Proceedings of the Second IASTED International Conference
HUMAN-COMPUTER INTERACTION
March 14–16, 2007, Chamonix, France
ISBN Hardcopy: 978-0-88986-654-6 / CD: 978-0-88986-655-3

MOBILE TEXT ENTRY METRICS FOR FIELD STUDIES

Matti Koivisto
Mikkeli University of Applied Sciences
P.O. Box 181, 50101 Mikkeli
Finland
matti.koivisto@mikkeliamk.fi

ABSTRACT
In recent years, mobile text entry has been a flourishing research area. Typically text entry studies have been carried out in a laboratory environment where precise measurement methods and established accuracy and efficiency metrics can be applied. Laboratory settings provide a high level of control but they have one serious limitation – the lack of realism. Field studies, on the other hand, provide a more realistic research environment. Unfortunately however, the established methods and metrics cannot be applied in them. In this paper the metrics used in laboratory based text entry studies are analysed and new accuracy metrics for field based evaluations are developed. These two new metrics Correction Rate (CR) and subjective MSD can be used to measure corrected and uncorrected errors in free text strings typical of field tests.

KEY WORDS
HCI in Mobile, Ubiquitous and Pervasive Computing Contexts, Mobile Text Entry

1. Introduction

During the last decade we have witnessed a fast growth of mobile networks and the stationary Internet, but a combination of these two technologies – the mobile Internet - has not yet proved successful. There are without any doubt many reasons for moderate development of mobile information services. One of the main reasons is users´ disappointing experiences with the mobile Internet resulting from the limitations that distinguish mobile devices from conventional desktop PCs [3]. It is commonly stated that the mobile Internet will only become successful after these usability problems have been overcome.

One of the key limitations of mobile information systems like the mobile Internet is text entry. Due to the small size of mobile phones and PDA devices full size QWERTY keyboards cannot be used and multiple alternative input methods have been introduced. Although the human computer interaction for mobile devices is a relatively young research area, mobile text entry has received a lot of interest among scientists.

In a typical mobile text entry study a group of users write some predefined sentences in a controlled environment with multiple input methods. During and after the input process some metrics are collected and input methods are compared with each other. This kind of experiment has typical advantages and disadvantages of a laboratory study. It provides good control, but it is not realistic.

In mobile text entry studies the special characteristics of mobility are often forgotten. Many scholars have pointed out that the mobile Internet can be used in various highly dynamic contexts whereas the stationary Internet is mostly used in predetermined environments like offices or homes [1], [5], [7]. One could conclude thus that laboratory studies are not a perfect research method for mobile information systems. Field-based evaluations seem like a more appealing alternative. However, field test are not easy to implement because maximized realism is often achieved at expense of precision and control.

The aim of the study is to develop suitable metrics for field-based text entry tasks. The paper begins with a short review of existing mobile text entry methods and metrics used in earlier studies. I go through 30 recent studies and compare the metrics used in them. In Chapter 3 I introduce the special characteristics of field and laboratory experiments. In Chapter 4 I carry out a small scale text entry study which includes tasks both in a lab and in a realistic environment. The hybrid method is used to reveal the differences between the results of laboratory and field studies. The collected data are further used for developing metrics that are suitable for text entry tasks carried out in real contexts. Chapter 5 discusses results. Finally, in Chapter 6, I provide conclusion and point out some limitations of the study.

2. Mobile Text Entry Metrics

Input methods are considered as one of the most problematic issues in the mobile Internet devices [2], [12]. Because the small size of a mobile phone or a PDA device prevents the use of a standard QWERTY keyboard many alternative input methods have been suggested. Some examples are multitap, predictive entry methods such as T9, and many virtual keyboard layouts.

The suitability of the input methods for mobile text entry has been studied heavily with laboratory experiments. Some recent studies are shown in Table 1. The metrics used in those studies vary but they fall into three different categories: preference, efficiency and accuracy metrics.

Many scholars have limited their studies only to performance measures. Preference analyses are included in less than 40 per cent of the studies (see Table 1). Preference has been measured either with simple test specific questions or with established usability or workload questionnaires like NASA-TLX, After-Scenario Questionnaire (ASQ) or System Usability Scale (SUS).

In all the efficiency studies listed in Table 1 some sort of text entry speed metric was used. Speed was measured either on character, word or task level but actually all of them can easily be converted to the same unit (e.g. WPM, Words Per Minute). For example, CPS (Characters Per Second) is converted to WPM by using an equation wpm = 60/5 * cps, because the definition of a word for this purpose is five characters, including spaces or any other characters in the inputted text.

