Technical Report

IBM

Predictive Keyboard Design Study: Effects of Different Word Populations, Number of Displayed Letters, and Number of Trigraph Tables

James R. Lewis
David J. Allard
Harold D. Hudson

May 31, 1994

TR54.846

Boca Raton, Florida

Unclassified

Predictive Keyboard Design Study: Effects of Different Word Populations, Number of Displayed Letters, and Number of Trigraph Tables

James R. Lewis
David J. Allard
Harold D. Hudson

TR54.846

Abstract

As touchscreen interfaces become smaller, emulating keyboard input becomes more difficult. One solution to on-screen keyboard emulation is a predictive keyboard -- one that only displays a subset of the set of keyboard keys, with predictions based on tables of letter frequencies (primarily letter triplets, or trigraph frequencies). It is possible to change these tables as a function of the text that a user types, but this adaptive strategy will be useful only if all the text that a user types comes from the same text population, where the distribution of its trigraphs defines a text population. If the words that a user types come from different text populations, then an adaptive strategy will result in digraph and trigraph tables that are suboptimal for any specific text population. Thus, one issue in the design of predictive keyboards is whether the words that users type come from one or more text populations. A second issue is the number of keys to display in a predictive keyboard. The more keys displayed, the more likely it is that the desired letter will be immediately available. However, displaying more keys means more visual search demand on the user (due to more letters to search among) and increased motor skill demand (due to smaller buttons). A third issue is the number of trigraph tables used to predict the most likely letters. The more tables used, the better the prediction will be. However, the more tables used, the greater is the demand on computer memory.

We randomly selected 100 names (first and last) from a telephone book, 100 words from a dictionary, and 100 words from an English-language novel. As a control, we constructed 100 random 5-letter strings. We also varied the number of trigraph tables used by the predictive keyboard from three to four, and varied the number of displayed letters from six to eight. Using a predictive keyboard optimized for standard English text, we measured the number of times users would have to press an "Other" button to change the displayed letters from the currently displayed letters to the next most-likely set of letters.

The results showed statistically significant effects for all terms in a complete linear model representing this experimental design. The words from the standard text in the novel required the fewest "Other" button presses, and names (both last and first) required more. Only the standard text from the novel seemed to benefit from having four rather than three trigraph tables. All text types benefited from increasing the number of displayed letters, although the standard text from the novel benefited the least. The results indicate that, across text types, there is little benefit gained from including a fourth trigraph table. There is a potential benefit from increasing the number of displayed letters from six to eight, although this change would require testing with users to ensure that the additional demands from increased visual search and motor skill requirement do not cancel the increased likelihood of having the desired character available immediately. Finally, the differences among the different word types indicate that these words represent different word populations, suggesting that an adaptive strategy for trigraph tables would be less effective than a multiple tables strategy.

Introduction

As touchscreen interfaces become smaller, emulating keyboard input becomes more difficult. One solution to on-screen keyboard emulation is a predictive keyboard -- one that only displays a most-likely subset of the set of keyboard keys, with predictions based on tables of letter (primarily letter triplet, or trigraph) frequencies. The software uses these tables to order the letters of the alphabet by most-likely frequency of occurrence, and displays the letters that the user is most likely to type next. If the desired letter is not in the initial set, users can touch an "Other" button to display the next most-likely letters.

It is possible to change these tables dynamically as a function of the text that a user types, but this adaptive strategy will be useful only if all the text that a user types comes from the same text population, where the distribution of its trigraphs defines a text population. If the text that a user types comes from different text populations, then an adaptive strategy will result in digraph and trigraph tables that are suboptimal for any specific text population. Thus, one issue in the design of predictive keyboards is whether the types of text that users type come from one or more text populations. If so, it would be better to base prediction on multiple sets of tables, each derived from different text populations and dynamically controlled by the application to optimize predictive accuracy.

A second issue is the number of keys to display in a predictive keyboard. The more keys displayed, the more likely it is that the desired letter will be immediately available. However, displaying more keys means potentially more visual search demand on the user (because there are more letters to search among) and increased motor skill demand (because the buttons are smaller). Therefore, unless displaying more keys substantially improves the likelihood of display of the desired letter, it would be better to display a smaller number of keys.

A third issue is the number of tables used to predict the most likely letters. The fundamental trigraph table for prediction contains the probability of a letter given the immediately preceding two letters (with no blanks). An important secondary table identifies the probability of a letter given an immediately preceding blank (in other words, the likelihood of a letter's occurrence at the beginning of a word). Other trigraph tables that can improve prediction include one that contains the probabilities of a letter given a blank and one letter, and one for a blank and two letters. The more tables used, the better the prediction should be. However, more tables place a greater demand on computer memory. Therefore, unless additional tables provide a substantial improvement in prediction, it would be better to use a smaller number of tables.

