devices under eyes-free conditions to obtain clues to system design.

Participants

Ten participants (nine males and one female) ranging in age from 21 to 24 took part in the experiment as volunteers. Their experience using mobile devices with touch screens ranged from 0 to 84 months ($mean = 21.9$, $SD = 23.3$).

Apparatus

We used a laptop computer (Apple MacBook Pro that had a 13-inch screen) and a mobile device with a touch screen (Apple iPhone 4S that had a 3.5 inch screen without either a screen sheet or device case).

Procedure

We placed the laptop computer on a desk. We asked the participants to sit on a chair and hold the mobile device with one hand (Figure 3). All the participants held the mobile device with their right hand since they were all right-handed. We mirrored the screen of the mobile device onto the laptop using Reflectio [1]. We also asked the participants to hold the mobile device under the desk and not look at its screen, but look at the screen on the laptop.

The experiment started when a participant pointed to (i.e., tapped on) any position on the mobile device's screen, and a gray rectangle (called "target" after this) was then presented on the mobile device (and on the laptop by mirroring). When the participant saw the target on the laptop, he or she pointed to the corresponding position on the screen of the mobile device. Regardless of the success or failure of pointing, the next target was presented when he or she pointed at the screen. A beep was played when the screen was pointed at to encourage the participant to begin the next trial.

We divided the screen into a 2×2, 3×3, 4×4, and 5×5 grid (screen conditions), and presented a target in a grid (Figure 4) in each trial. One target was presented four times in random order (e.g., when the screen condition was 2×2, targets were presented 4 times $\times$ 4 areas = 16 times). All screen conditions were presented in random order. As a result, the target was presented 216 times (4 times $\times$ (4+9+16+25) areas) per participant. Participant took about 10 minutes to complete this experiment. We recorded the positions they tapped for each trial.

[1]Reflectio `http://www.reflectionapp.com/`

Figure 3. Experimental setup for preliminary experiment 1.

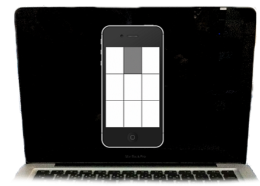

Figure 4. Example of target presented on laptop when screen condition was 3 × 3.

Results and Analysis

We measured the pointing accuracy, calculated as (the number of successful pointed positions) $\div$ (the number of trials) $\times$ 100 (%). The pointing accuracy of all screen conditions is shown in Table 3. This indicates that participants could point to a grid accurately in eyes-free when the screen condition was 2×2.

We analyzed the distribution of pointed positions. Figure 5 shows all the positions participants pointed at under the 2×2 screen condition. The blue points represent the pointed positions, the centers of the gray ellipses represent the centroids of pointed positions for each target, and the radii of gray ellipses represent the standard deviations. Note that the centroids of pointed positions tend to deviate from the centers of the targets to the bottom, and that the centroids in the targets on the right tend to deviate from the centers of the targets to the right. Also note that the rightmost points of the blue points in the targets on the left are located near the vertical boundary and the leftmost points of the blue points in the targets on the right are located at the far right from the vertical boundary.

PRELIMINARY EXPERIMENT 2: VARIATION IN FLICKING IN EYES-FREE

We conducted another preliminary experiment to investigate the behavior of flic input in eyes-free to further understand the properties of eyes-free interaction on touch screens of mobile devices. We focused on the 2×2 screen condition in this experiment, since the results from preliminary experiment 1 revealed that the 2×2 screen condition had better potential to accomplish accurate eyes-free interaction.

Screen condition	Pointing accuracy
2×2	100.0%
3×3	83.1%
4×4	57.5%
5×5	48.5%

Table 3. Pointing accuracy under each screen condition.

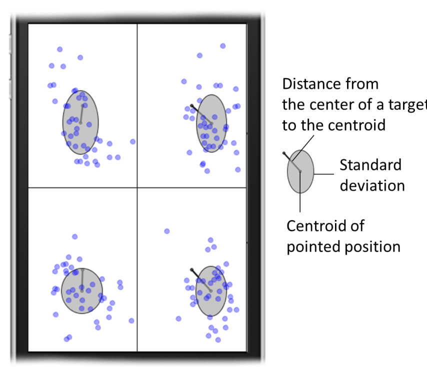

Figure 5. Distribution of pointed positions.

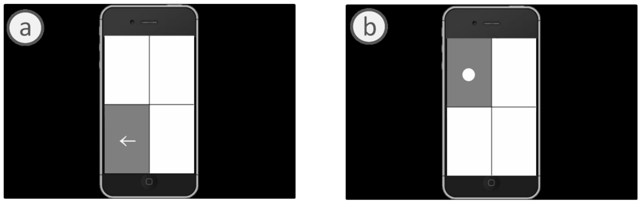

Figure 6. Example targets: (a) Target to instruct flickin to the left in bottom left area of screen, (b) Target to instruct tapping in top left area of screen.

