Modeling Learning Effects in Mobile Texting

Arindam Das, Wolfgang Stuerzlinger
Department of Computer Science and Engineering,
York University, Toronto, Ontario, Canada
{arindam,wolfgang} @ cse.yorku.ca

ABSTRACT
No work on mobile text messaging so far has taken into account the effect of learning on the change in visual exploration behavior as users progress from non-expert to expert level. We discuss within the domain of multi-tap texting on mobile phone and address the process of searching versus selecting a letter on the keypad interface. We develop a simulation model that forecasts the probability of letter location recall by non-expert users and thereby models learning, as the user acquires expertise in recalling, with practice, session after session. We then plug-in this probability within a model of visual strategy that combines the effect of different ways visual exploration: non-expert users *search* for a letter while expert users *select* a letter. The observed non-expert non-motor time preceding a key press (for a letter) correlates extremely well with the simulation results.

Categories and Subject Descriptors
H.5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/methodology.

General Terms
Human Factors.

Keywords
Texting, learning, visual search, non-expert/novice user, mobile phone

1. INTRODUCTION
While the empirical work evaluating mobile keypad layout for text entry is strong, it has either concentrated solely on expert behavior, see e.g. the survey [12], or solely on non-expert behavior, see e.g. [14], [6]. In essence, all previous work ignores a user's gradual skill development from novice to expert level.

MUM'2008, December 3-5, 2008, Umeå, Sweden.

This is not surprising, as this is a difficult transition to analyze. First, finding pure novice users who never used texting before is getting harder and harder in many countries. Most previous studies were performed in geographical regions where a sizeable subset of the subjects had prior experience in texting. Second, performing experiments on learning in text entry is difficult, as monitoring the transition between novice and expert adequately requires longitudinal evaluation over a prolonged period of time (e.g. the experiment in [13] took 20 sessions of 45 minutes each to compare learning in two soft keyboards). The logistics of scheduling such an experiment for several users as well as the boredom factor (e.g. [3] found that even a 30 minute session could be "tedious and frustrating" for a typical subject) make this next to impossible. Last, but not least, reimbursing participants for their time becomes expensive for such long-running studies.

One solution for this is to develop cognitive models that predict user performance transitioning from novice to expert level in texting on mobile phones. Towards this end, [5] developed a first model, although it had a few limitations: (a) it is a coarse ACT-R model at the *symbolic* level, which advocates instantaneous availability of knowledge. This ignores all the neural effort that takes place in acquiring that knowledge. (b) it completely ignores the stochastic nature of human behavior. (c) it did not account for the time that is usually spent by novice users in visually scanning the frontal surface in order to identify a letter location on the phone keypad.

In this paper, we present a new simulation model that addresses the limitations of [5]. We make the following assumptions about the process: When the user is a pure novice with respect to a given keypad layout, she performs an explicit *visual search* to find each letter. However, as she gains expertise with practice over time, she gradually starts recalling letter locations and spends less time in visual search. With this increase in prior knowledge of letter locations, she now begins to spend more time *deciding* which letter location to select out of the remembered ones. Overall, the objective of the new simulation model is to be able to take into account this gradual transition from a *searching* process to a *selection* (i.e. decision) process. We simulate this transition by predicting the amount of time that *precedes* the key press for a letter. We refer to this time as *non-Fitts time,* which implies that this is the fraction of the user's total task performance time after subtracting the movement time predicted by Fitts law [7]. In previous work, the non-Fitts time for novice users was measured in an empirical study in [14]. Similar to that study, our model does not provide for error checking or error correction behavior. Finally, we apply our model to predict the curve of simulated user's text entry speed and show how her entry speed changes with the change in her accuracy of letter location recall.

2. BACKGROUND

2.1 Multi-tap Text Entry on Mobile Phone

[6] were the first to investigate multi-tap text entry on cell phones. Shortly after that, Silfverberg et *al*. performed an extensive empirical study on multi-tap text entry [17]. This was followed by other studies [11] and [3].

While [6] and [17] concentrated on expert users, the following studies [11] and [3] analyzed both novice and expert users. These studies also point out that [17] is an overly optimistic predictive model, as it focuses solely on the motor part. It effectively ignores any potential cognitive component (which is non-zero even for expert behavior as discussed in [11]). [14] demonstrated the existence of this cognitive component in novice users and contributed a model for novice user behavior in text entry.

Our work is motivated by the fact that no work on mobile text messaging so far has considered the impact of learning. In other words, we focus on the change in visual exploration behavior as well as the effects of increasing recall accuracy as users progress from novice to expert level.

