Theoretical Upper and Lower Bounds on Typing Speed Using a Stylus and Soft Keyboard

R. William Soukoreff and I. Scott MacKenzie

Department of Computing and Information Science
University of Guelph
Guelph, Ontario, Canada, N1G 2W1
will@snowhite.cis.uoguelph.ca
mac@snowhite.cis.uoguelph.ca

Abstract
A theoretical model is presented to predict upper- and lower-bound text-entry rates using a stylus to tap on a soft QWERTY keyboard. The model is based on the Hick-Hyman law for choice reaction time, Fitts' law for rapid aimed movements, and linguistic tables for the relative frequencies of letter-pairs, or digrams, in common English. The model's importance lies not only in the predictions provided, but in its characterization of text-entry tasks using keyboards. Whereas previous studies only use frequency probabilities of the 26×26 digrams in the Roman alphabet, our model accommodates the space bar — the most common character in typing tasks. Using a very large linguistic table that decomposes digrams by position-within-words, we established start-of-word (space-letter) and end-of-word (letter-space) probabilities and worked from a 27×27 digram table. The model predicts a typing rate of 8.9 wpm for novices unfamiliar with the QWERTY keyboard, and 30.1 wpm for experts. Comparisons are drawn with empirical studies using a stylus and other forms of text entry.

1. INTRODUCTION

The efficiency of different manual data entry methods has been of interest since the field of information processing began. Entry methods such as typing, printing, hand-writing, or selecting with a mouse have been investigated extensively (e.g., Card, Moran, & Newell, 1983; Van Cott & Kinkade, 1972).

With the advent of pen-based computers, there is renewed interest in printing or hand-writing using a pen or stylus as a form of input (Gibbs, 1993). In fact, the lure of cursive handwritten input is so strong that alternate forms of entry are often ignored. In particular, we feel an entry method worthy of serious evaluation is tapping with a stylus on a graphic representation of a QWERTY keyboard. We call this a "soft keyboard".

In the following sections, we will attempt to answer the following question: "How fast can one type using a stylus to tap on a soft keyboard". We have calculated a lower bound, which is the typing speed expected for walk-up, novice users unfamiliar with the QWERTY layout; and an upper bound, a rate attainable after considerable practice on a QWERTY soft keyboard. Our model is useful, not only through the predictions provided, but also because of its behavioural description of text-entry tasks using various forms of keyboards.

To answer the question posed above, we draw on Fitts' law for rapid aimed movements (Fitts, 1954), the Hick-Hyman law for choice reaction time (Hick, 1952; Hyman, 1953), and linguistics tables for characterizing the text-entry task. Work by Epps (1986), Kerr (1977), MacKenzie, Sellen, and Buxton (1991) and others has demonstrated that Fitts' law is applicable to the movement of a stylus as an input device. By extension, Fitts' law is applicable when stylus movements are between "keys" on a QWERTY keyboard simulated on a pen-based computer's display.

The applicability of the Hick-Hyman law has been established in numerous choice reaction tasks, such as pressing buttons in response to lights. Welford (1968, p. 64) notes that the act of sorting cards can be modeled as two separate acts: a movement time and a visual scan time modeled by the Hick-Hyman law. More recently, Landauer and Nachbar (1985) showed that a computer input task requiring a choice followed by a movement-plus-selection can be modeled using the Hick-Hyman law for the choice component of the task followed by Fitts' law for the movement-plus-selection component of the task.

These simple and intuitive relationships suggest that we can predict novice performance by adding a visual scan time to the movement component of each entry. As the novice gains familiarity with the QWERTY layout, the visual scan time diminishes (eventually to zero) and performance improves to expert levels, whereby the movement time for each entry is the sole component of the task. Although novice-to-expert migration may reveal an improvement in the motor component of the task, this is considered negligible in comparison to the reduction in the visual scan time. Evidence of this has been noted in novice-to-expert transitions for shorthand, wherein improvements follow from reductions in inter-stroke hesitations rather than reductions in stroking time (Gregg, Leslie, & Zoubek, 1972).

To express entry rate in "words per minute", we assume that the task is text entry of common English. Linguistic tables compiled from large samples of representative English assist in characterizing the text-entry task. Briefly, a 26-character alphabet contains 26×26 letter-pairs, or digrams. For each digram, we can predict the movement time to enter the second letter given the first, and this prediction is weighted by the probability of the digram's occurrence. We will develop and refine this concept later, particularly with respect to the role of the space bar (which is of little relevance to linguists).

After deriving our upper-bound and lower-bound models, we will compare our predictions against data from empirical studies using a stylus and other forms of text entry. Weaknesses in our model are identified and assessed; and, in conclusion, qualitative comparisons are drawn between stylus-tapping and other forms of text entry.