Participants

The same participants as those in preliminary experiment 1 took part in the second experiment as volunteers. Eight participants used Grid Flick as Japanese text input method on a daily basis. Two participants had never used Grid Flick before. Their experience using Grid Flick ranged from 0 to 36 months ($mean = 15.5$, $SD = 11.7$).

Apparatus

We used the same apparatus (a laptop computer and a mobile device) as that in preliminary experiment 1.

Procedure

The experiment started when a participant tapped on any position of the mobile device's screen, and a target with a gray background was presented on the mobile device (and on the laptop by mirroring) as seen in Figure 6. The target was a white arrow or a white circle.

A white arrow against a gray background (Figure 6a) instructed the participant to flic toward the direction of the arrow in the gray area, and a white circle against a gray background (Figure 6b) instructed him or her to tap on the gray area. When the participant saw the target on the laptop, he or she tapped or flic ed according to the target's instructions. Regardless of the success or failure of tapping or flicking the next target was presented when he or she tapped on the screen. A beep was played when the screen was tapped on to encourage the participant to begin the next trial.

Participants carried out tasks under three posture conditions:

Sitting posture (Figure 7a)
The participant sat on a chair and carried out the task holding the mobile device under the desk.

Standing posture (Figure 7b)
The participant stood in front of the laptop and carried out the task with his right hand holding the mobile device near his waist.

Walking posture (Figure 7c)
The participant input phrases while walking behind a researcher. We designed this posture condition by referring to [7, 19].

One target was presented four times under the three posture conditions in random order (i.e., 4 times $\times$ 4 areas $\times$ 5 kinds of flick $= 60$ times). As a result, the targets were presented 180 times (60 times $\times$ 3 posture conditions $= 180$ times). Participants took about 10 minutes to complete this experiment. We recorded the start and end positions of the flic gestures for all trials.

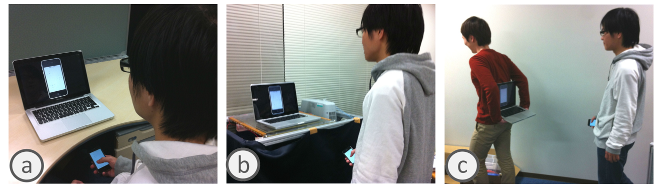

Figure 7. (a) Sitting posture, (b) standing posture, and (c) walking posture.

Results and Analysis

We analyzed the variations in flicking Figure 8 shows the centroids of the start positions for each kind of flicking We numbered the kinds of flic input from 1 to 20 as annotated in Figure 8. We found that the start positions varied depending on the kinds of flicking For example, the start positions for flickin to the right deviated from the center of each area to the left while the start positions for flickin to the left deviated from the center of each area to the right. Variations in the flickin start positions were observed for the three posture conditions in the same way.

We measured the accuracy of flicking calculated as (the number of successful flicks $\div$ (the number of total flicks $\times$ 100 (%). We evaluated flickin as having been successful in these measurements if it was the same as what the target instructed. Figure 9 shows the accuracy of flickin under three posture conditions. One-way repeated measures ANOVA found significan differences between the three posture conditions ($F_{2,18} = 4.591$, $p = .024 < .05$). To analyze these further, we conducted post-hoc analysis with Bonferroni correction under the three posture conditions. Accuracy under the sitting condition in pairwise comparison was significantl higher than that under the walking condition ($p = .007 < .01$).

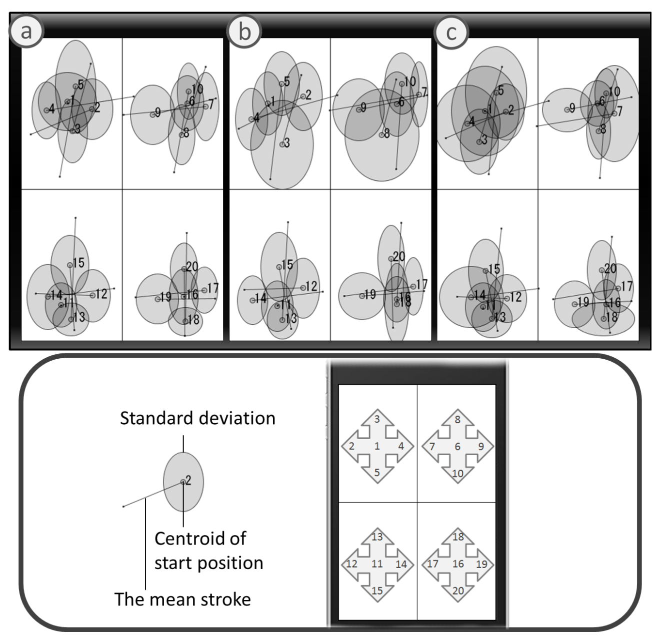

Figure 8. Variation in flickin (Annotated numbers represent kinds of flic inputs that were presented by target), (a) in sitting posture, (b) in standing posture, and (c) in walking posture.

