A new Performance Measure Taking into Account the Mental Load in Mobile Text Entry Tasks

Franck Poirier, Hamed H. Sad
Université de Bretagne-Sud
VALORIA - Centre de recherche – 56000 Vannes-France
franck.poirier, sad@univ-ubs.fr

Abstract. Text entry research has received a lot of attention in recent years because of the need for more effective and usable entry methods on mobile devices. Technical limitations such as screen size have led to the design of entry interfaces that mentally load the user in order to obtain better performances. Current evaluation methodologies of these interfaces focus on text entry speed and error rate but don't pay enough attention to the mental load. In this paper, we concentrate on the evaluation of the load's effect on text entry process and we present a comparative evaluation of three mobile text entry methods with and without the application of a secondary task. We also define a performance measure that takes into account the mental load characteristic for a given text entry interface.

Keywords: Mobile text entry, mental load, evaluation, secondary task.

1 Introduction

Mobile text entry process can be viewed as the integration of two sub-processes. The first is the text creation in which an idea should be translated into words and sentences in a particular language. The second is the process of converting these words and sentences to an electronic form using a text entry technique with minimal resources (small screen, tiny soft keyboard...). In mobile context, these two sub-processes are always executed concurrently, which causes an increase in the mental load compared to the desktop interaction context. Moreover, in the mobile environment, the user may be on the move while executing the two processes, which also increases the load.

In the literature of human-computer interaction (HCI) many words and expressions have been used to describe "How busy is the operator?". In interface design, the term "*cognitive load*" is used. But, in this domain, this word is more general and indicates one or more of the following three components: perceptual load, central processing and motor load [3]. In this paper, we use the term "*mental load*" to indicate the amount of resources used for the perceptual and central processing activities, where memory operations and decision-making are involved.

2 Mental load in mobile text entry

A review of mobile text entry methods shows that, for their evaluation, the motor activities required for the interaction have received the most attention [4], [6]. Most of the evaluations of text entry methods are based on the entry speed expressed in Words per Minute (WPM). The mental load corresponding to a particular entry technique affects the entry speed and also the error rate, but these quantities do not totally reflect this load.

Comparisons of different text entry methods based only on motor components of the user-interaction are not sufficient. The mental operations associated with different methods may be very dissimilar, so the evaluation should take into account the mental load. For example, using a menu augmented soft keyboard [2] for suggesting the next character after each keypress, increases the speed, but also the cognitive load while typing. This increase cannot be reflected by the theoretical input rate based on motor movement computing. The mental load can be measured by the application of a secondary task [1]. The degradation of performance after using the secondary task reflects the difficulty of operating the tested technique.

In this work, we chose as secondary task the digit monitoring task. While operating the entry technique, the user should monitor voiced digits regularly separated in time, but randomly in order. The time period between two numbers is the factor that controls the level of difficulty. The user should respond to the repetition of two identical consecutive digits. This task requires auditory perception, working memory, and decision making. This will constitute a simple central executive task.

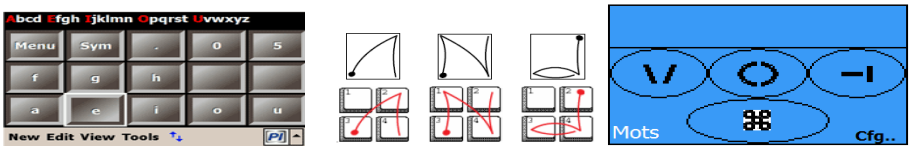
We think that the digit monitoring task is clearly a secondary task in relation to the text entry operation. Humans are usually capable of interacting with a visual based task like an entry method while manipulating auditory secondary task.

3 Experimental work

We compared three text entry methods. The first one was Phraze-It (www. prevalentdevices.com/) where a letter is entered by exactly two keystrokes (fig.1-a). The second method was 4-Key EdgeWrite (4KEW) [5] where the entry of a character is done by 2 to 7 keystrokes on four keys in a specific sequence (fig. 1-b). The third method, UniGlyph (fig. 1-c), developed in our lab, is based on a 3-key ambiguous keyboard where letter association with keys is based on graphic information.

We decided to apply the secondary task test for comparing these three methods. The subjects, 5 Master students, had already completed 5 sessions on the three methods. So, we assumed that the mental load caused by learning is lowered enough, and the application of the random digits monitoring secondary task gives us information about the mental load associated only with the method operation.