Whereas text entry speed is relatively easy to measure, the accuracy is not. According to Table 1, accuracy was typically measured with error rate but multiple variants of it were used, including Minimum String Distance (MSD), Total Error Rate (TER), Corrected Error Rate (CER), Not Corrected Error Rate (NCER), and Cost per Correction (CPC). In some studies Key Strokes per Character (KSPC) is also used as an accuracy metric for multi-tap input methods. See [4], [10], and [11] for details of these metrics.

All accuracy metrics compared two text strings the original or presented one and the written or transcribed one with each other. Any differences between these two strings were then considered errors. This kind of method suits for controlled laboratory tests where users carry out limited tasks and write predefined sentences. In field test tasks sentences are typically not restricted and this kind of comparison of text strings is difficult or impossible.

3. Pros and Cons of Laboratory and Field Tests

As mentioned above, laboratory studies have been the main method in mobile text entry research. Kjeldskov and Graham [6] have pointed out that laboratory experiments have been a dominating research method in mobile usability studies in general. Their study revealed that more than 70 per cent of mobile system evaluation studies have been conducted by means of laboratory experiments.

Table 1. Metrics used in Text-entry studies

Study	Preference Metrics	Efficiency Metrics	Accuracy Metrics
Bouteruche et al. (2005)	-	Writing time	error rate
Butts & Cockburn (2002)	Subjective learnability evaluation	WPM, efficiency evaluation	subjective accuracy evaluation
Canesta (2002)	discomfort level	WPM	error rate
Clarkson et al. (2005)	comfort level questionnaire	WPM	error rate (total)
Clawson et al. (2005)	-	WPM	accuracy rate
Cockburn & Siresena (2003)	NASA-TLX[1]	WPM	-
Commarford (2004)	After Scenario Questionnaire [2]	WPM	error rate (uncorrected)
Dwigdor & Balakrishnan (2004)	-	WPM	error rate
Evreinova et al. (2004)	frustration level	WPM	confusion matrix
Fleetwood et al. (2002)	input method preference questionnaire	WPM	MSD
Fleetwood & Fick (2004)	-	CPM	error rate
Gong et al. (2005)	-	WPM	TER,CER, NCER
Gong & Tarasewich (2005)	-	WPM	CER, NCER
Gong & Tarasewich (2006)	-	-	CPC
Green et al. (2004)	-	WPM	-
Isokoski & Käki (2002)	-	CPS	error rate
Isokoski & Linden (2004)	-	WPM	MSD, KSPC
Isokoski & Raisamo (2004)	-	WPM	MSD, KSPC
Koivisto & Urbaczewski (2005)	SUS	WPM	MSD, TER, CER, NCER
MacKenzie & Soukoreff (2002)	-	-	MSD, confusion matrix
Miniotas et al. (2003)	-	Entry time	MSD
Moore (2004)	-	WPM, CWPM	error rate
Oniszczak & MacKenzie (2004)	-	WPM	error rate, KSPC
Pavlovych & Stuerzlinger (2003)	-	WPM	error rate, KSPC
Roeber et al. (2003)	fatigue questionnaire	WPM	error rate (average)
Sirisena (2002)	NASA-TLX[1]	WPM	KSPC
Soukoreff & MacKenzie (2004)	-	-	TER,CER , NCER
Soukoreff & MacKenzie (2004)	-	WPM	MSD, TER, CER, NCER
Tarasewich (2003)	not reported	Writing time	error rate, KSPC
Wobbrock & Myers (2005)	-	WPM	error rate

[1] NASA-TLX is a subjective work load measurement tool rather than preference metric
[2] Only first two items of the ASQ was used

The benefits of laboratory studies are well known. The advantages include a large degree of control, the opportunity to focus on specific phenomena of interest and the possibility to use precise metrics [6], [13]. On the other hand, many scholars have argued that laboratory studies have many serious limitations when applied to mobile information systems. According to Zhang and Adipat [12] the main limitation of the laboratory testing method is that it ignores mobile context and the unreliable connection of the wireless network. Generally, the most important problem of laboratory experiments is their lack of realism.

Because mobility is the key issue in mobile information systems, field-based evaluations seem like an appealing, or even indispensable approach, for mobile text entry studies. The major advantage of field experiments is increased realism [6]. In field tests perceived usability is based on users' experiences in a real environment. Thus, the results are likely to be more realistic and reliable.