2.2 Using the ACT-R Cognitive Architecture

The ACT-R cognitive architecture [2] is a discrete-event simulation framework that embodies current theories of cognition, perception and motor behavior. It enables the creation of models that simulate the temporal (and often simultaneous) progression of cognitive processes, such as attentional changes and memory retrieval, as well as motor processes, such as movement of fingers. It has been used as a basis for several models in Human-Computer-Interaction [15]. An ACT-R based simulation model can be "run" at two levels: *symbolic* and *sub-symbolic*. The symbolic level in ACT-R is an abstract characterization of how brain structures encode knowledge [1]. At this level, the natural randomness inherent in human behavior is not simulated. The sub-symbolic level is an abstract characterization of the role of neural computation in making the aforementioned knowledge available to different cognitive processes [1]. This level simulates the randomness in human behavior.

3. USER STRATEGIES: *Visual search* versus *Selection*

The vision module in ACT-R is an attentional system - it does not account for time spent in visual exploration (visual search) in the field-of-view. This is a shortcoming of ACT-R, which affects our work, as users at the novice level perform a *visual search* to find the key for a given letter. On the other hand, expert behavior, involves *selecting* (i.e. deciding, not searching) a key by using stored knowledge.

In previous work, [16] observed experimentally that placing multiple letters on each key of a keypad results in a longer visual search time per letter compared to a single key for each letter. The same authors also observed a noteworthy shortcoming of a previous work on this issue: The formula to compute choice-reaction time for a key on the keypad in [18] is correctly based on Hick-Hyman law [8, 10]. However, [18] then mistakenly utilized the Hick-Hyman Law to calculate *visual search* time instead of *selection* (i.e. decision) time. To further clarify this distinction, [16] mentions that novice users, by definition, do not know the mapping between letters and keys for a given keypad layout. Therefore, novices cannot *select* (i.e. decide) between alternative responses, which is required for usage of the Hick-Hyman law.

4. A NEW COGNITIVE MODEL FOR MOBILE TEXTING

In the subsections below, we present the pieces of our new cognitive simulation model. We first describe the task that we consider in this work. Then we describe how we adapt ACT-R and discuss how we model the average visual *search* and *selection* (i.e. decision) time for a letter. We then utilize these two quantities in an equation that predicts an important part of the non-Fitts time of text entry on cell phones.

4.1 The task

In this work, we focus on the task of performing a key press on the cell phone. More precisely, we focus on a simulation model that predicts the non-Fitts time preceding the key press for a letter in a letter-group containing multiple distinct letters.

4.2 Modifications to ACT-R

We extend the ACT-R 6.0 framework for our work as follows: The motor module of the standard ACT-R framework contains only a model of a QWERTY computer keyboard. To develop our model of text entry on cell phones, we added a model for the keypad interface of a Nokia 5190 handset to the motor module of ACT-R. This is the same handset used in previous work [14]. For this, we created a virtual grid containing all key locations as well as a start and recoil position for the right thumb. See Figure 1, where columns 0-2 contain the keys, and column 3 contains the start and the recoil positions. Although the recoil position may vary and hence affect the movement time predicted by Fitts' law, this work focuses only on the non-Fitts component of user's total task performance time and is hence unaffected by the exact location of the recoil position. We further make the following simplifying assumptions: a) all the keys on the keypad are of the same size, b) the width of a key is one "key unit", c) the horizontal and vertical distance between adjacent keys on the keypad is one "key unit", and d) that the (right-handed) user holds the handset in the right hand and uses only the thumb to press keys.

We also added several motor movement styles. The first is called thumb-recoil-to-location and models the movement of the right-hand thumb from a key to the recoil location (3, 2) in Figure 1. The second movement style addresses the fact that in multi-tap text entry more than one character is mapped to the same key. The default "peck" movement style of ACT-R (a directed movement of a finger to a new location followed by a keystroke, all as one continuous movement) is only appropriate for keyboards where a single letter is mapped to each key. To adapt this for multi-tap text entry, we extended the ACT-R system to allow the modeler to specify the location of the target key as well as the character the cognitive model would be pecking for. We named the new movement style peck-to-location-for-char.

Finally, we modified the ACT-R motor module to disable the inclusion of Fitts law time in the task performance time during simulation. This allows us to focus only on the cognitive aspect of text entry.