2. MODEL DERIVATION

2.1 Characterizing the Text-Entry Task

Before modeling a task, we must define it. We characterize "text entry of common English" using tables for single-letter and letter-pair (digram) frequency counts in English. Our approach is similar to the study of touch typing speeds for different keyboards presented by Card et al. (1983), who used a table by Underwood and Shultz (1960). For reasons explained shortly, we used a different but similar table by Mayzner and Tresselt (1965). Mayzner and Tresselt's (1965) table was compiled from 100 samples of 200 words. Each sample was drawn at random from a variety of sources, including newspapers, magazines, and fiction and non-fiction books. Mayzner and Tresselt established that the relative frequencies of letters in their sample correlated highly with the same frequencies in Underwood and Schulz's (1960) sample. And so, the claim that the sample is representative of English is made, notwithstanding caveats for non-text input, cultural bias, etc.

Mayzner and Tresselt's table, which spans 18 pages, is particularly useful because diagram positions within words were also provided, unlike Underwood and

Schulz's (1960) table. This allowed us to identify digrams at the beginning and ending of words, and therefore, to extend the 26 × 26 diagrams to 27 × 27 digrams. The "alphabet" in our model consists of 26 letters plus the space character. Figure 1 gives the frequency counts of the 27 × 27 diagrams used to characterize the text-entry task for our model. The data in figure 1 were entered into a large spreadsheet with 27 × 27 rows. Columns were added incrementally to build our model, as described herein.

First Letter	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
A	2	144	308	382	1	67	138	9	322	7	146	664	177	1576	1	100	–	802	683	785	87	233	57
B	136	14	–	–	415	–	–	–	78	18	–	98	1	–	240	–	–	88	15	7	256	1	1
C	368	–	13	–	285	–	–	412	67	–	178	108	–	1	298	–	1	71	7	154	34	–	–
D	106	1	–	37	375	3	19	–	148	1	–	22	1	2	137	–	–	83	95	3	52	5	2
E	670	8	181	767	470	103	46	15	127	1	35	332	187	799	44	90	9	1314	630	316	8	172	106
F	145	–	–	–	154	86	–	–	205	–	–	69	3	–	429	–	–	188	4	102	62	–	–
G	94	1	–	–	289	–	19	288	96	–	–	55	1	31	135	–	–	98	42	6	57	–	1
H	1164	–	–	–	3155	–	–	1	824	–	–	5	1	–	487	2	–	91	8	165	75	–	8
I	23	7	304	260	189	56	233	–	1	–	86	324	255	1110	88	42	2	272	484	558	5	165	–
J	2	–	–	–	31	–	–	–	9	–	–	–	–	–	41	–	–	–	–	–	56	–	–
K	2	–	–	–	337	–	–	–	127	–	–	10	1	82	3	1	–	–	50	–	3	–	–
L	332	4	6	289	591	59	7	–	390	–	38	546	30	1	344	34	–	11	121	74	81	17	19
M	394	50	–	–	530	6	–	–	165	–	–	4	28	4	289	77	–	–	53	2	85	–	–
N	100	2	98	1213	512	5	771	5	135	8	63	80	–	54	349	–	3	2	148	378	49	3	2
O	65	67	61	119	34	80	9	1	88	3	123	218	417	598	336	138	–	812	195	415	1115	136	398
P	142	–	1	–	280	1	–	24	97	–	–	169	–	–	149	64	–	110	48	40	68	–	3
Q	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	66	–	–
R	289	10	22	133	1139	13	59	21	309	–	53	71	65	106	504	9	–	69	318	190	89	22	5
S	196	9	47	–	626	–	1	328	214	–	57	48	31	16	213	107	8	–	168	754	175	–	32
T	259	2	31	1	583	1	2	3774	252	–	–	75	1	2	331	–	–	187	209	154	132	–	84
U	45	53	114	48	71	10	148	–	65	–	–	247	87	278	3	49	1	402	299	492	–	–	–
V	27	–	–	–	683	–	–	–	109	–	–	–	–	–	33	–	–	–	–	–	1	–	–
W	595	3	–	6	285	–	–	472	374	–	1	12	–	103	264	–	–	35	21	4	2	–	–
X	17	–	9	–	9	–	–	–	10	–	–	–	–	–	1	22	–	–	–	23	8	–	–
Y	11	10	–	–	152	–	1	1	32	–	–	7	1	–	339	16	–	–	81	2	1	–	2
Z	3	–	–	–	26	–	–	–	2	–	–	4	–	–	2	–	–	–	3	–	–	–	–
Space	1882	1033	864	515	423	1059	453	1388	237	93	152	717	876	478	721	588	42	494	1596	3912	134	116	1787
Total	7069	1418	2059	3770	11645	1549	1906	6739	4483	131	932	3885	2163	5241	5781	1339	66	5129	5278	8536	2701	870	2507

(Second Letter spans columns A–W)

Figure 1. 27 × 27 digrams for the text-entry task. The core 26 x 26 digrams are from Mayzner and Tresselt's (1965) table 2. The space digrams (shaded) were compiled from Mayzner and Tresselt's frequency counts for start-of-word and end-of-word digrams.

Note: A careful inspection of figure 1 reveals slight discrepancies in the row and column totals. These are due to small errors noted by Mayzner and Tresselt (1965, p. 13) and attributed to human error in tabulating the original data.