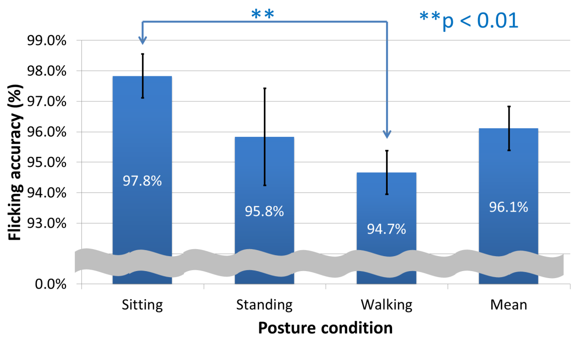

Figure 9. Accuracy of flickin under three posture conditions.

This was consistent with previous studies [24, 19] that also found reduced accuracy while walking.

Overall, the grand mean flickin accuracy was 96.1%. Therefore, users could flic with 96.1% accuracy in the 2 × 2 grid in eyes-free using our algorithm.

Our conclusion from preliminary experiment 2 was that the method of flic input with our flic recognition algorithm had the potential to accomplish accurate eyes-free interaction with the 2 × 2 key layout.

NO-LOOK FLICK

We designed No-look Flick, which is a Japanese kana text input system, based on the insights from the preliminary experiments.

Key Layout and Input Method

Figure 10 shows the key layout for No-look Flick. We located two consonant keys and one vowel key on the touch screen.

As described in the previous section, most kana letters can be transcribed into a consonant and a vowel, and a special letter can be transcribed into a consonant, a vowel, and a symbol. Therefore, we adopted two-stroke input in most cases, and three-stroke input for special letters.

Users input one kana letter in two strokes: the firs is a flic to input the consonant of a kana letter and the second is a flic to input its vowel. If they input consonants twice or more in succession, the last consonant is adopted. If they input a vowel before inputting a consonant, no kana letters are input. When they finis inputting these two strokes, one kana letter is input, and the vowel key changes to the key for a special letter. At this time, users can change the kana letter to its voiced letter, its p-sound letter, or its small letter by inputting the key for a special letter. That is, users can input a voiced letter, a p-sound letter, or a small letter in three strokes. They can backspace by swiping the screen from the right to the left edge.

Whenever a kana letter is input (i.e., after a vowel is input, after a kana letter has changed to its special letter), the user is given tactile feedback through vibration. Similarly, the same tactile feedback is given whenever backspacing occurs.
Figure 11 is a state transition diagram for accepting a kana letter, where “special input” means the input of a symbol for a voiced letter, a p-sound letter, or a small letter. Vowel input

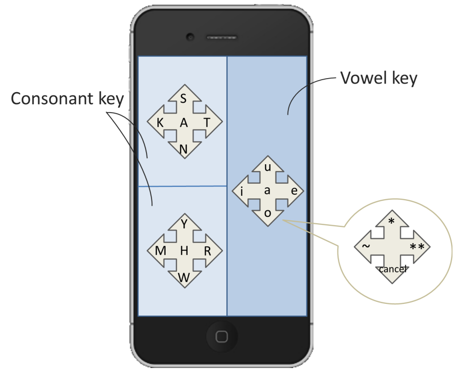

Figure 10. Key layout. (These keys were not displayed on screen on mobile device.)

1 means the input of a kana letter that cannot be changed to a voiced letter, a p-sound letter, or a small letter. Vowel input 2 means the input of a kana letter that can be changed to special letters. Specificall , Vowel input 1 includes letters in the ‘N’ , ‘M’, ‘R’, and ‘W’ rows. Vowel input 2 includes letters in the ‘A’, ‘K’, ‘S’, ‘T’, ‘H’, and ‘Y’ rows.

Figure 12a shows the input of “も”, which has been transcribed into consonant ‘M’ and vowel ‘o’. It is input in two strokes. Figure 12b shows the input of “じ”, which has been transcribed into consonant ’S’ , vowel ‘i’, and symbol ‘*’. It is input in three strokes.

Design Principles

To realize single-handed and eyes-free input with accuracy, we adopted the following design principles:

Flick gesture

We adopted location-dependent flic gestures into No-look Flick. We found that users could make flic gestures with 96.1% accuracy (equal to 3.9% error rate) in the 2 x 2 key layout from the results of preliminary experiment 2. That means if two flick are required to input a letter, error would occur approximately 8% of the time. Although 8% sounds like a high error rate for a text entry system, we feel that it is reasonable because no-look is a relatively new paradigm, and because it is a low error rate given that a user’s eyes are free for other tasks.