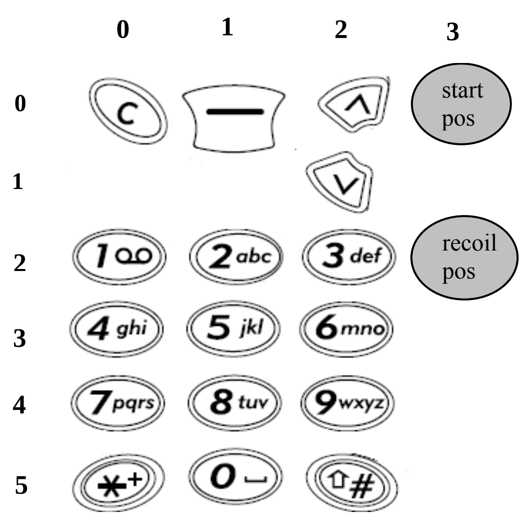

Figure 1. Virtual grid for the Nokia 5190 keypad.

4.3 Visual Search Time for Letter

To model the behavior of a novice user, we predict the visual search time for a letter on the multi-tap keypad as follows.

Let,

NFT_{av}^{L} denote Average Non-Fitts time for letter,
NFT_{av}^{N} denote Average Non-Fitts time for number,
$CogT_{av}^{L}$ denote Average Cognitive Time for letter,
$CogT_{av}^{N}$ denote Average Cognitive Time for number,
$VSRT_{av}^{L}$ denote Average Visual Search Time for letter,
$VSRT_{av}^{N}$ denote Average Visual Search Time for number

Then, for a novice user,

$$NFT_{av}^{L} = CogT_{av}^{L} + VSRT_{av}^{L}$$

where Cognitive Time, $CogT_{av}^{L}$, is the mental processing time preceding the *visual search* for an alpha-numeric character (i.e. letter or number).

Next, we make two assumptions: (i) For a given instance of the task execution (i.e. for a given trial), the difference between the novice cognitive times for any two distinct alpha-numeric characters (i.e. letter or number) on the keypad is zero, and (ii) Visual Search Time for a number on the cell phone keypad (1 to 9 and 0) is zero. We believe that these assumptions are valid due to the following reasons: In case of the first assumption, all the alpha-numeric characters are treated uniformly as symbols; they are not distinguished by their features (i.e. lines, angles and curves that make up an alpha-numeric character). Hence the cognitive times for all alpha-numeric characters are conjectured to be the same. In case of the second assumption, the numbers are thought to be arranged in a standardized layout that is common to all telephones, including landlines. Hence, we assume that users are very familiar with this layout. Furthermore, numbers on the keypad are mapped one per key and the font size of numbers is usually are fairly large compared to letters. Hence a typical user is assumed not to spend any time for searching a number on the cell phone keypad.

We can therefore write,

$$\begin{aligned} VSRT_{av}^{L} &= VSRT_{av}^{L} - VSRT_{av}^{N} \\ &\quad \text{(since } VSRT_{av}^{N} \text{ is assumed to be 0 as per } \textit{assumption} \text{ (ii) } \textit{above}\text{)} \\ &= (NFT_{av}^{L} - CogT_{av}^{L}) - (NFT_{av}^{N} - CogT_{av}^{N}) \\ &= (NFT_{av}^{L} - NFT_{av}^{N}) + (CogT_{av}^{N} - CogT_{av}^{L}) \\ &= NFT_{av}^{L} - NFT_{av}^{N} \\ &\quad \text{(since } CogT_{av}^{N} - CogT_{av}^{L} \text{ is assumed to be 0 as per } \textit{assumption} \text{ (i) } \textit{above}\text{)} \end{aligned}$$

Thus,

$$VSRT_{av}^{L} = NFT_{av}^{L} - NFT_{av}^{N}$$

We use the empirical data from [14] to compute $VSRT_{av}^{L}$ as follows: Using the data from the first 75 trials of the study, the NFT_{av}^{L} is 1679.79ms and NFT_{av}^{N} is 999.20ms. Therefore, using the above equation, we find that $VSRT_{av}^{L} = NFT_{av}^{L} - NFT_{av}^{N} = 1679.79 - 999.20 = 680.59$ ms.

4.4 Selection Time for Letter

In case of an expert user, we use the Hick-Hyman Law to compute the *selection* time (i.e. decision time) for a letter on the keypad. The Hick-Hyman law is:

Selection time = $a + b \log_2 (n)$

where n is the number of already known items to select (i.e. decide) from. The items being already known, ideally there is no searching involved in this case. The coefficients a and b are constants. Since text entry is continuous, there would be no surprises on stimulus arrival and therefore, as suggested by [20], we set the constant, a, to 0. [20] also maintains that the speed of key presses in response to stimulus presentation would range between 5 to 7 bits per second. We assume that the fastest selection (or choice) processing speed would be appropriate for a pure expert user, and therefore we set the constant, b, to 1/7 seconds per bit. [16] had suggested that, the number of alternatives, n, should be based upon the number of keys (i.e. reactions) on the keypad rather than the number of letters (i.e. stimuli). Hence we set $n = 8$ since the traditional cell phone keypad has letters appearing only on eight buttons.