However, it is not easy to make field studies. Kjeldskov and Stage [7] have pointed out three challenges of field studies. Firstly, it is quite complicated to establish realistic field studies that capture key situations in the use-context. Secondly, it is not easy to apply to field settings the evaluation techniques typically used in laboratories. And thirdly, field evaluations make data collection more complicated and less controlled.

The reduced control in data collection means that you cannot always use the metrics developed in a laboratory environment with field bases evaluations. As mentioned earlier, accuracy metrics used in laboratory experiments compare presented and written sentences with each other. However, in a typical field test there are no presented sentences but users have either total or some freedom to decide the contents of the messages. Therefore, we need different accuracy methods and metrics for tests carried out in labs as opposed to field tests.

4. Developing Mobile Text Entry Metrics for Field Tests

In order to analyze and develop mobile text entry methods and metrics for field study I carried out a simple experiment. In the experiment test users were asked to send text messages with a mobile phone both in laboratory and in field environments. Because I wanted the users to be unaware of the data collection I used a special SMS program for Symbian mobile phones. What the users regarded as a normal SMS application was in fact a program that collected valuable text entry data during the message creation process. This way the data collection did not jeopardize unbiased results. The automatically collected metrics are the length of the message, the time to write the message and the amount of correction key presses.

In many text entry studies new input methods or virtual keyboard layouts have been tested. Because I wanted the test to be as realistic as possible I also applied a new input method to the users. Therefore I decided to use Nokia 3650 mobile phones in the experiment. This model has a typical 12 key keyboard but the layout of the keys is different from the usual one as can be seen from Figure 1. None of the test users had prior experience of this kind of keyboard.

Figure 1. Tested and typical mobile phone keyboards

In the field test part of the experiment I gave a mobile phone to the test users and asked them to send some text messages with our application. The number of messages, contents or receivers were not limited in any way. The users had the phone from an hour to one day and during that time they sent 2 - 7 messages. The average length of these messages was 42.7 characters.

I also collected some reference data in a laboratory test. I asked users to write three predefined sentences with the same device in a controlled environment. The average length of the laboratory messages was 24 characters. To avoid any learning effect, every second user did the laboratory part before and every second user after the field test. The later analyses indicated, however, that the order of the field and laboratory tests did not have any effect on the results.

The total number of users was 15. They represented three different age groups: early teenagers (13 – 16 years), young adults (23 – 26 years) and middle aged people (40 - 49 years). Each group had five users and the total number of messages sent was 104. All users were familiar with mobile phones and text messaging. All of them send messages every day or several times a week.

The results of the field and laboratory studies are shown in Table 2. According to the results the text entry speed was significantly slower in a field test as opposed to a lab test. The obvious reason for this was that in a laboratory test all sources of disturbance were minimized and the participants could concentrate on their tasks only. In real environments, test users were for various reasons often distracted away from what they were doing. This, of course, affected their performance. The users also made more corrections in the field than in a laboratory, but the difference was not as significant as in speed.

Table 2. Test results

	Lab test	Field test	T-test value	Sign. level
Text entry speed (wpm)	5.59 (1.79)	4.66 (1.73)	2.68	> 0.99
Correction rate (corrections per character)	4.3 % (0.073)	6.5 % (0.067)	1.55	0.88

As can be seen from Table 2 the accuracy metric used here is Correction Rate (CR) and it is calculated according to Equation 1.

$$CR = \frac{\text{Number of correction characters}}{\text{Total number of characters}} \quad (1)$$

The reason why the correction rate was used instead of the error rate is simple. The error rate metrics would have been applicable to lab messages but not to the field test ones. When people type text messages in a real environment, they create the text during the writing process. Because field based studies do not have reference messages, error rates cannot be calculated.

Although the correction rate can give us valuable information about corrected errors, it does not tell us anything about the errors that are not corrected. In order to evaluate the uncorrected errors as well some kind of reference text is needed. Because reference strings do not exist in field tests, we must create them. In laboratory tests the reference string is the starting point, and the transcribed text is derived from it. In field tests I suggest a reverse method where the reference is created afterwards from the transcribed text.

Following the classification used for example in machine translation [9] we can use either an objective or a subjective method in reference creation. In the objective method the reference does not depend on personal preferences whereas the subjective criteria are based on individual quality judgement.