The importance of the space character is illustrated as follows. Mayzner and Tresselt's (1965) sample contained 20,000 words with a total of 87,199 characters. If figure 1 was constructed only showing 26 × 26 digrams, then the most-prominent letter would be e with a frequency probability of 11,612 / 87,199 = .133 and the most-prominent digram would be h-e with a frequency probability of 3155 / 87,155 = .0362 . However, by adding 20,000 spaces to the sample (one space per word), the frequency probability of e drops to 11,620 / 107,199 = .108 and the space character becomes the most prominent with a frequency probability of 20,000 / 107,199 = .187. The most prominent digram becomes e-space with a frequency probability of 4904 / 107,199 = .0457. Table 1 illustrates the calculations for the ten most-frequent digrams in the sample.

Table 1
Ten most frequent diagrams

Digram	Count	Probability
e–space	4904	.0457
space–t	3912	.0365
t–h	3774	.0352
h–e	3155	.0294
d–space	2627	.0245
s–space	2348	.0219
t–space	2228	.0208
space–a	1882	.0176
space–w	1787	.0167
a–n	1576	.0147

Note: Probability = Count / 107,199

2.2 The Upper Bound

For each of the 27 × 27 digrams in figure 1, we used Fitts' law to predict the movement time to select the second key, given the first. Fitts' law is expressed as

$$MT = a + b \times ID \qquad (1)$$

where *MT* is the predicted movement time, and *ID* is the index of difficulty of the task being modeled. The coefficients *a* and *b* are constants obtained through linear regression. The index of difficulty, in bits, is defined using the Shannon formulation (MacKenzie, 1992) as

$$ID = \log_2\left(\frac{A}{W} + 1\right) \qquad (2)$$

A and *W* are parameters in the task being modeled. Specifically, *A* is the amplitude of the movement and *W* is the width of the target (ideally, measured in the direction of motion).

We are interested in movements from an initial character on the keyboard *i*, to a subsequent character *j*. Combining Equations 1 and 2 results in

$$MT_{ij} = a + b \times \log_2\left(\frac{A_{ij}}{W} + 1\right) \tag{3}$$

where the variable A_{ij}, the amplitude, represents the distance from the centre of the i key to the center of the j key across the keyboard. The variable W, the width, represents the width of the target key, which is the same for all alpha keys (see below). The variable MT_{ij} is the predicted time to move from the i key to the j key.

Notice that amplitude and width are measured in the same units and form a ratio in equation 3. This implies that a QWERTY keyboard will yield the same movement time regardless of its size, so long as the same relative scale is maintained. This will be true only within limits. For example, as keyboard size decreases, precision will be compromised as size of the targets diminishes. As keyboard size increases, movement times will increase (and precision may decrease) because forearm motion will be required where wrist-only motion was previously used.

There exists a problem with the predicted movement time as defined above. Fitts' law is only valid for movements from a start location to a target some non-zero distance away. When entering a character twice in succession, the starting character i is the same as the target character j, and the amplitude A is zero. Fitts' law does not apply. To circumvent this difficulty we modify our model:

$$MT_{ij} = \begin{cases} a + b \times \log_2\left(\frac{A_{ij}}{W} + 1\right) & \text{if } i \neq j \\ MT_{Repeat} & \text{if } i = j \end{cases} \tag{4}$$

where MT_{Repeat} is the time to activate a key a second time after it has already been activated. A discussion of how a value for MT_{Repeat} was obtained is given later.

To weight each MT_{ij}, we define P_{ij} as the probability of occurrence of the digram ij, (the character i followed by the character j). Note that the sum of all probabilities must be one; that is

$$\sum_{i,j \in \text{Alphabet}} \sum P_{ij} = 1$$

Then we define

$$\overline{MT} = \sum_{i,j \in \text{Alphabet}} \sum \left(P_{ij} \times MT_{ij}\right) \tag{5}$$

where $\overline{MT}$ is the mean time between keys with an alphabet of 27 symbols, as noted earlier. $\overline{MT}$ is a minimum because it is based solely on the time to move between keys and assumes no pause for thinking or finding keys.

Taking the reciprocal of the average movement time yields the average number of characters per second, which is easily transformed into words per minute:

$$CPS_{MAX} = \frac{1}{\left(\overline{MT}\right)} \tag{6a}$$

$$WPM_{MAX} = \frac{CPS_{MAX} \times 60}{5} \tag{6b}$$

Note that typists define a word as five characters, including spaces and punctuation. Equation 6 is our upper-bound model.

2.3 The Lower Bound

To compute a lower-bound typing rate, we use the Hick-Hyman law to add a visual scan time onto each entry. The Hick-Hyman law is expressed as

$$RT = a' + b' \times \log_2(n) \tag{7}$$

where n is the number of items to choose from, and RT is the choice reaction time, or the time required to make a choice among n items. The coefficients a' and b' are slope and intercept constants, similar to the coefficients a and b in Fitts' law.

To calculate the lower bound, we simply add the choice reaction time to the movement time per character:

$$CPS_{MIN} = \frac{1}{\left(\overline{MT} + RT\right)} \tag{8a}$$

$$WPM_{MIN} = \frac{CPS_{MIN} \times 60}{5} \tag{8b}$$

The minimum typing speed considers not only the movement time between keys, but the time it takes on average to find each key by someone who is unfamiliar with the QWERTY layout. Equation 8 is our lower-bound model.