Let,

SLT_{av}^{L} denote Average Selection time for letter

Then,

$SLT_{av}^{L} = b \log_2 (n) = (1/7) * \log_2 8 = 428.57$ ms.

In case of an expert user, the Average Non-Fitts time for letter, NFT_{av}^{L}, could be written as

$$NFT_{av}^{L} = CogT_{av}^{L} + SLT_{av}^{L}$$

where Cognitive Time, $CogT_{av}^{L}$, is the mental processing time preceding the *selection* of a letter.

5. INTEGRATING VISUAL EXPLORATION STRATEGIES

In this section, first we briefly discuss the typical behavior of a novice user during multi-tap text entry in cell phones; then we define few terms that we use in this work; finally we provide a strategy adaptation equation that we use during the simulation.

5.1 Novice Behavior during Multi-tap Text Entry

Let us assume that the task of a novice user is to copy some letters into the cell phone. Let there be three main areas viz. display area, text input area and the keypad area from top to bottom respectively on the frontal surface of the handset. Let each of those letters be pre-displayed one at a time (i.e. one per trial) in the display area so that the user can see the letter clearly before copying it into the text input area. Let us further assume that no typing error is committed during the text entry. Initially, the user holds the handset in her right hand with her right thumb roughly on location (3, 0), the start position in Figure 1. Then she roughly carries out the following actions: (i) She looks in the display area at the letter to be copied. (ii) She tries to recall the position of the letter in the keypad area. (iii) If the recall fails, she does one of two things: (a) If her thumb is on the start position, she visually searches the keypad to find the letter on it and then pecks the key containing the letter with her right thumb. (b) If her thumb is not on start position, we assume that it will be on a key that she pecked last. In that case, she first recoils her thumb to some location in column 3 of Figure 1 (this helps her not to block the keypad area with her thumb so that she can see the keypad area clearly). Next, she visually searches the keypad to find the letter on it. Once she finds the letter, she pecks the key containing that letter with her right thumb. (iv) If the recall succeeds, she does not spend time searching the keypad for the letter; rather, she directly pecks the key (containing that letter) since she already remembers its position on the keypad. (v) The next letter, then, gets pre-displayed in the display area. (vi) She saccades back to the display area to look at the next letter to be copied. She then repeats the above steps starting from (ii) again until all the letters are copied into the input text area.

5.2 Definitions

In the aforementioned sections, we explained two cases: (a) When the user is novice, her Non-Fitts time would be composed of Cognitive time and Visual Search Time. (b) When the user is expert, her Non-Fitts time would be composed of Cognitive time and Selection Time. In order to generalize these two cases, let us consider that the Non-Fitts time is composed of the Cognitive time and *Visual Exploration Time* where Visual Exploration Time is either the Visual Search Time or the Selection Time or a combination of both. As a user gradually transitions from novice to expert level with practice, the proportion of her *visual search* time versus *selection* time at any given instance of task execution is determined by a quantity known as *Recall Accuracy* which is defined as the ratio of the number of successful recalls to the number of recall attempts (per letter position on the keypad).

5.3 Recall Accuracy Computation

During a simulation run, our ACT-R model attempts to recall the position of a letter (to be texted) on the keypad at every trial. The recall attempt either fails or succeeds. We divide all the trials from the run equally into blocks. Let each block constitutes n trials. Let the number of successful recalls per block be x where $x <= n$. Let Ra denote the Recall Accuracy. Then,

$$Ra = x / n$$

Recall Accuracy, thus, ideally varies from 0 corresponding to visual search only by pure novice, to 1 corresponding to selection only by pure expert.