This paper describes an experiment designed to investigate these three issues. (1) Do the types of text that users might type with a predictive keyboard come from the same or different text populations? (2) Does displaying more keys substantially increase the likelihood that the desired next key will be immediately available for typing? (3) Does using more trigraph tables substantially improve the accuracy of letter prediction?

Method

We randomly selected 100 names (first and last) from a telephone book, 100 words from a dictionary, and 100 words from an English-language novel. As a control, we constructed 100 random 5-letter strings. We also varied the number of trigraph tables used by the predictive keyboard from three to four (without and with the trigraph table for a blank and two letters), and varied the number of displayed letters from six to eight. Using a predictive keyboard optimized for standard English text, we measured the number of times users would have to press an "Other" button to change the displayed letters from the currently displayed letters to the next most-likely set of letters if users typed these strings. The predictive keyboard could accept lists of words stored in files, so users did not actually have to type the words into the predictive keyboard. A separate program entered the word lists into the predictive keyboard and kept track of the number of times a user would have to press an "Other" key to display the desired letter, and provided as output the average "Others" per character for each word.

Results

Table 1 shows the means and 95% confidence intervals for each combination of conditions in the experiment. The results of an analysis of variance (see Table 2) showed statistically significant effects for all terms in a complete linear model representing this experimental design. (We dropped the random set from the analysis of variance because the means from the random set were clearly different from the other sets, and including them in the analysis could have obscured other more meaningful comparisons.) The design included one between-words variable (word source) and two within-words variables (number of displayed letters and number of trigraph tables used), and all their interactions. The words themselves formed a random variable; all others were fixed.

Because the three-way interaction (word source x number of displayed letters x number of trigraph tables used) was significant, the first task of interpreting the analysis was to study that interaction. Figure 1 is a graph of the interaction, and shows that the manipulations of number of letters displayed and number of trigraph tables used affected all text types in the same direction, but with differing magnitudes.

Figure 2 shows the main effect of word source, Figure 3 shows the word source by number of characters displayed interaction, and Figure 4 shows the word source by number of trigraph tables used interaction. The words from the standard text in the novel required the fewest "Other" button presses, and names (both last and first) required more. Only the standard text from the novel seemed to benefit substantially from having four rather than three trigraph tables. All text types benefited from increasing the number of displayed letters, although the standard text from the novel benefited the least.

Table 1. Experimental Results (Average "Other" Presses per Letter, and Upper and Lower Limits of a 95% Confidence Interval Around the Average)

		Six Letters Displayed		Eight Letters Displayed	
		Tables Used		Tables Used	
Word Source	Calculation	Three	Four	Three	Four
Standard Text	Upper limit	0.53	0.42	0.32	0.26
	Average	**0.45**	**0.34**	**0.26**	**0.20**
	Lower limit	0.37	0.27	0.20	0.14
Dictionary	Upper limit	0.66	0.64	0.40	0.40
	Average	**0.58**	**0.56**	**0.33**	**0.33**
	Lower limit	0.49	0.49	0.27	0.27
First Names	Upper limit	0.85	0.81	0.57	0.53
	Average	**0.76**	**0.73**	**0.50**	**0.47**
	Lower limit	0.68	0.66	0.44	0.41
Last Names	Upper limit	0.82	0.78	0.51	0.47
	Average	**0.74**	**0.70**	**0.44**	**0.41**
	Lower limit	0.65	0.62	0.38	0.35
Random Letters	Upper limit	2.08	2.09	1.45	1.46
	Average	**1.96**	**1.97**	**1.35**	**1.36**
	Lower limit	1.84	1.85	1.25	1.26

Table 2. Analysis of Variance for Experimental Results

Source of Variance	Sums of Squares	Degrees of Freedom	Mean Square	F-Score	P(>F)
Between Words					
Word Source	22.47	3	7.49	16.32	0.000
Words w/Groups	181.75	396	0.46		
Within Words					
Letters Displayed	22.96	1	22.96	659.07	0.000
Letters Displayed x Word Source	0.84	3	0.28	8.08	0.000
Letters Displayed x Words w/Groups	13.79	396	0.03		
Tables Used	0.56	1	0.56	19.96	0.000
Tables Used x Word Source	0.32	3	0.11	3.84	0.010
Tables Used x Words w/Groups	11.12	396	0.03		
Letters Displayed x Tables Used	0.03	1	0.03	9.97	0.002
Letters Displayed x Tables Used x Word Source	0.03	3	0.01	4.27	0.006
Letters Displayed x Tables Used x Words w/Groups	1.06	396	0.003		