2.4 Coefficients in the Model

The derivation presented thus far, culminating in Equations 6 and 8, expresses our upper-bound and lower-bound models in general terms, without assigning numerical coefficients. Predictions follow once the coefficients in the models are assigned values.

2.4.1. Fitts' Law Intercept, *a*.

The form of Fitts' law expressed by equation 1 contains coefficients for the intercept, *a*, and slope, *b*. It is generally accepted that the intercept should be zero. If the intercept were positive then the movement time for a task with $ID = 0$ is not zero (see equation 1). This seems unreasonable. Similarly, with a negative intercept, a task of zero difficulty has a negative predicted movement time, which also seems unreasonable. Fitts' (1954) original relationship excluded the intercept altogether. For these reasons, we used $a = 0$.

2.4.2. Fitts' Law Slope, *b*.

Choosing an appropriate value for the slope is a problem. The reciprocal of the slope is the bandwidth, measured in bits per second (bps). Fitts (1954) original work with a stylus reported bandwidth in the range 9.5 to 11.5 bps. In later work, Fitts and Peterson (1964) suggested a range of 14 to 22 bps. In our opinion, these values are too high. (A justification appears later.) The study by MacKenzie et al. (1991) measured the bandwidth for pointing tasks using a stylus as a computer input device and found a rate of 4.9 bps. The pointing task in the latter experiment was similar to the form of typing we are investigating.

Many reasons for the lack of consensus on bandwidth relate to diverse and inconsistent experimental methods, as identified in an earlier paper (MacKenzie, 1992). The calculations that follow were performed in duplicate using 14 bps, as suggested by Fitts and Peterson (1964), and 4.9 bps, as suggested by MacKenzie et al. (1991).

2.4.3. Key Distances, A_{ij}.

We used a facsimile of a QWERTY keyboard taken from IBM hardware documentation (IBM, 1984, pp. 4-6) to find the inter-key distances. Using an arbitrary origin, *x* and *y* coordinates were assigned to each key and inter-key distances were calculated using the Pythagorean identity. As an example, the distance across the keyboard from Q to P in our template was 18.9 cm.

As noted previously the actual size of the keyboard is irrelevant, within limits. It is assumed that the simulated QWERTY keyboard matches the proportions of the keyboard taken from the IBM documentation.

The position of the space bar cannot assume a fixed *x-y* coordinate, as with alpha keys, since the space bar spans the width of the keyboard. As illustrated in figure 2, the distances for the diagrams involving spaces depend on three keys, rather than two. For our lower-bound model we assume users do not optimize movements and proceed directly to the space bar at the end of each word (see figure 2a). For our upper-bound model we assume a strategy that minimizes the total distance of the letter-space-letter sequence; that is, the angle of approach to the space bar is the same as the angle of departure (see figure 2b).
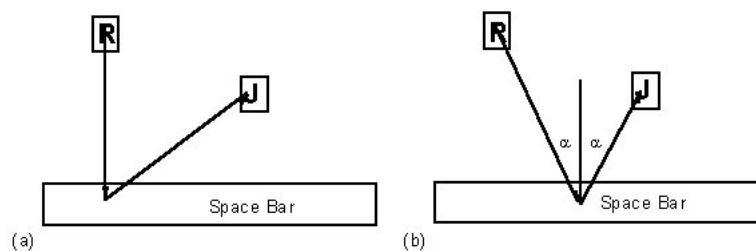
Figure 2. Distances to and from the space bar. (a) non-optimized movements for novices (b) optimized movements for experts.

Instead of computing distances explicitly for each letter-space-letter trigram, we computed each letter-space distance as a weighted average considering all possible letters that might follow, and, similarly, computed each space-letter distance as a weighted average considering all possible preceding letters. Although this sounds messy, the formulas, once established, are easily added to the spreadsheet containing the 27 × 27 digrams.

Note that the space bar of many modern computers does not extend from the edge of the left side of the alphabetic keys to the edge of the right side. In our model, we assume the space bar spans the width of the alpha keys. This a minor adjustment; however, the assumption simplifies the calculations of letter-space-letter distances. Since the space character is the most common, it is reasonable to assume that a soft keyboard would accommodate the prominence of the space bar.

2.4.4. Key Width, *W*.

Recall that the width, *W*, is measured in the direction of motion. Because the keys of the QWERTY keyboard are rectangular, the magnitude of *W* differs with the angle that the stylus approaches the target key. This complicates the model; however a pragmatic solution (with empirical support) is to set *W* to the minimum of the height and width of rectangular targets (MacKenzie & Buxton, 1992). This applies to all keys including the space bar. Thus, we used $W = 2.12$ cm for the width of the alpha keys and $W = 2.38$ cm for the width (actually, the height) of the space bar.

2.4.5. Hick-Hyman Law Intercept, *a*'.

Most interpretations of the Hick-Hyman law associate the intercept, *a*', with a transport lag resulting from the subjects' immediate reaction to the onset of the stimulus. However, when there is no uncertainty as to when the stimulus signal arrives, as in continuous text-entry, Welford (1968) suggests that the constant, *a*', is zero. Therefore, we set $a' = 0$.