5.4 Visual Strategy Adaptation Equation

During a simulation run, corresponding to a given block of trials, the Average Visual Exploration Time, VET_{av}^{L}, for letter position on the keypad can be computed by interpolating between the Average Visual Search Time ($VSRT_{av}^{L}$) and Average Selection Time (SLT_{av}^{L}) for letter position as follows:

$$VET_{av}^{L} = (1 - Ra) * VSRT_{av}^{L} + Ra * SLT_{av}^{L}$$

where *Ra* is the Recall Accuracy. We term this equation Visual Strategy Adaptation Equation. The equation reflects that with practice, as the user becomes more familiar with the arrangement of the keys on the keypad, she is able to remember more letter positions and hence her *search* time for a letter position decreases towards zero. With the increase in familiarity of keypad layout, she adapts to spend more time in *selecting* (i.e. choosing) a letter position out of all the letter positions she remembers so far and therefore her *selection* time dominates. We adapted this idea from [4] who had applied it for searching menus in graphical user interfaces. For our simulation, we substitute $VSRT_{av}^{L}$ and SLT_{av}^{L} of the equation with the values obtained in the previous section in order to obtain the average visual exploration time (in ms) as follows:

$$VET_{av}^{L} = (1 - Ra) * 680.59 + Ra * 428.57$$

During the simulation run, for every block of simulated trials, we compute the recall accuracy, *Ra*, and then plug-in that value in the above equation to calculate the average visual exploration time, VET_{av}^{L} per block of simulated trials.

6. REAL USER'S NON-FITTS TIME

In order to validate our simulation results, we need real user data to compare against. This section discusses how we obtained the real user data related to novice non-Fitts time and expert non-Fitts time.

6.1 Real Novice Non-Fitts Time

One of the user studies carried out in [14] identified the non-Fitts time for the task considered in our work. Using the non-Fitts time collected from the first 75 user trials of that empirical study, we created 15 blocks of real non-Fitts time where each block is an average of 5 trials (a trial being one instance of task execution), in the temporal order of trials. Table 1 has the real non-Fitts time of novice users for those 15 blocks.

6.2 Real Expert Non-Fitts Time

We use data from two previously reported user studies, [11] and [17], to derive a lower bound on expert user data.

[11] found an overall expert measure of 7.93 words per minute (a word being a group of five letters) without differentiating between the time-out or time-out-kill feature for the multi-tap mode of a cell phone and also without differentiating between the use of either the index finger or the thumb. Using the common assumption of 5 letters per word, the average total task performance time to enter a letter by an expert user would therefore be (60/7.93) /5 = 1513.24 ms.

From [17], we obtain the mean text entry speed of 23.75 words per minute by averaging over 22.5 wpm (time-out, index finger), 20.8 wpm (time-out, thumb), 27.2 wpm (time-out-kill, index finger) and 24.5 wpm (time-out-kill, thumb), all of which were predicted utilizing Fitts law. The reason we take the average of all the expert entry speeds from [17] is because we want to avoid differentiating between various modes (i.e. index finger, thumb, time-out or time-out-kill) at the expert level, thereby staying compatible with [11]. Using the value 23.75 wpm computed above, we derive the average time to enter a letter by an expert user as predicted by Fitts law to be (60/23.75) / 5 = 505.26 ms (assuming 5 letters per word).

We, therefore, derive the expert non-Fitts time for entering a letter to be (Average Total Task Performance Time – Average Fitts Law Time) = (1513.24 − 505.26) = 1007.98 ms.

7. SIMULATED USER'S NON-FITTS TIME

Our ACT-R simulation model uses a single modeler-defined chunk-type. The chunks created from this chunk-type helps the model to keep track of the state of search of a letter on the display area as well as on the keypad, the last letter searched and found, the current letter being searched, the location of the current letter on the keypad and the location of the current letter on the display area. The procedural knowledge of our model is represented using the production rules which are similar to those in [5].

The key rules of the model are as follows:

- *can-recall-letter-location-on-keypad* matches if the keypad coordinates of the current letter (that has just been encoded from the display area) is same as the information present in the retrieval buffer and fails to match if it doesn't. If the match occurs, the model will execute a motor action directly, without any attention shift, to enter the letter.
- *cannot-recall-letter-location-on-keypad* matches if the keypad coordinates of the current letter (that has just been encoded *from* the display area) is not same as the information present in the retrieval buffer (more specifically when the retrieval buffer is empty). If the match occurs, it will lead to the shift of visual attention, to the keypad area, for the current letter.

Similar to [14], our model is constructed in such a way that it avoids repeated key presses required to scroll for a letter. We do this in order to stay compatible with the user study in [14] that we model; however we point out that our model can easily simulate the effect of scrolling.

The model performs text entry using only right thumb. Although this is to stay theoretically compatible with the user study in [14] for novice users, however it must be noted that the motor module of ACT-R 6.0 models the action of every finger of either hands in the same way meaning that the non-Fitts times obtained from them at any given trial will all be the same. We further assume that at expert level, the difference between the non-Fitts times obtained from right or left thumb, right or left index finger is negligible.