Three-key layout

We adopted a three-key layout (Figure 10), although the results of preliminary experiment 2 implied the use of 2 × 2

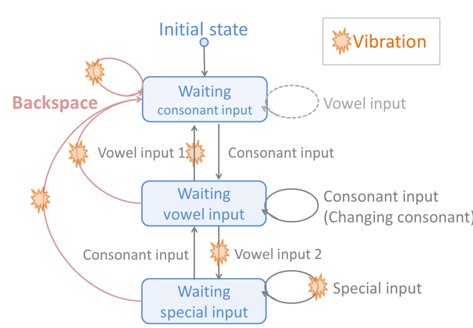

Figure 11. State transition diagram for accepting kana letter.

key layout. While the layouts were different, they were sufficientl similar that we believe the three-key layout would perform similarly, if not a little better, because the single button would be larger.

Our three-key layout fit the properties of the Japanese language. Concretely, Japanese has 10 basic consonants and f ve basic vowels. 50 basic letters (in Table 1) are transcribed into combinations of a consonant and a vowel. Users of our three-key layout can input 10 kinds of consonants with two keys on the left of the screen (2 keys $\times$ (4 flick $+$ 1 tap)) and input 5 kinds of vowels with one key on right-side of the screen (1 key $\times$ (4 flick $+$ 1 tap)). Therefore, this layout fit the properties of the Japanese language, making it easy for those fluen in Japanese to learn our system.

Separation of consonant input and vowel input

We separated consonant keys from a vowel key. This design allows users to retype a consonant as many times as they want before finishin the input of a kana letter. If it were not for this retyping function, they had to delete the consonant when they fin that they mistyped it. However, as the deletion of vowels can be confused with the deletion of letters, this made it difficul for users to be aware of the number of letters that had been deleted under conditions where they could not see what they had typed. Therefore, this separation is effective in avoiding such confusion.

Near-edge interaction

We designed a gesture for backspacing to start and end at the edge of the screen, similar to Bezel Swipe [20], whose starting position for operation was located at the edge of the screen. This design was based on the fact that near-edge interaction can be operated accurately even in eyes-free [13, 5]. In addition, it prevented conflic with the flic operation for inputting kana letters.

Tactile feedback

Whenever a letter was input or deleted, the device vibrated, enabling users to understand that a letter had been input or deleted. More specificall , tactile feedback by vibration is provided with the timing given in the state transition diagram in Figure 11. No vibration is provided when the firs stroke was input, which only inputs a consonant. When the second stroke, third stroke, or backspace were input, same vibration that lasts 0.4 seconds is provided.

EVALUATION

We implemented the prototype of No-look Flick as an iOS application in Objective-C that was operated on iPhone 4S

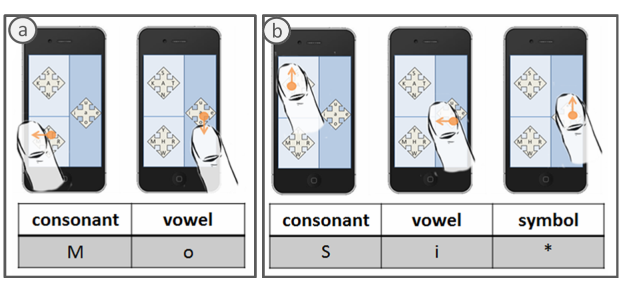

Figure 12. (a) Example of inputting kana letter in two strokes ("も" in this case), (b) example of inputting kana letter in three strokes ("じ" in this case).

(iOS 6.0), and we evaluated No-look Flick by using this prototype. No-look Flick was evaluated in three parts:

Part I

Part I was a longitudinal study designed in accordance with conventional studies on text input systems [23, 17, 30, 12] to measure the basic performance of No-look Flick. This part consisted of 10 sessions.

Part II

Part II was a study under totally eyes-free conditions with one session to measure performance under more realistic conditions. This part consisted of one session, and was carried out after participants had become used to No-look Flick through Part I.

Part III

Part III was a test to read transcribed phrases in Part II to measure the readability of the text input with No-look Flick. Participants were informed of this part more than 48 hours after they completed Part II.

Participants

Six participants (f ve males and one female) ranging in age from 21 to 23 took part in the experiment and received an incentive for their participation. They were all right-handed. Their experience using smartphones ranged from 5 to 36 months (*mean* $=$ 18.8, *SD* $=$ 10.8). All the participants were constant users of Grid Flick and their experience with it ranged from 5 to 36 months (*mean* $=$ 18.8, *SD* $=$ 10.8). None took part in the preliminary experiments or had experienced No-look Flick previously.