The objective method used in the study was based on the spell checker of Microsoft Word 2003 (Finnish version). The messages were transferred from the mobile phone to Word. Word then underlined words with spelling mistakes with a red zigzag line. After that I made changes necessary to fulfill the spelling requirements of Word and then calculated the Minimum String Distance between these two strings: the original one and the one accepted by Word. The MSD calculation algorithm was based on [10]. I ignored all errors suggested by Word in names, abbreviations and greetings.

The subjective error criteria were based on the writer's own evaluations. Their own messages were shown to them afterwards, and they were asked to identify all spelling mistakes in their text. Again MSD values were calculated between the original and the subjectively corrected messages.

The average MSD values for the three different user groups (teenagers, young adults and middle aged users) are shown in Table 3. The subjective MSD values seem to be quite similar to all three groups in both settings. Objective values, in contrast, show remarkable differences. Interestingly, the high objective MSD values in field tests are typical of teenage users. This is because adult users tend to use standard written language in their messages. The younger generation uses more colloquial spoken language. According to the results Finnish teenagers "talk" with text messages as their parents "write" them.

Table 3. Measured objective and subjective MSD values

	Field test		Lab test	
Group	Objective MSD	Subjective MSD	Objective MSD	Subjective MSD
Teenagers	15.2 %	0.51 %	0.27 %	0.27 %
Young adults	2.93 %	0.19 %	0.00 %	0.00 %
Middle aged	1.72 %	0.00 %	0.27 %	0.27 %

5. Discussion

Soukoreff and MacKenzie have developed text entry methods for laboratory studies. Their latest method is based on delineating participants´ keystrokes into four classes (correct, incorrect, incorrect but fixed and fixes) and several metrics derived from them. Their most important addition to earlier methods was the idea of two kinds of errors: corrected and not corrected ones, which are measured with Corrected Error Rate (CER) and Not Corrected Error Rate (NCER). Other scholars have also pointed out the importance of the error types. For example Koivisto and Urbaczewski [8] have suggested that if accuracy is measured only with not corrected errors the accuracy measure is no longer correlating with the used input method. For this reason it is important to find a way of analyzing both corrected and uncorrected error in field tests.

Although the lack of original texts prevents the use of these metrics in field tests, the same logic should also be used in real environment experiments. Based on the findings of the experiments I suggest Correction Rate (CR) as a metric for corrected errors and subjective MSD for not corrected ones. It is also important to note that subjective evaluations should be done by someone belonging to the same demographic group with the test users. For example a middle aged researcher may have huge difficulties in analysing the errors in teenagers' messages.

The results clearly indicated that objective metrics are not suitable for error analyses in field tests. Although adult writers mostly use standard language in their messages, younger writers in particular use dialect and spoken language with colloquialisms. The varying use of language registers means that there is no single standard or reference for all writers.

Efficiency can be measured in field tests with text entry speed. I find WPM a standard that can well be applied to both lab and field settings. However, it must be remembered that there are more sources of disturbance in a field than in a laboratory environment. For example text message writing may be interrupted by a phone call or a friend's visit, and thus affecting the time measured.

6. Conclusion

Both laboratory and field studies have distinct pros and cons. They complement each other in usability testing of mobile applications. Laboratory experiments provide us with precise data collected in controlled but unrealistic settings. Field test metrics are not as accurate, but that is the price you have to pay for realism.

Experiences from laboratory experiments offer valuable information for field studies. Although the same metrics cannot always be used, established methods of laboratory experiments can often be applied. Based on the earlier work done in a laboratory I suggest two accuracy metrics for text entry experiments carried out in uncontrolled settings. Correction Rate (CR) and subjective MSD can be used to evaluate corrected and uncorrected errors. In efficiency measurements WPM gives a possibility to compare laboratory test results with those of field tests.

I admit that this study has some limitations. Firstly, the sample size in the experiment was quite small. However, I think that it was large enough to identify the essential differences between laboratory and field studies as well as the challenges of objective error evaluations in uncontrolled field tests.

Secondly the study was limited to only one language: Finnish. The differences between objective and subjective MSD values for teenagers cannot be generalised to other languages without further studies.

Further studies are still needed to analyse the reasons for the different efficient and accuracy results between laboratory and field tests. This was out of the scope of this study. Research work is also needed to reveal the meaning of mobile context to the usability of mobile information systems.