We set the ACT-R 6.0 sub-symbolic parameters for the model as follows: enable sub-symbolic computation (*esc*) to *t*, retrieval threshold (*rt*) to -0.70, latency factor (*lf*) to 0.00035, activation noise (*ans*) to 0.05 and base level learning (*bll*) to 0.5. We leave the rest of the parameters at their default values. We estimate the above parameter values following the ACT-R modeling procedure so as to fit the novice non-Fitts time data from simulation to the real non-Fitts time data observed from novice users.

We collected data from 9 simulation runs. Each run consisted of 150 blocks of trials. At the start of execution of every block, 5 distinct letters were randomly chosen out of 26 letters in order to simulate 5 trials per block. Overall, each run involved 21 simulated users * 150 blocks * 5 trials = 15750 simulated trials in total. Each simulation run took around ten minutes on Windows XP Home Edition running on a Pentium 4 CPU, 3.20 GHz and 448 MB RAM.

Table 1 shows the real and simulated non-Fitts time for the 15 blocks when the user is at novice level. The simulated times are the average of the data collected from 9 simulation runs. The 15 blocks of real data constitute the first 75 trials from [14], each block being average of 5 trials, in the temporal order of the trials. Considerable oscillation is evident in real data from blocks 1 to 7 and blocks 13 to 15, as illustrated in figure 2. This is due to the associated large variance in the first 75 real data points observed in [14], possibly owing to the short test. As a direct correlation between the real and the simulated novice data makes no real sense due to the noise, we generated trendlines for these two novice data sets (real and simulated) using MS-Excel. The mean slope of the simulated non-Fitts time data from 9 simulation runs was –5.6581 (*SD*=0.7355) as shown in Figure 2. The slope of the real data being -5.662, differed only by 0.069% from the simulated slope. This small percentage of difference between the trendline slopes clearly indicates that, in general, the moving average (i.e. trend) of the simulated data correlates extremely well with that of the real data.

In Figure 3, we show how the Recall Accuracy drives the gradual shift of visual exploration behavior from full visual search towards full selection. When the recall accuracy is near 0, the visual search dominates whereas when it starts approaching 1, selection begins dominating. It is also apparent from the graph that a combination of search and selection occurs during the transition from novice to expert level which is in compliance with the normal human behavior as discussed in [4].

Table 1: Real and Simulated non-Fitts time for 15 blocks of data when the user is at novice level. The 15 blocks of real data constitute the first 75 trials from [14], each block being average of 5 trials, in the temporal order of the trials. Simulated data is rounded off to three decimal places.

Block No.	Real Novice non-Fitts time (ms)	Simulated Novice non-Fitts time (ms)
1	1748.22	1931.590
2	1617.57	1952.065
3	1890.37	1932.935
4	1810.90	1922.622
5	1591.47	1903.082
6	1607.80	1891.836
7	1620.65	1876.617
8	1691.20	1902.466
9	1628.02	1896.852
10	1650.55	1878.440
11	1687.87	1880.802
12	1616.57	1873.837
13	1623.32	1866.414
14	1798.27	1876.137
15	1614.07	1856.225

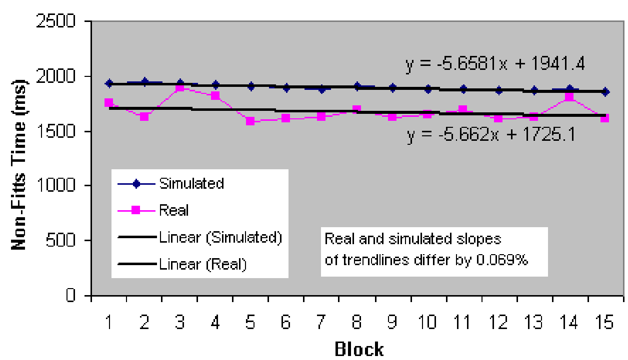

Figure 2. Novice non-Fitts time (simulated vs real) over 15 blocks, 5 trials per block. The slopes of the trendlines for the real and simulated data differ by 0.069%.

8. PREDICTING ENTRY SPEED OF SIMULATED USER

Entry speed is usually considered to be the traditional measure of performance in text entry. In order to predict this measure we would need both the non-Fitts time as well as the Fitts time to enter one letter. In this work, we consider handheld cell phones whose 12-button keypads are fairly small and narrow in size; we consider only 8 keys that are laid out very close to each other on that miniature-sized keypad so that the distance between any two keys is not very large compared to the width of the target key (say, unlike the case of a QWERTY computer keyboard where the distance to width ratios are fairly large). Moreover, the Fitts coefficients for 12-button keypad of a cell phone are also quite small as reported in [17].