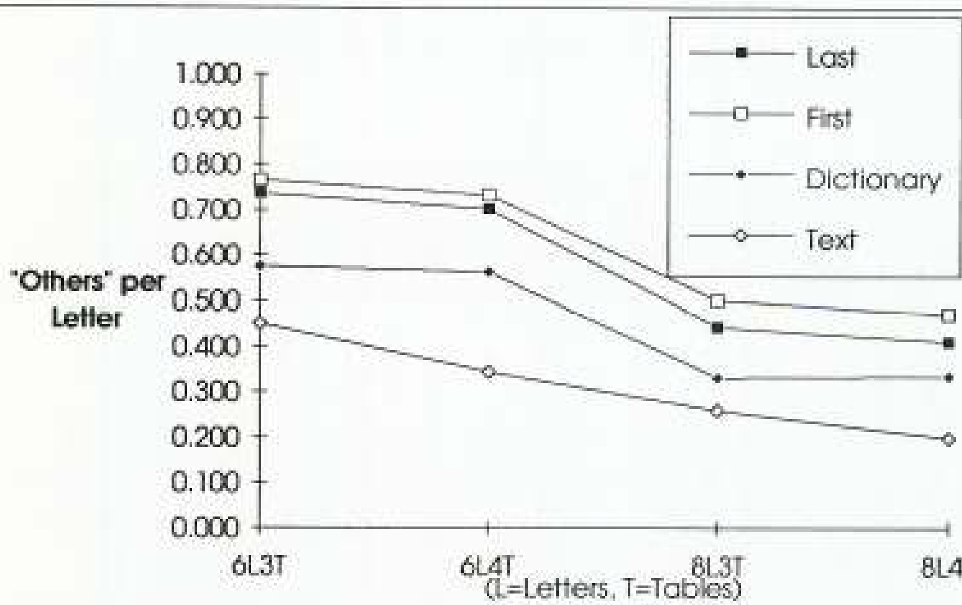

Figure 1. Word Source x Number of Letters Displayed x Number of Tables Used

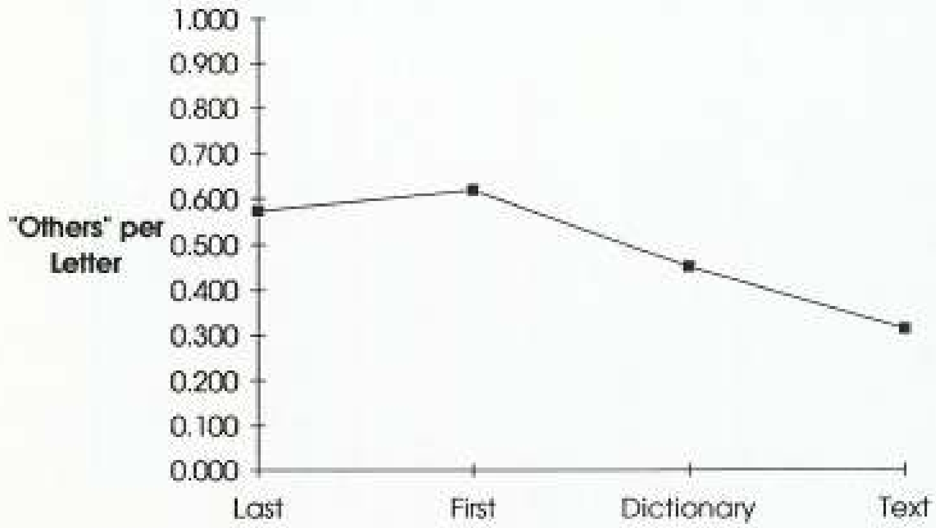

Figure 2. Main Effect of Word Source

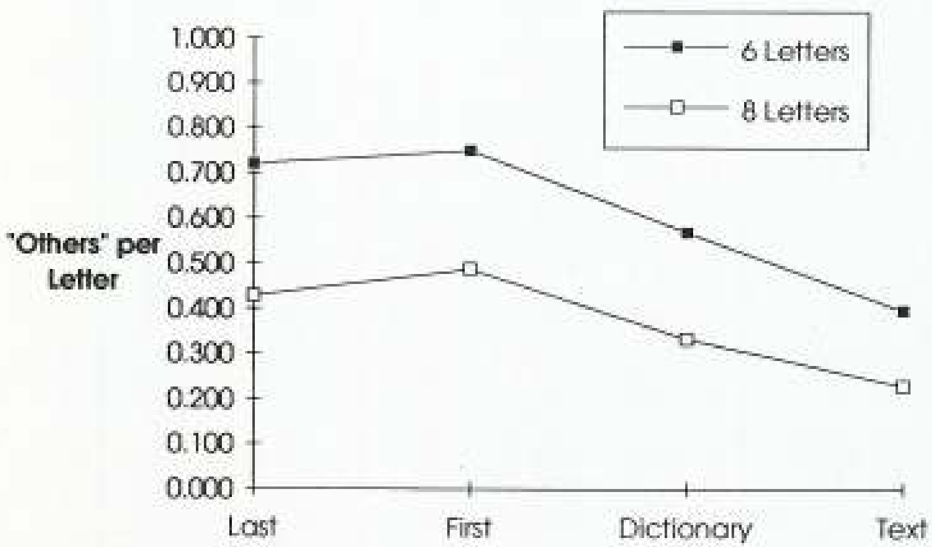

Figure 3. Word Source x Number of Letters Displayed

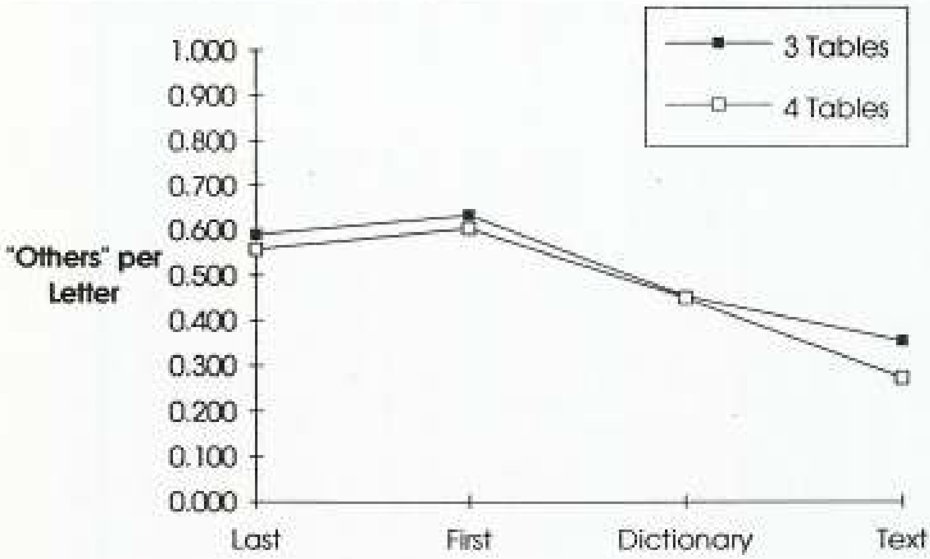

Figure 4. Word Source x Number of Tables Used

Discussion

The purpose of this experiment was to investigate three issues related to the design of predictive keyboards. (1) Do the types of text that users might type with a predictive keyboard come from the same or different text populations? (2) Does displaying more keys substantially increase the likelihood that the desired next key will be immediately available for typing? (3) Does using more trigraph tables substantially improve the accuracy of letter prediction?

The statistically significant differences among the different text types (word sources) indicate that these words represent different word populations. This suggests that an adaptive strategy for trigraph tables would be less effective than a multiple tables strategy because a single set of tables cannot represent the different text populations as effectively as multiple sets. This is especially important if users will type both normal text and names (names of people and names in addresses). In practice, it is important to balance the desire for maximized prediction accuracy with a device's memory constraints. Smaller devices may not have enough memory to accommodate multiple sets of tables. In this case, an adaptive strategy might be better than a multiple tables strategy. If adaptation is not feasible, then the cells of the tables should contain weighted probability averages from all the types of text that users might type with the device (such as normal text, names, and addresses).