Acknowledgement

I want to thank Mr. Vesa Hartonen for the SMS test program. He implemented software as a part of his Bachelor Thesis in Mikkeli Univeristy of Applied Sciences under the supervision of the main author of the article.

References

[1] P. Bhagwat & S. Tripathi, Mobile computing. *Proc. IFIP International Conf. on Networks,* Madras, India, 1994.

[2] G. Buchanan, S. Farrant, M. Jones, H. Thimbleby, G. Marsden, & M. Pazzani, Improving mobile Internet usability. *Proc. 10th International Conf. on World Wide Web*, Hong Kong, 2001.

[3] M. Chae, J. Kim, H. Kim & H. Ryu, Information quality for mobile Internet services: a theoretical model with empirical validation, *Electronic Markets*, 12 (1), 2002, 38-46.

[4] J. Gong & P. Tarasewich, A new error metric for text entry method evaluation. *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, Montreal, Canada, 2006.

[5] H. Kim, J. Kim, Y. Lee, M. Chae & M. Choi, An empirical study of the use contexts and usability problems in mobile Internet. *Proc. 35th Annual Hawaii International Conf. on System Sciences*, Big Island, Hawaii, USA, 2002.

[6] J. Kjeldskov & C. Graham, A review of mobile HCI research methods. *Proc. Mobile HCI*, Udine, Italy, 2003.

[7] J. Kjeldsjov & J. Stage, New techniques for usability evaluation of mobile systems, *International Journal of Human-Computer Studies*, 60(5-6), 2004, 599-620.

[8] M. Koivisto & A. Urbaczewski, Accuracy metrics in mobile text entry. *Proc. 1st IASTED Conf. on Human-Computer Interaction*, Phoenix, USA, 2005.

[9] F. Och, Statistical machine translation: from singleword models to alignment templates, *Technical report*, RWTH Aachen 2003.

[10] W. Soukoreff & S. MacKenzie, Metrics for text entry research: an evaluation of MSD and KSPC, and a new unified error metric. *Proc. ACM Conference on Human-Factors in Computing Systems*, Ft. Laudertale, USA, 2003.

[11] W. Soukoreff & S. MacKenzie, Recent developments in text entry error rate measurement. *Proc. ACM Conference on Human Factors in Computing Systems*, Vienna, Austria, 2004.

[12] D. Zhang & B. Adipat, Challenges, methodologies, and issues in the usability testing of mobile applications, International Journal of Human-Computer Interaction, 18(3), 2005, 293-308.

[13] J. Wynekoop & S. Conger, A review of computer aided software engineering research methods. Proc. IFIF TC8 WG 8.2 Working Conf. on the Information Systems

Research Arena of the 90's, 1990.

Appendix: Text Entry Studies in Table 1

[1] F. Bouteruche, G. Deconde, E. Anquetil, & E. Jamet, Design and evaluation of handwriting input interfaces for small-size mobile devices. *Proc. 1st Workshop on Improving and Assessing Pen-Based Input Techniques*, Edinburgh, Scotland, 2005.

[2] L. Butts & A. Cockburn, An evaluation of mobile phone text input methods. *Proc. 3rd Australasian User Interfaces Conf.*, Melbourne, Australia, 2002.

[3] Canesta Usability Lab, The mobile input usability threshold, *White Paper*, 2002.

[4] E. Clarkson, J. Clawson, K. Lyons, & T. Starner, An empirical study of typing rates on mini-QWERTY keyboards. *Proc. ACM Conf. on Human Factors in Computing Systems*, Oregon, USA, 2005.

[5] J. Clawson, K. Lyons, T. Starner & E. Clarkson, The impacts of limited visual feedback on mobile text entry for the Twiddler and mini-QWERTY Keyboards. *Proc. 9th IEEE International Symposium on Wearable Computers*, Osaka, Japan, 2005.

[6] A. Cockburn & A. Siresena, Evaluating mobile text entry with the Fastap keypad. *Proc. British Computer Society Conf. on Human Computer Interaction*, 2003.

[7] P. Commarford, An investigation of text throughput speeds associated with Pocket PC input method editors, *International Journal of Human-Computer Interaction*, 17(3), 2004, 293–308.

[8] D. Dwigdor & Balakrishnan, A comparison of consecutive and concurrent input text entry techniques for mobile phones. *Proc. ACM Conference on Human Factors in Computing Systems*, Vienna, Austria, 2004.