2.4.6. Hick-Hyman Slope, *b*'.

As in Fitts' law, the reciprocal of the slope coefficient in the Hick-Hyman law, *b*', is called the bandwidth and is measured in bits per second. Bandwidth, in this context, is the rate at which humans process choices.

Welford (1968) maintains that for subjects in their twenties using key presses to signal choices, the reciprocal of the slope in the Hick-Hyman law lies in the range 5 to 7 bps. Since we are searching for a lower bound, we assume that the slowest choice processing speed is appropriate, and set $b' = 1/5 = 0.2$ seconds per bit. This is also the figure reported by Hick (1952).

With a 27-character alphabet, $n = 27$, so the visual scan time for novices in the lower-bound model is

$$RT = 0.2 \times \log_2(27) = 0.951 \text{ seconds}$$

2.4.7. Key Repeat Time, MT_{Repeat}.

A small informal experiment was conducted to determine a reasonable value for the mean key repeat time, MT_{Repeat} Six subjects entered one character repeatedly into an editor for one minute. They used a stylus to tap on a soft keyboard displayed on a Wacom *PL-100V* digitizing display/tablet. Every ten seconds a carriage return was entered by the operator from the physical keyboard. (This was done to simplify the subsequent analysis of the number of characters entered in each ten second period.) Because the physical keyboard was used to enter the carriage returns, the participants were not disturbed from their task. Of the 2,347 characters entered by all subjects, the average number depressed per second was 6.52, with a standard deviation of 1.15 characters. No significant difference was noted in the number of characters depressed across the six different ten-second periods. The results indicate that $MT_{Repeat} = 1 / 6.52 = 0.153$ seconds. This is close to the figure of 140 ms cited by Card et al. (1983, p. 60) for a typist repetitively pushing a key with a finger.

3. RESULTS OF MODEL DERIVATION

We have now completely defined all equations and coefficients for the upper-bound and lower-bound models. Since the models were developed using a spreadsheet, the results were available to us immediately upon entering the formulas. Table 2 summarizes the results. Using a bandwidth of 4.9 bps, we expect subjects can type (i.e., tap) at about 8.9 wpm without any prior experience with a QWERTY keyboard. With practice, this rate should increase to about 30.1 wpm. Using a bandwidth of 14 bps, the range is from about 11.0 wpm to 84.6 wpm. We feel the latter figure is much too high.

Table 2
Upper-Bound and Lower-Bound Predictions

|	Lower Bound		Upper Bound	
Bandwidth	Characters Per Second	Words Per Minute	Characters Per Second	Words Per Minute
4.9 bps	0.74	**8.9**	2.51	**30.1**
14 bps	0.92	10.98	7.05	84.56

4. DISCUSSION

Before comparing our results with empirical data from other studies, we present an observation that illustrates why a bandwidth of 14 bps is too high. Consider entering the sequence Z-X, which corresponds to neighbors on a QWERTY keyboard. Using $1 / b = 14$ bps, the predicted time to tap X following Z is

$$
\begin{aligned}
MT_{ZX} &= b \times \log_2\left(\frac{A_{ZX}}{W} + 1\right) \\
&= \tfrac{1}{14} \times \log_2\left(\frac{2.12}{2.12} + 1\right) \\
&= 0.071 \\
&= 71 \text{ ms}
\end{aligned} \tag{9}
$$

Clearly this is unreasonable because it is less than half the MT_{Repeat} time of 153 ms given earlier. For this and other reasons (see MacKenzie, 1992), we feel the bandwidth figure of 4.9 bps is more accurate.

4.1. Comparisons With Other Text Entry Methods

Figure 3 compares the results of our model with empirical data from several text-entry experiments.
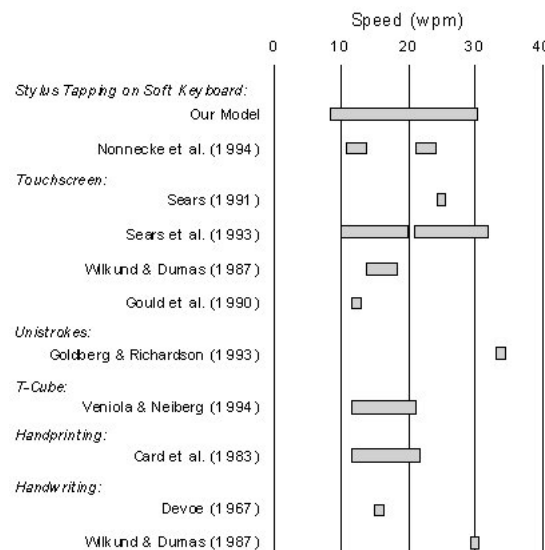
Figure 3. Performance comparison for stylus-tapping on a soft keyboard and several other text-entry methods.