There is a potential benefit from increasing the number of displayed letters from six to eight, which resulted in an average 40% decrease in the number of "Others" per letter. The words from the novel received the least raw benefit from increasing the number of displayed letters. This is reasonable because the original source for the tables was standard text, so the current tables should be better at predicting letters in standard text relative to words from other sources. Therefore, the correct letter is more likely to appear in the first six letter choices for standard text, reducing the benefit of displaying two additional letters for standard text relative to other text sources. The gain for standard text was large enough (42.3% fewer "Other"s per letter) to justify further investigation of the feasibility of increasing the number of displayed letters in the predictive keyboard. However, this change would require testing with users to ensure that the additional demands from increased visual search and motor skill requirement do not cancel the increased likelihood of having the desired character available immediately.

The results indicate that, across all word sources, there is little benefit gained from including the additional trigraph table. Only the normal text showed a substantial reduction in "Others" per letter (from .355 to .272, a reduction of 23%). The average reduction for the other text types was only 4%. However, because the additional table did improve predictive accuracy for its text type, it would be worthwhile to re-evaluate the value of the extra table given additional table sets derived from different word sources. As with the issue of an adaptive strategy versus multiple table sets, dropping the extra table is reasonable if its inclusion places too much demand on limited memory resources. Otherwise, it is better to include it in a predictive keyboard system.