Short Phrases

We prepared 750 short phrases that consisted of six to eight characters based on two criteria:

- Exhaustively including letters that users can input with this system.
- Short phrases that conceivably could be input in the assumed scenario.

One example is "れぽーとしめきり", which means "deadline for the report." This phrase is a reminder that means a report needs to be finishe as the deadline is drawing nearer.

The Kullback-Leibler divergence of the distribution of our phrase set from that of a Japanese language dictionary [15] was .07684.

Procedure

Participants were instructed about the method of input for No-look Flick and the goal of our study. The participant input some phrases for practice that contained all the kana letters available in No-look Flick. After the explanation and practice, the three parts of the evaluation were conducted according to the following procedures:

Part I

The 10 sessions of Part I were scheduled with one or two sessions a day, allowing zero to two intervening days between sessions. Each session was divided into 12 blocks with f ve phrases per block (i.e., 60 phrases per session; a total of 600 phrases). Participants could freely take breaks between blocks. We used 600 phrases from the phrase set in this part. Participants never input the same phrase twice. Each session lasted 20-35 minutes.

We used the same laptop computer and mobile device as we had in the preliminary experiments. Participants sat down on a chair and held a mobile device in their right hand. The experiment started when they touched any position on the screen of the mobile device, and they transcribed the presented phrase while watching the screen of laptop computer (Figure 13). After they had transcribed the phrase, they input "Enter", and then the next phrase was presented. "Enter" could be input with consonant "W" and then vowel 'o' to ensure consistency in the method of input. We asked participants to transcribe phrases as quickly and accurately as possible, and to correct errors they noticed by backspacing. The text entry speed and the number of errors were presented on a pop-up window to motivate the participants every f ve transcriptions.

Part II

Participants participated in Part II after Part I, which consisted of one session divided into three blocks with 25 phrases per block (i.e., a total of 75 phrases). They carried out tasks under the same three posture conditions (i.e., sitting, standing, and walking) as those in preliminary experiment 2. We only presented phrases to be input, which was different from Part I. Thus, participants could not know what they actually input. Since participants could not determine whether the input letters were correct or not without looking, we asked them to transcribe phrases by focusing on accuracy rather than speed. In this part, we used the remaining 150 phrases from the phrase set, i.e., 75 phrases for P1-P3 (Group A) and another 75 phrases for P4-P6 (Group B). Participants never input the same phrase twice. This session lasted 25-30 minutes.

Part III

Participants participated in Part III more than 48 hours after they had completed Part II. They read their own transcribed 75 phrases and another participant's transcribed 75 phrases. More specificall , a participant in Group A read the phrases transcribed by a participant in Group B, and vice versa. We asked participants to guess the originally presented phrases. We indicated deleted letters in gray, and "E" represented "Enter" (Figure 14) in the phrases that participants read. Because we observed high error rates in Part II (see below), we presented not just the characters entered, but also the characters deleted in Part III (Figure 14). Although this is unusual for existing text entry systems, we think it is a likely adaptation for applications using no-look input in the future.

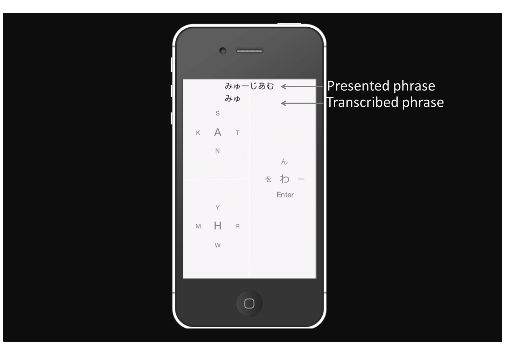

Figure 13. Screenshot in evaluation Part I.

あなうんあさーE
えいせほうそうE

Figure 14. Transcribed phrases with deleted letters in gray.

Metrics

We used the following metrics to evaluate the performance of our method in Part I and Part II:

Speed

We calculated the text entry speed in characters per minute (cpm) by the number of transcribed characters over the time it took to transcribe phrases. More specificall , the time began when participants touched the screen to enter the consonant of the firs letter and ended when they released their finge from the screen after inputting "Enter". Because the typing speeds of Japanese text input systems are usually evaluated in cpm, we calculated cpm as the typing speed.

Error rate

We classifie errors into substitutions (incorrect letters), omissions (omitted letters), or insertions (added letters). The error rate over characters is calculated by (the sum of substitutions, omissions, and insertions) $\div$ (the number of the letters in the presented phrases).

In part III, in addition to cpm and error rate over characters, we measured the error rate of reading over phrases, which we devised as a novel metric to measure the readability of transcribed phrases. The error rate of reading was calculated by (the number of phrases that were incorrectly understood) $\div$ (the number of presented phrases) (%). The high error rate of reading means that transcribed phrases had too many errors for their meanings to be understood.