The only experiment we are aware of that tested stylus-tapping on a soft keyboard is reported in MacKenzie, Nonnecke, McQueen, Riddersma, and Meltz (1994). Two different keyboard arrangements were used: a QWERTY layout and an alphabetic (ABC) layout. In the latter case, keys were arranged alphabetically in two horizontal rows with a space bar across and below the bottom row. The ABC layout was tested because of its good use of screen real

estate. Typing speeds for the QWERTY layout ranged from 21 wpm for the first block to 24 wpm for the ninth and last block. With prolonged practice, subjects would, no doubt, improve, perhaps leveling off near 30 wpm, as predicted in our upper-bound model.

Although the movement component of our model is inaccurate for the ABC layout (because the key distances were different), it is the visual scan time that dominates for novices. Therefore, the results for the ABC layout, which was unfamiliar to users, merit comparison with our lower-bound model. Typing speeds ranged from 12 wpm for the first block to 14 wpm for the ninth block. The entry rate of 12 wpm is slightly above our model's lower-bound rate of 8.9 wpm. We attribute this to the following. First, the alphabetic ordering of the ABC layout is not random; in fact, it gives subjects a good sense of where characters are located. This reduces the visual scan time that would occur for a random layout. Second, learning is "immediate". That is, after the first few entries, subjects are already gaining familiarity with the layout and cease to be novices in the strictest sense. In fact, an experiment to test our lower-bound model would require a completely new and random key assignment after each key stroke.

The closest entry method to stylus-tapping is the use of a soft keyboard on a touchscreen. Gould, Greene, Boies, Meluson, and Rasamny (1990) found entry rates of 12 wpm for experienced users. This rate seems low; however, the measurement included the extra touches to back-up the cursor, correct errors, and the time to review the input text mid-way through a block if subjects forgot the input string. Wilkund and Dumas (1987) found rates of 14-18 wpm in a similar experiment. Sears, Revis, Swatski, Crittenden, and Shneiderman (1993) measured entry rates with novice and expert users on two sizes of keyboards. Novice rates were 10 wpm on the small keyboard and 20 wpm on the large keyboard. Expert rates were 21 wpm on the small keyboard and 32 wpm on the large keyboard.

Sears (1991) found a rate of 25.4 wpm using expert subjects for a text-entry task with a touchscreen. The task was similar to that modeled here, since the only characters entered were the 26 alpha keys, a space bar, and a return key. The experiments described by Sears (1991), Sears et al. (1993), and Wilkund and Dumas (1987) allowed subjects to use two hands on the touchscreens, so comparisons with stylus-tapping are limited.

The text-entry method for pen-based computers that receives the most press is hand printing or hand writing with built-in recognition software. A "perfect" recognizer will be transparent to the user; so traditional research into hand printing and writing gives some insight into the best possible scenario for this provocative form of input. Hand printing speeds are well known to lie in the range of 12-23 wpm (Bailey, 1987; Card et al., 1983). So, even with highly accurate recognition software, entry speeds will not match experts tapping on a soft keyboard with a stylus. Although cursive handwriting rates can range from 16 wpm (Devoe, 1967) to over 30 wpm (Wilkund and Dumas, 1987), recognition errors exceed those for block printing (Wolf, 1987) and therefore exacerbate the algorithmic recognition of input.

At least two studies have attempted to re-design the Roman alphabet with simplified strokes. Goldberg and Richardson's (1993) *Unistrokes* assign each letter to a single stroke. This greatly simplifies the recognition software and permits eyes-free input with each stroke entered on top of preceding strokes in a dedicated scripting region. A formal experiment was not reported; however, one user attained rates of 33.6 wpm for "peak error-free" input. Veniola and Neiberg's (1994) *T-Cube* maps each letter to a flick gesture entered on top of a displayed radial menu. An experiment reported entry rates from 12 to 21 wpm. Notwithstanding the potential of techniques such as *Unistrokes* or *T-Cube*, the fact that a technique must be learned is a serious drawback.

Based on the results displayed in figure 3, our upper-bound prediction of 30 wpm fares quite well as a text-entry technique.

4.2. Weaknesses in our Model

Every effort was made to make our model comprehensive. Still, the predictions must be considered rough estimates. Although weaknesses in our model are easy to spot, we feel these represent less dominant aspects of the task than those that are accommodated. A few examples are offered.

Our model is limited to a 27-key soft keyboard with 26 alpha keys and a space bar. In practice, soft keyboards must accommodate the full spectra of text entry, including upper- and lower-case letters, punctuation symbols, function keys, etc. We limited our model to 27 keys because it made the task easy to define. Implementing the full complement of QWERTY keyboard features in a soft keyboard is easy, but performance predictions are difficult since they are highly task-dependent and they require a much larger table of digrams. Adding the period and comma to our table would be possible only if we undertook a very arduous process of data collection and tabulation similar to that reported by Mayzner and Tresselt (1965). Adding the less common graphic symbols (e.g., ~&"#) to our model is not worth considering, since they are not used in a consistent manner in "common English".

To obtain upper-case letters, a shift function can be implemented several ways. The simplest is to include a shift key which is touched once to transpose the next alpha key to uppercase. This method was implemented by Plaisant and Sears (1992) for a touchscreen. Caps lock may be implemented by a separate caps lock key or by double-tapping the shift key. In the latter case, caps lock ceases upon the next tap of the caps lock key.