The facts mentioned above results in reducing the impact of Fitts law on the key press time on a 12-button cell phone keypad to a very small fraction compared to the non-motor time spent by a user for a key press; this has been discussed in one form or other in several works such as [9], [19], [14], [11] and [3]. On top of all this, it must be noted that we are simulating the observations from those participants who, at least having used the 12-button keypad of the landline phones as part of their day to day living, cannot be considered pure novices in terms of the visual familiarity of the keypad layout.

We, therefore, propose to combine the average value of Fitts law time with the simulated average value of non-Fitts time for every block to arrive at the total task performance time corresponding to that block as follows:

$$TPT_{av}^{L} = NFT_{av}^{L} + FT_{av}^{L}$$

where FT_{av}^{L} denotes the average Fitts law time (in ms) and TPT_{av}^{L} denotes the simulated average total task performance time (in ms) for entering a letter. So far as Fitts time computation in the ACT-R motor module is concerned, we still leave it disabled so as to avoid any conflict during the TPT_{av}^{L} calculation.

In our case, FT_{av}^{L} is 505.26 ms as derived in section "Real User's Non-Fitts Time". Hence for this work, the above TPT_{av}^{L} equation takes the form

$$TPT_{av}^{L} = NFT_{av}^{L} + 505.26$$

For every block of trials, we utilize the TPT_{av}^{L} value (calculated per block) to compute the simulated entry speed in terms of the traditional *words per minute* (WPM) metric:

$$WPM = (1/5) * (1/ TPT_{av}^{L}) * 1000 * 60$$

where the convention of five letters per word is assumed.

Figure 4 shows the curve of simulated entry speed in words per minute (wpm) and recall accuracy, over 150 blocks of trials. The simulated novice speed from our model is found to be about 5 wpm which is close to the novice speed of 5.87 wpm predicted in [14] or 5.59 wpm observed in [11]. Besides, while acknowledging the real expert entry speed of 7.93 words per minute [11], we predict through our model that the simulated expert level speed of 7.926 words per minute (obtained on inspecting the simulation data) would be reached roughly at around 111th block. The predicted expert speed is, thus, fairly close to the real one as well.

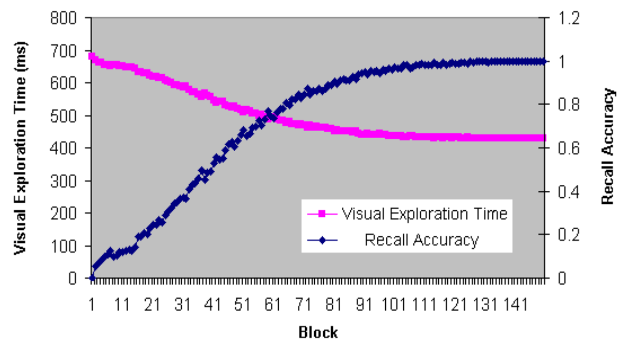

Figure 3: Simulated visual exploration time (in ms) and the recall accuracy (0 to 1) over 150 blocks as the simulated user transitions from novice to expert level.

9. LIMITATIONS OF THE SIMULATION MODEL

There are quite a few limitations of our model: (i) Our work does not model the potential errors that may be committed by entering unexpected characters while texting. (ii) In the visual strategy adaptation equation, we use average values for the visual search time ($VSRT_{av}^{L}$) and selection time (SLT_{av}^{L}) for every block of trials. The value of $VSRT_{av}^{L}$ was derived from the real user data observed on a traditional 12-button multi-tap phone keypad in [14]. This is a layout that a typical phone user is very familiar with. If a different keypad layout is used, it is possible that the $VSRT_{av}^{L}$ may turn out to be different from the value used in this work. We suggest that longitudinal study be undertaken on other cell phone keypad layouts to investigate this possibility.

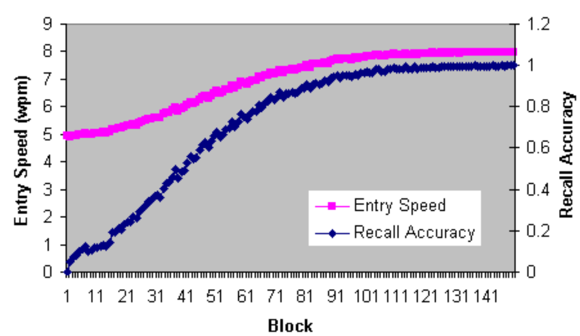

Figure 4: Simulated Entry Speed (in wpm - words per minute) and the recall accuracy (0 to 1) over 150 blocks as the simulated user transitions from novice to expert level. The simulated user reaches the expert level at around 111th block.