RESULTS AND DISCUSSION

Part I - Basic Performance

The mean text entry speed in Session 1 at the beginning of Part I was 27.2 cpm ($SD = 4.9$) and the mean speed in Session 10 at the end of Part I was 51.2 cpm ($SD = 7.7$) with an increase of 88.3% (Figure 15). The black bold line in Figure 15 illustrates linear regression ($R^2 = .6147$).

We compared the results to those from related Japanese text entry systems. Previous studies have reported that beginners can type kana letters at 21.0 cpm using QWERTY on mobile devices with touch screens [10], and at 22.2 cpm using Grid Flick [16] (they were not carried out under eyes-free conditions). The entry speed for beginners of No-look Flick in comparison with these text entry systems was higher (27.2 cpm in Session 1) even under eyes-free conditions. A previous study [10] has also reported that a regular user of Grid Flick recorded 75.0 cpm by using Grid Flick to type kana. In our experiment, P3 recorded 60.4 cpm in Session 10; this speed was lower than Grid Flick by only 19.4% even under eyes-free conditions. These results imply that in terms of text

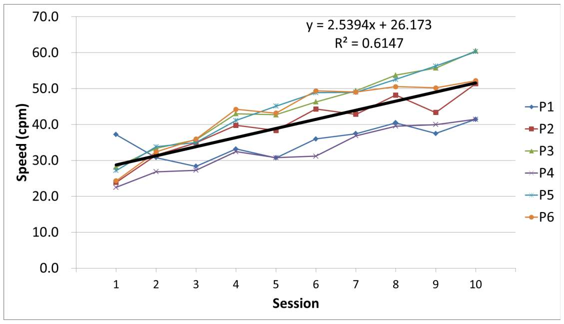

Figure 15. Text entry speeds for participants in Part I.

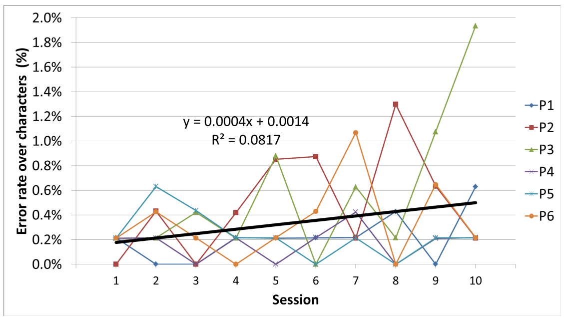

Figure 16. Error rates over characters for participants in Part I.

entry speed, No-look Flick recorded sufficientl high performance for a practical Japanese kana input method.

The mean error rate over 10 sessions was 0.3% of all characters (Figure 16). All the mean error rates across all participants in each session were lower than 1%. A slight increase in error rates over sessions was observed ($R^2 = .0817$). This trend predominantly appeared in the results for P3 in the last two sessions (sessions 9 and 10). P3 recorded 60.4 cpm (the highest typing speed in this study) and 1.9% of error rate (the highest error rate in this study) in session 10. On the other hand, P5 recorded a higher typing speed (56.3 cpm in session 9 and 60.2 cpm in session 10) and a lower error rate (0.2% in sessions 9 and 10) than average. Although the results differed between participants, some of them might have become tired or focused on speed rather than accuracy over the sessions.

Part II - Performance under Totally Eyes-free Conditions

The mean text entry speed in Part II was 33.9 cpm ($SD = 6.5$), which was 33.8% lower than the speed in session 10 of Part I (51.2 cpm). Since we asked participants to transcribe phrases by focusing on accuracy rather than speed in Part II, the reduced entry speed was unsurprising. One-way repeated measures ANOVA shows no significan difference among posture conditions ($F_{2,10} = .177$, $p = .841$) as shown in Figure 17.

The mean error rate in Part II was 4.8% of all characters. This value means that the transcribed phrases contain less than 1 error out of 20 characters. The results showed No-look Flick is a reasonably accurate eyes-free typing system. Although the error rate under walking condition is higher than other two conditions (sitting, standing), one-way repeated measures ANOVA revealed no significan differences in the three posture conditions ($F_{2,10} = 1.480$, $p = .274$) as seen in Figure 18.

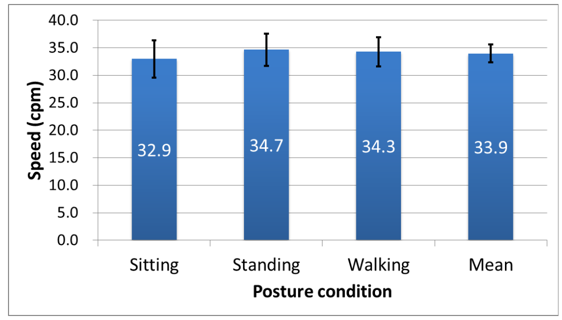

Figure 17. Text entry speeds under three posture conditions in Part II.