The space bar on a QWERTY keyboard is much larger than the alpha keys and it is the most frequently entered key. Our model's formula for choice reaction time (equation 7) is a rough estimate; it assumes all choices are equally probable and are equally presented. Obviously, the space bar poses a problem in predicting visual scan time. Detailed treatments of choice reaction time replace $\log_2(n)$ in equation 7 with an information metric for the entropy of the distribution of choices. This would tend to reduce *RT* in our lower-bound model and increase the predicted entry rate for novices.

The predictions are also weakened by the fact that the user's hand will partially obscure the keyboard after certain key-taps (for example, in the middle or upper-left regions of the keyboard for right-handers). This has a slight impact on both the *RT* and *MT* predictions in the lower-bound model. Other weaknesses have already been noted, such as the validity and accuracy of the Fitts' law model predicting the movement time for each entry.

4.3. Alternate Key Layouts

Since our model is based on a spreadsheet in which each key is assigned an *x-y* coordinate and movement distances are calculated using the Pythagorean identity, it is easy to predict the performance attainable with alternate key layouts. For example, it is often claimed that the Dvorak layout is preferable to the QWERTY layout because key positions were assigned along principles of time-motion studies and scientific measurements of efficiency for two-handed touch typing (Potosnak, 1988). These benefits may not necessarily transfer to typing with a stylus on a soft keyboard, however. We tested this by re-assigning the *x-y* coordinate for each key to match the Dvorak layout. Using our preferred bandwidth figure of 4.9 bps, we predict a lower limit of 8.6 wpm and an upper limit of 27.1 wpm, slightly below the predictions for the QWERTY layout.

Since our model is for a soft keyboard, we can go beyond simply rearranging the key positions, however. We can also adjust key sizes, with the most probable symbols assigned to "big keys" and the least probable assigned to "small keys". Such a keyboard would have a large space bar in the middle, "big" keys around the space bar around the space bar for the most probable letters (e.g., e, a, t, o), and "small" keys around the outside for the least probably letters (e.g., q, x, z).

Despite a lot of interesting research with alternate key layouts, or even chord keyboards, the momentum of the QWERTY layout is substantial. The promise of payoffs in terms of expert performance levels has had little effect in convincing users to shun the venerable QWERTY. For this reason, we have not pursued alternate key layouts further than the brief exploration noted above.

4.4. Stylus Tapping and Soft Keyboards

Stylus tapping on a soft keyboard is an entry method well-suited to pen-based computers. However, it is often ignored in the popular press which focuses on software that accepts the sloppy scrawl of the ambivalent user. The performance of recognition software is quite another issue. The gap between manufacturer's claims and the goods delivered was expounded by Kurtenbach, Moran, and Buxton (1994), for example.

From a qualitative viewpoint, one drawback of stylus tapping is the lack of kinesthetic feedback: Users must maintain eye-fixation on the screen. This is also true of hand printing if an input grid is presented. Cursive handwriting recognition promises to ease this constraint by allowing input anywhere on the input device. The input is recognized, converted to ASCII, and delivered to the application's insertion point. Another drawback of soft keyboards is that they consume screen real estate. This is particularly troublesome with the small form-factor of personal digital assistants. As with current implementations on graphical user interfaces, the soft keyboard is usually displayed only while in use. There is also the "feel" of the LCD surface. Users often note that the touch is too slippery — unlike the texture of pen-on-paper. Parallax is also a problem, as noted by Goldberg and Goodisman (1991).

In our view, pen-based computers are unsuited to applications that require extensive text input. Veniola (1994) also raises this point. Indeed, the promise of pen-based computers lies in vertical markets, wherein the application is so highly understood and constrained that input is primarily through on-screen buttons, menus, etc. For input-intensive tasks, the venerable QWERTY keyboard — like the one used to write this paper — remains unchallenged. With input rates well beyond 40 wpm for expert users, the alternate forms of input cited herein are reduced to niche applications, such as information kiosks or banking machines, or to occasional input on pen-based computers.

5. CONCLUSIONS

We have presented a model that gives upper-bound and lower-bound predictions for entry rates using a pen or stylus to tap on a soft QWERTY keyboard. The model is based on a characterization of the text-entry task based on linguistic tables of frequency probabilities of digrams (including spaces) in common English. Visual scan time is predicted by the Hick-Hyman law for choice reaction time, and movement time is predicted by Fitts' law for rapid aimed movements.

With predicted rates ranging from about 9 wpm for novices to 30 wpm for experts, stylus-tapping is a viable input method. Considering the accuracy problems that plague systems supporting hand writing recognition, stylus tapping on a soft keyboard represents a fast and easy alternative for text entry on pen-based computers.

REFERENCES

Bailey, R. W. 1989, *Human performance engineering* (2nd ed.), (Prentice Hall, Englewood Cliffs, NJ).

Card, S. K., Moran, T. P. and Newell, A. 1983, *The psychology of human-computer interaction* (Erlbaum, Hillsdale, NJ).