We carried out one new session for each participant, with the addition of the digit monitoring secondary task. We have presented the necessary instructions on the secondary task to each user before she/he began. The time period between two consecutively voiced numbers was set to 1.75 second. This period allowed the user to interact with entry and in the same time to be able to monitor the secondary task.



a b c

Fig. 1. The Phraze-it Keyboard (a), the 4KEW Keyboard (b) and the UniGlyph Keyboard (c)

Results

The result of secondary task application is shown in table 1 where the intersection of a participant and a text input method is the percentage secondary task missing rate. It is defined as the ratio of the number of detected repetitions to the total number of repetitions in percentage. A repetition is detected if the subject responded to it before the voicing of the following number. There is no significant difference between UniGlyph and 4KEW ("Average missing rate" row of table 1). There is more evident difference in performance degradation in WPM and accuracy. Phraze-It gives the lowest charge and the lowest percentage performance degradation.

Table 1. Secondary task missing rate percentage, WPM degradation and error rate increase. The error rate is the ratio of the correctly entered characters to the total number of characters.

Method / Participant	UniGlyph	Phraze_It	4 Key EdgeWrite
Participant 1	25	9,76	16
Participant 2	14,85	16,98	19,67
Participant 3	20	20	20
Participant 4	35	27	30
Participant 5	57	31	53
Average missing rate	30,75	20,94	27,73
Mean WPM degradation (%)	-41,84	-21,39	-24,8
Mean Error increase (%)	43,1	2,72	10,55

Analyses

First, it is clear that the application of the random number monitoring task leads to a remarkable performance degradation in the primary task. This indicates that the working memory subcomponent tapped by the secondary task is involved in the performance of the primary task and validates the applicability of our choice for the secondary task. The random number monitoring task seems to be suitable to discriminate between different methods based on the mental operating load.

We define a Performance Measure (PM) that takes into account the efficiency and the mental load characteristics for a given interface:

$$PM= (ES*EA)(1-MLWF*STMR*DIS*DIA)$$

where:

- ES: Entry Speed expressed in words per minute (wpm),
- EA: Entry Accuracy is (1- error rate), between 0 (error free) and 1
- MLWF: Mental Load Weighting Factor determines the importance of mental load effect in the evaluation (>0),
- STMR: Secondary Task Missing Rate between 0 (all repetitions were detected) to 1(all repetitions were missed),
- DIS: Degradation In Entry Speed as a result of secondary task application.
- DIA: Degradation In Accuracy as a result of secondary task application

The greater the performance measure is, the lower the user is cognitively loaded and more efficient is the text entry method.

The calculated values of PM for UniGlyph, Phrase-It, 4KEW are respectively 11.2, 9.0, 5.3, respectively, at MLWF=2 and 6.3, 8.8, 4.3 at MLWF=20. For small values of MLWF (MLWF =2), i.e. we are interested in the performance rather than the mental charge effect, it is found that UniGlyph is the best method. For greater values of MLWF (MLWF=20), i.e. the mental load effect is the more important, Phrase-It is the more suitable method. As all the variables change with practice, the performance measure should be determined at the same level of experience for all tested methods.

4 Conclusion

In this paper we argue the necessity of evaluating the operating mental load involved in mobile text entering by the secondary task technique. This approach allows a more complete evaluation and fair comparison of different entry methods. We propose to identify the mental load, at a certain amount of practice, by applying the digit monitoring secondary task. We defined a performance measure in order to rank mobile text entry methods according to the efficiency and the level of mental load.

References

1. Hegarty, M., Shah, P., & Miyake, A. (2000). Constraints on using the dual-task methodology to specify the degree of central executive involvement in cognitive tasks. *Memory & Cognition*, 28(3), 376-385.
2. Isokoski, P. (2004). Performance of menu-augmented soft keyboards. In *Proceedings of the SIGCHI conference on Human factors in computing systems*.
3. John, B. E., & Kieras, D. E. (1996). Using GOMS for user interface design and evaluation: which technique? *ACM Trans. Comput.-Hum. Interaction, 3*(4), 287-319.
4. Soukoreff, R. W., & MacKenzie, I. S. (1995). Theoretical upper and lower bounds on typing speed using a stylus and soft keyboard. Behaviour & Information Technology, 14.
5. Wobbrock, J., Myers, B., & Rothrock, B. (2006). Few-key text entry revisited: mnemonic gestures on four keys. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*.
6. Zhai, S., Hunter, M., & Smith, B. A. (2000). The metropolis keyboard - an exploration of quantitative techniques for virtual keyboard design. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*.