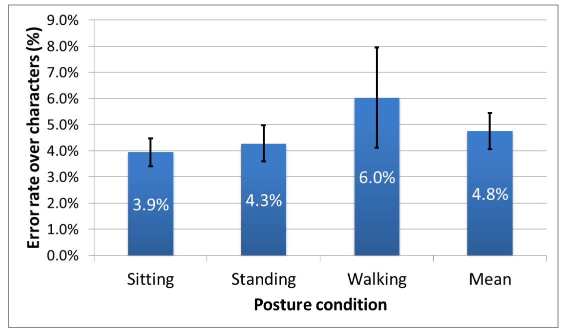

Figure 18. Error rates over characters under three posture conditions in Part II.

Part III - Readability

The grand mean error rate of reading was 5.7% of all phrases. This value means that approximately one phrase out of 18 phrases input with No-look Flick was misread. The mean error rate of reading one's own transcribed phrases was 5.1% of all phrases and the mean error rate of reading other's transcribed phrases was 6.2% of all phrases. A paired t-test revealed no significan differences between the two error rates ($t_5 = 1.000$, $p = .363$). We believe that the error rates were reasonably low to support personal notes (i.e., reading one's own phrases) as well as texting (i.e., reading another's phrases).

We found that participants frequently failed to guess correct phrases from those that had two or more errors (e.g., mistyping "し**ら**し**ふ**し**る**う" instead of "しゃしんしゅう") by examining the errors in detail. In contrast, although phrases had errors in some cases, participants guessed the phrases correctly. For example, a participant mistyped "こく**の**い**そ**ん" instead of "こくないさん" ("Ko Ku N*o* Ai S*o* Wu" instead of "Ko Ku Na Ai Sa Wu" in the alphabet). That is, although there were two incorrect letters, the participant could guess the phrase correctly. This is because the two letters only had an incorrect vowel. In addition, some participants commented "I could guess the correct phrase from the consonant or the vowel." when they had finishe Part III. This means that our design principle of "separation of consonant input and vowel input" was not only successful in avoiding confusion in deletion of a vowel with the deletion of a letter, but also in improving the readability of phrases input with our system. This also raises the possibility of enabling computers to predict correct phrases from user's phrases containing incorrect letters. Analyses of pointed positions and flic tra-jectories would also similarly help this.

Limitation

In our current prototype of No-look Flick, users can input only kana letters because we have not implemented kana-kanji conversion, which generally requires visual attention focused on the screens of devices. Although the current prototype without kana-kanji conversion can be used to take personal notes as the result of the evaluation Part III showed, we think that allowing users to convert kana to kanji later is reasonable in some cases. Although this would mean not being purely eyes-free, occasional visual attention for kana-kanji conversion might be acceptable for users.

CONCLUSIONS AND FUTURE WORK

We have presented No-look Flick, the Japanese text input system for eyes-free typing on screens of mobile devices to take personal notes. We implemented this system as an iOS application, and evaluated its basic performance (text entry speed and error rate over characters), its performance under totally eyes-free conditions, and the readability of transcribed phrases. As a result, we observed the mean entry speed of 33.9 cpm (54.8% slower than the speed of a regular user of Grid Flick under non eyes-free condition) and the error rate of 4.8% of all characters in the 11th session under a totally eyes-free condition, and observed the error rate of 5.7% of all phrases regarding the readability of transcribed phrases. Although our approach leveraged properties of the Japanese language whose kana letters have a vowel and a consonant, our design principles for eyes-free text input will be applicable for other languages, especially for languages which have vowels and consonants.

In future work, we intend to implement iOS applications that No-look Flick is applied to, which will allow users to tweet or text eyes-free.