10. SUMMARY

In this paper we presented a new ACT-R based cognitive model that simulates the effect of recall accuracy (of letter positions in a cell phone keypad) on user behavior in finding a letter in the keypad during multi-tap text entry. The simulation predicts that as the recall accuracy increases with practice, the user gradually changes her visual exploration strategy from *visual search* to *selection*.

Our cognitive model is important for two reasons. First, from a theoretical standpoint, there has been surprisingly no work in cell phone text entry that successfully considers cognitive time, visual search time, Hick-Hyman selection time and Fitts' time. Our success will hopefully stimulate further work on modeling text entry tasks on other types of mobile devices that include all three – cognitive workload, visual exploration and aimed movement components. Second, nine simulation runs, each equivalent to a longitudinal study, taking only 10 minutes per run as opposed to several weeks, amply proves that such modeling effort will drastically cut down on time.

11. ACKNOWLEDGMENTS

We thank Andriy Pavlovych of York University, Canada.

12. REFERENCES

[1] Anderson, J. R. 2007. *How Can the Human Mind* Occur *in the Physical Universe*? Oxford University Press.

[2] Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Quin, Y. 2004. *An integrated theory of the mind*. Psychological Review, 111(4), 1036–1060.

[3] Butts, L. & Cockburn, A. 2002. *An evaluation of mobile phone text input methods*, Australian Computer Science Communications, Vol 24, Issue 4, 55-59.

[4] Cockburn, A., Gutwin, C., & Greenberg, S. A. 2007. *A Predictive Model of Menu Performance*, CHI 2007, 627-636.

[5] Das, A., *& Stuerzlinger, W.* 2007. *A Cognitive Simulation Model for Novice Text Entry on Cell Phone Keypads*, European Conference on Cognitive Ergonomics, ECCE 2007, 141-147.

[6] Dunlop, M. D., & Crossan, A. 1999. *Dictionary based text entry method for mobile phones*. Second Workshop on Human-Computer Interaction with Mobile Devices, 5-7.

[7] Fitts, P. M. 1954. *The information capacity of the human motor system in controlling the amplitude of movement*. J. Exp. Psychology, 47(6), 381-391.

[8] Hick, W. E. 1952. *On the rate of gain of information*. Quarterly Journal of Experimental Psychology, 4, 11-26.

[9] Holleis, P., Otto, F., Hußmann H., and Schmidt, A. 2007. *Keystroke-Level Model for Advanced Mobile Phone Interaction*. CHI 2007, 1505-1514.

[10] Hyman, R. 1953. *Stimulus information as a determinant of reaction time*. J. Exp. Psychology, 45, 188-196.

[11] James, C. L., and Reischel, K. M. 2001. *Text input for mobile devices: Comparing model predictions to actual performance*. CHI 2001, 365-371.

[12] MacKenzie, I. S., and Soukoreff, R. W. 2002. *Text Entry for Mobile Computing: Models and Methods, Theory and Practice*. Human-Computer Interaction, 17, 147-198.

[13] *MacKenzie*, I. S. and Zhang, S. X. 1999. *The design and evaluation of a high-performance soft keyboard*. CHI 1999, 25-31.

[14] Pavlovych, A., and Stuerzlinger, W. 2004. *Model for non-Expert Text Entry Speed on 12-Button Phone Keypads*. CHI 2004, 351-358.

[15] Ritter, F. E., and Young, R. M. 2001. *Embodied models as simulated users: Introduction to this special issue on using cognitive models to improve interface design*. Int. J. of Human-Computer Studies 55, 1–14.

[16] Sears A., Jacko J. A., Chu J., & Moro F. 2001. *The role of visual search in the design of effective soft keyboards*. Behav. & Info. Tech., Vol. 20, No. 3, 159-166(8).

[17] Silfverberg, M., MacKenzie, I. S., and Korhonen, P. 2000. *Predicting Text Entry Speed on Mobile Phones*, CHI 2000, 9-16.

[18] Soukoreff, W. & MacKenzie, I. S. 1995. *Theoretical upper and lower bounds on typing speed using a stylus and soft keyboard*. Behav. & Info. Tech., 14, 370-379.

[19] St. Amant, R., Horton, T. E., and Ritter, F. E. 2007. *Model-Based Evaluation of Expert Cell Phone Menu Interaction*. ACM Tran. on Computer-Human Interaction, Vol. 14, No. 1, 1–24.

[20] Welford, A. T. 1968. *Fundamentals of Skill*. Methuen, London.