[9] T. Evreinova, G. Evreinov, & R. Raisamo, Four-key text entry for physically challenged people. *Proc. 8th ERCIM workshop*, Vienna, Austria, 2004.

[10] M. Fleetwood, M. Byrne, P. Centgraf, K. Dudziak, B. Lin & D. Mogilev, An evaluation of text entry in Palm OS - Graffiti and the virtual keyboard. *Proc. 46th Annual Meeting Human Factors and Ergonomics Society*, Baltimore, USA, 2002.

[11] M. Fleetwood & C. Fick, Input rates for a one-handed input device (OHAI) for Chinese text entry. *Proc. 48th Annual Meeting Human Factors and Ergonomics Society*, New Orleans, USA, 2004.

[12] J. Gong, B. Haggerty & P. Tarasewich, An enhanced multitap text entry method with predictive next-letter highlighting. *Proc. ACM Conf. on Human Factors in Computing Systems*, Oregon, USA, 2005.

[13] J. Gong & P. Tarasewich, Testing predictive text entry methods with constrained keypad designs. *Proc. International Conf. on HCI*, Las Vegas, USA, 2005.

[14] J. Gong & P. Tarasewich, A new error metric for text entry method evaluation. *Proc. SIGCHI Conference on Human Factors in Computing Systems*, Montreal, Canada, 2006.

[15] N. Green, J. Kruger, C. Faldu, & R. St. Amant, A reduced QWERTY keyboard for mobile text entry. *Proc. ACM Conf. on Human Factors in Computing Systems*, Vienna, Austria, 2004.

[16] P. Isokoski & M. Käki, Comparison of two touchpadbased methods for numeric entry, *CHI Letters: Human Factors in Computing Systems*, 4(1), 2002, 25 – 32.

[17] P. Isokoski & T. Linden T., Effect of foreign language on text transcription performance: Finns writing English. *Proc. NordiCHI´04*, Tampere, Finland, 2004.

[18] P. Isokoski & R. Raisamo, Quikwriting as a multidevice text entry method. *Proc. NordiCHI´04*, Tampere, Finland, 2004.

[19] M. Koivisto & A. Urbaczewski, Accuracy Metrics in Mobile Text Entry. *Proc. 1st IASTED Conference on Human-Computer Interaction*, Phoenix, USA, 2005.

[20] S. MacKenzie & W. Soukoreff,, A character-level error analysis technique for evaluating text entry methods. *Proc. NordiCHI´02*, Aarhus, Denmark, 2002.

[21] D. Miniotas, O. Spakov & Evreinov G., Symbol creator: an alternative eye-based text entry technique with low demand for screen space. *Proc. 9th IFIP TC13 International Conf. on Human-Computer Interaction*, Zürich, Switzerland, 2003.

[22] R. Moore, Modelling data entry rates for ASR and alternative input methods. *Proc. 8th International Conf. on Spoken Language Processing*, Jeju Island, Korea, 2004.

[23] A. Oniszczak & S. MacKenzie, A comparison of two input methods for keypads. *Proc. NordiCHI´04*, Tampere, Finland, 2004.

[24] A. Pavlovych & W. Stuerzlinger, Less-tap: a fast and easy-to-learn text input technique for phones, *Graphics Interface online papers*, 2003.

[25] H. Roeber, J. Bacus & C. Tomasi, Typing in thin air: the Canesta projection keyboard. *Proc. ACM Conf. on Human-Factors in Computing Systems*, Ft. Laudertale, USA, 2003.

[26] A. Sirisena, Mobile text entry, Department of Computer Science University of Canterbury Christchurch, New Zealand, 2002.

[27] W. Soukoreff & S. MacKenzie, Metrics for text entry research: an evaluation of MSD and KSPC, and a new unified error metric. *Proc. ACM Conf. on Human-Factors in Computing Systems*, Ft. Laudertale, USA, 2003.

[28] W. Soukoreff & S. MacKenzie, Recent developments in text entry error rate measurement. *Proc. ACM Conf. on Human Factors in Computing Systems*, Vienna, Austria, 2004.

[29] P. Tarasewich, Evaluation of thumbwheel text entry methods. *Proc. ACM Conf. on Human-Factors in Computing Systems*, Ft. Laudertale, USA, 2003.

[30] J. Wobbrock & B. Myers, Gestural text entry on multiple devices. *Proc. 7th ACM Conf. on Computers and Accessibility*, Baltimore, USA, 2005.