Devoe, D. B. (1967). Alternatives to handprinting in the manual entry of data, *Transactions on Human Factors*, **HFE-8**(1), 21-32.

Epps, B. W. 1986, Comparison of six cursor control devices based on Fitts' law models, *Proceedings of the Human Factors Society 30th Annual Meeting*, 327-331.

Fitts, P. M. 1954, The information capacity of the human motor system in controlling the amplitude of movement, *Journal of Experimental Psychology*, **47**, 381-391.

Fitts, P. M. and Peterson, J. R. 1964, Information capacity of discrete motor responses, *Journal of Experimental Psychology*, **67**, 103-112.

Gibbs, M. 1993, March/April, Handwriting recognition: A comprehensive comparison, *Pen*, pp. 31-35.

Gregg, J. R., Leslie, L. A. and Zoubek, C. E. (eds). 1972, *Gregg shorthand dictionary* (McGraw-Hill, New York).

Goldberg, D. and Goodisman, A. 1991, Stylus user interfaces for manipulating text, *Proceedings of the ACM SIGGRAPH and SIGCHI Symposium on User Interface Software and Technology*, 127-135.

Goldberg, D. and Richardson, D. 1993, Touch-typing with a stylus, *Proceedings of the INTERCHI'93 Conference on Human Factors in Computing Systems*, 80-87.

Gould, J. D., Greene, S. L., Boies, S. J., Meluson, A. and Rasamny, M. 1990, Using a touchscreen for simple tasks, *Interacting with Computers*, **2**, 59-74.

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

Hyman, R. 1953, Stimulus information as a determinant of reaction time, *Journal of Experimental Psychology*, **45**, 188-196.

IBM Corporation. 1984, *IBM personal computer hardware reference library: Technical reference, revised edition* [Report #6361453] (IBM Corp., Boca Raton, Florida).

Kerr, B. A. & Langolf, G. D. 1977, Speed of aiming movements. *Quarterly Journal of Experimental Psychology*, **29**, 475-481.

Kurtenbach, G., Moran, T. P. and Buxton, W. 1994, Contextual animation of gestural commands, *Proceedings of Graphics Interface '94*, 83-90.

Landauer, T. K. and Nachbar, D. W. 1985, Selection from alphabetic and numeric menu trees using a touch screen: Breadth, depth, and width, *Proceedings of the CHI'85 Conference on Human Factors in Computing Systems*, 73-78.

MacKenzie, I. S. 1992, Fitts' law as a research and design tool in human-computer interaction, *Human-Computer Interaction*, **7**, 91-139.

MacKenzie, I. S. and Buxton, W. 1992, Extending Fitts' law to two-dimensional tasks, *Proceedings of the CHI'92 Conference on Human Factors in Computing Systems*, 219-226.

MacKenzie, I. S., Nonnecke, B., McQueen, C., Riddersma, S. and Meltz, M. 1994, A comparison of three methods of character entry on pen-based computers, *Proceedings of the 1994 Human Factors and Ergonomics Society 38th Annual Meeting*.

MacKenzie, I. S., Sellen, A. and Buxton, W. 1991, A comparison of input devices in elemental pointing and dragging tasks, *Proceedings of the CHI '91 Conference on Human Factors in Computing Systems*, 161-166.

Mayzner, M. S. and Tresselt, M. E. 1965, Tables of single-letter and digram frequency counts for various word-length and letter-position combinations,

Psychonomic Monograph Supplements, **1**(2), 13-32.

Plaisant, C. and Sears, A. 1992, Touchscreen interfaces for flexible alphanumeric data entry, *Proceedings of the 36th Annual Meeting of the Human Factors Society*, 293-297.

Potosnak, K. M. (1988). Keys and Keyboards. In M. Helander (Ed.). *Handbook of human-computer interaction* (pp. 475-494). Amsterdam: Elsevier.

Sears, A. 1991, Improving touchscreen keyboards: Design issues and a comparison with other devices, *Interacting with Computers*, **3**, 252-269.

Sears, A., Revis, D., Swatski, J., Crittenden, R. and Shneiderman, B. 1993, Investigating touchscreen typing: The effect of keyboard size on typing speed, *Behaviour & Information Technology*, **12**, 17-22.

Underwood, B. J. and Schulz, R. W. 1960, *Meaningfulness and verbal learning* (Lippincott, New York).

Van Cott, H. P. and Kinkade, R. G. (eds.) 1972, *Human engineering guide to equipment design* (U.S. Government Printing Office, Washington).

Veniola, D. and Neiberg, F. 1994, T-cube: A fast, self-disclosing pen-based alphabet, *Proceedings of the CHI'94 Conference on Human Factors in Computing Systems*, 265-270.

Welford, A. T. 1968, *Fundamentals of Skill* (Methuen, London).

Wilkund, M. E. and Dumas, J. S. 1987, Optimizing a portable terminal keyboard for combined one-handed and two-handed use, *Proceedings of the 31st Annual Meeting of the Human Factors Society*, 585-589.

Wolf, C. G. and Morrel-Samuels, P. 1987, The use of hand-gestures for text-editing, *International Journal of Man-Machine Studies*, **27**, 91-102.