Chapter 2
Soft Keyboard Evaluations:
Integrating User's Background in Predictive Models

Bruno Merlin
UFPA, Brazil

Mathieu Raynal
Université Paul Sabatier, France

ABSTRACT

Predictive models, based on cognitive and motor laws (modeling some aspect of human behaviors), were supposed to provide an efficient tool to compare quickly soft keyboards performances, but these models are confronted with some limits. Moreover, they failed to predict efficiently the performances during the first usage that got an important impact on soft keyboard acceptability. To improve these models, the authors integrated into them the character search strategies oriented by the user's background. They illustrate their purpose by modeling three keyboards and comparing the results with experimental ones. The models integrating the user's background predicted results close to the experimental ones. However, those models must be adapted to the function of the keyboard and the targeted population. The applicability of the models as a rapid comparison tool for soft keyboards must be questioned.

INTRODUCTION

With the expansion of mobile phones with touch screens, increasing the input speed with soft keyboards became an important challenge. Several authors demonstrated that the classical soft keyboards, such as mini-QWERTY (Goldberg & Richardson, 1993; Raynal & Vigouroux, 2005; Merlin & Raynal, 2009) or phone keyboards (Foulds, et al., 1987; Lesher & Moulton, 2000; Levine, et al., 1987) are non-optimal solutions to input text with mobile devices or as assistive technologies. However, in spite of several alternatives elaborated to improve the soft keyboard performances, these classical input solutions remain the most popular ways to input text with a pointing device (finger, stylus, mouse, etc.) until now.

DOI: 10.4018/978-1-4666-1628-8.ch002

Toward this difference between the evaluated performances and the usage reality, we implicitly observe that the variables studied to compare soft keyboard performances are not able to reflect the motivation for these users' preferences. A brief state of art illustrates that, in general, the variables used to compare the different input systems refer to the expert performances: text input speed in word per minute (wpm) or character per second (cps) and error rate obtained after a long period of usage. However, some other researches (Isokoski, 2004; Raynal & Vigouroux, 2005) point out that, out of experimental contexts, a user may not accept a regression of his performances during the first usages. Therefore, a new artifact should provide beginner performances better or at least close to the performances obtained with the classical soft keyboards. More other, some characteristics (such as a high cognitive cost required to use the keyboard), may prejudiced a solution acceptation. However, these criteria are ignored in the global comparative reviews (MacKenzie & Zhang, 1999) of soft keyboards.

Long term usage is the only way enabling a user to reach expert performance. Out of an experimental context, it means that a tool must get a good acceptability at short term to enable a usage during a long term. Thus, the beginner performances should be one of the main comparison criteria.

In order to ease this comparison, Soukoreff and MacKenzie (1995) proposed a model establishing predictive performances for both expert and novice users: the upper and lower-bound. This model is based on behavior models such as Fitts' law (Fitts, 1954) and Hick-Hyman's law (Hick, 1952; Hyman, 1953). However, they concluded that the model is bad for novice performances. Das (Das & Stuerzlinger, 2008) improved this model, but identified that the notion of beginner is very relative because the user "learn[s] some features of a technique very quickly," prejudicing the reliability of the results. Therefore, an improved model to compare the beginner performances should be very valuable.

Out of the problem identified by MacKenzie and Das, we think that the user's background may influence users' performances in the first usages of a new soft keyboard. The background of the targeted population is frequently used (Isokoski, 2004) to justify design choices for soft keyboards. Therefore, it would be reliable to integrate it in a model predicting beginner performances.

In this chapter, we relate the practical and theoretical problems encountered in the application of the predictive models. Then, we propose to improve the model predicting the novice performances. We integrated the features representing user cultural experience toward the keyboards into this model.

To illustrate our study, we selected three soft keyboards as examples and compared the performances calculated by the initial predictive models, by our predictive models and performances measured experimentally.

Then we discuss the constraints to apply the model, and moreover the global limits of predictive model in HMI so as to keyboard evaluation. At last, we present our future axes of research to improve soft keyboard evaluation and comparison.

BACKGROUND: PREDICTING HMI PERFORMANCES AND TEXT INPUT SPEED WITH SOFT KEYBOARDS

In HMI, several cognitive models (such as KLM model [Card, et al., 1980] and GOMS [Card, et al., 1983]) and cognitive laws (such as Fitts's [Fitts, 1954] and Hick-Hyman's law [Hick, 1952; Hyman, 1953]) enable quick comparisons of different tasks and interaction patterns by predicting the time required to perform the interaction / the task. However, these models do not take into account complex cognitive processes or design criteria. Their results remain reliable for repetitive