REFERENCES

1. Aoki, R., Hashimoto, R., Seko, S., Kataoka, Y., Ihara, M., Watanabe, M., and Kobayashi, T. Drag&flick A text entry method on a touch screen for visually impaired people. In Interaction2013, IPSJ (2013), 72–79. (in Japanese).
2. Azenkot, S., Wobbrock, J. O., Prasain, S., and Ladner, R. E. Input finge detection for nonvisual touch screen text entry in perkinput. In GI '12, CIPS (2012), 121–129.
3. Banovic, N., Yatani, K., and Truong, K. N. Escape-keyboard: A sight-free one-handed text entry method for mobile touch-screen devices. IJMHCI 5, 3 (2013).
4. Bonner, M. N., Brudvik, J. T., Abowd, G. D., and Edwards, W. K. No-look notes: accessible eyes-free multi-touch text entry. In Pervasive '10, Springer-Verlag (2010), 409–426.
5. Bragdon, A., Nelson, E., Li, Y., and Hinckley, K. Experimental analysis of touch-screen gesture designs in mobile environments. In CHI '11, ACM (2011), 403–412.
6. Frey, B., Southern, C., and Romero, M. Brailletouch: mobile texting for the visually impaired. In UAHCI'11, Springer-Verlag (2011), 19–25.
7. Goel, M., Findlater, L., and Wobbrock, J. Walktype: using accelerometer data to accomodate situational impairments in mobile touch screen text entry. In CHI '12, ACM (2012), 2687–2696.
8. Google handwrite (in japanese). `http://www.google.com/insidesearch/features/search/handwritinginput/mobile.html`.
9. Grid flic (in Japanese). `http://ja.wikipedia.org/wiki/フリック入力`.
10. Hasegawa, A., Hasegawa, S., and Miyao, M. An evaluation of input system on the touch panel phone. In Symposium on Mobile Interactions, Society of Mobile Interactions (2010), 73–76. (in Japanese).
11. Ikawa, Y., and Miyashita, H. *"Direction Only" Flick-Input Method for Shorthand with Eyes-Free*. Bachelor thesis, Meiji University, 2013. (in Japanese).
12. Isokoski, P., and Raisamo, R. Device independent text input: a rationale and an example. In AVI '00, ACM (2000), 76–83.
13. Jain, M., and Balakrishnan, R. User learning and performance with bezel menus. In CHI '12, ACM (2012), 2221–2230.
14. Karlson, A. K., and Bederson, B. B. Understanding single-handed mobile device interaction. Tech. rep., Dep. of Computer Science, Univ. of Maryland, 2006.
15. Katagiri, H., and Sezaki, N. Statistics of japanese syllabary. Tech. rep., National, 1968. (in Japanese).
16. Kato, T., Mannaka, Y., and Hasegawa, A. Operation skill and the speed of character input by flic method of mobile terminals. In Symposium on Mobile Interactions, Society of Mobile Interactions (2010), 137–140. (in Japanese).
17. MacKenzie, I. S., Soukoreff, R. W., and Helga, J. 1 thumb, 4 buttons, 20 words per minute: design and evaluation of h4-writer. In UIST '11, ACM (2011), 471–480.
18. Mascetti, S., Bernareggi, C., and Belotti, M. Typeinbraille: quick eyes-free typing on smartphones. In ICCHP'12, Springer-Verlag (2012), 615–622.
19. Nicolau, H., and Jorge, J. Touch typing using thumbs: understanding the effect of mobility and hand posture. In CHI '12, ACM (2012), 2683–2686.
20. Roth, V., and Turner, T. Bezel swipe: conflict-fre scrolling and multiple selection on mobile touch screen devices. In CHI '09, ACM (2009), 1523–1526.
21. Sánchez, J., and Aguayo, F. Mobile messenger for the blind. In ERCIM'06, Springer-Verlag (2007), 369–385.
22. Saponas, T. S., Harrison, C., and Benko, H. Pockettouch: through-fabric capacitive touch input. In UIST '11, ACM (2011), 303–308.
23. Sato, D., Shizuki, B., Miura, M., and Tanaka, J. Menu-selection-based japanese input method with consonants for pen-based computers. In APCHI '04, Springer-Verlag (2004), 399–408.
24. Schildbach, B., and Rukzio, E. Investigating selection and reading performance on a mobile phone while walking. In MobileHCI '10, ACM (2010), 93–102.
25. Sears, A., Lin, M., Jacko, J., and Xiao, Y. When computers fade...pervasive computing and situationally-induced impairments and disabilities. In HCII 2003, 1298–1302.
26. Siek, K. A., Rogers, Y., and Connelly, K. H. Fat finge worries: how older and younger users physically interact with pdas. In INTERACT'05, Springer-Verlag (2005), 267–280.
27. Silfverberg, M. Using mobile keypads with limited visual feedback: Implications to handheld and wearable devices. In MobileHCI '03, Springer-Verlag (2003), 76–90.
28. Masui, T., Okamoto, A., Itoh, K., and Uchida, H. Interfaces for text input. *Journal of Human Interface Society 4*, 3 (2009), 131–144. (in Japanese).
29. Apple - accessibility - iphone - voiceover. `http://www.apple.com/accessibility/voiceover/`.
30. Wobbrock, J., Myers, B., and Rothrock, B. Few-key text entry revisited: mnemonic gestures on four keys. In CHI '06, ACM (2006), 489–492.
31. Yatani, K., and Truong, K. N. Semfeel: a user interface with semantic tactile feedback for mobile touch-screen devices. In UIST '09, ACM (2009), 111–120.
32. Yi, B., Cao, X., Fjeld, M., and Zhao, S. Exploring user motivations for eyes-free interaction on mobile devices. In CHI '12, ACM (2012), 2789–2792.