Attention Demands in Text Entry Interfaces

I. Scott MacKenzie
Department of Computer Science and Engineering
York University
Toronto, Ontario Canada M3J 1P3
mack@cs.yorku.ca

This work seeks to identify, quantify, and model attention demands in text entry interfaces. While it is relatively straight-forward to model the motor-component of text entry [6-8, 10], the situation is quite different, and much more complex, if the interaction involves pauses or hesitations to visually attend to or consider options in the interface.

Attention demands exist in some models of interaction. For example Card et al.'s keystroke level model [2, 3] includes an operator for mentally preparing for an input action. Their model does not specifically acknowledge or quantify a switch in visual attention, however.

Others have acknowledged the need to more carefully consider the effect of attention demands in text entry interfaces. Bouteruche et al. [1] offer six design principles for mobile devices, and two pertain directly to attention: principle #1 is "minimize the number of attention areas", and principle #2 is "minimize the distance between attention areas to avoid switch of attention".

Attention demands were also addressed in previous work specifically on text entry where, for example, a distinction was drawn between text copy tasks and text creation tasks [7]. In a text copy task, the user's attention switches between the source text and the interface. Experimentally, the shift poses a problem because the time taken for the gaze shift and for visual and cognitive processing serves to push the entry speed down, yet the additional time is not intrinsic to the interface. In a text creation task, the user enters text from memory. Similar attention shifts are not needed and, presumably, do not occur. Free form text creation (e.g., "Write a paragraph about your weekend.") is problematic, because the user will no doubt hesitate to think about, or ponder, what to enter, and this also pushes entry speed down. Most text entry evaluations use a controlled text creation task with short phrases of memorized text. It is felt that the benefit in eliminating or minimizing attention demands out weighs the cost of reduced external validity (due to a procedure that is atypical of usage). Often the phrase disappears with the first keystroke to further ensure that the user does not switch the focus of attention between the source text and the entered text.

While minizing of avoiding attention demands seems reasonable, at least in experimental research, some text entry methods inherently require attention operation. Examples include methods involving keyboard disambiguation or word prediction/completion. If a user is sending an SMS message saying "today is the first day of autumn", and the entry

Dagstuhl Seminar Proceedings 05382
Efficient Text Entry
http://drops.dagstuhl.de/opus/volltexte/2006/518

method is T9, the key presses required for the word "autumn" are 288866*. The final asterisk is necessary because the numeric key presses do not map to a single word in the dictionary. The pattern also maps to a more common word, "button", so an additional key press is necessary. Since the user – even an expert user – is unlikely to know this, the input method poses an attention demand on the user.

To quantify or model entry speed or time, then a first-level approach is to use Fitts' law to describe and predict the movement of the thumb, finger, or stylus about a keyboard, and there are many examples [8, 10-12]. But, this is to ignore the time cost to visually attend to the display. During text entry on a mobile phone, most users are likely to focus primarily on the keypad as keys are pressed, but attention shifts are needed, as illustrated shortly. Models of text entry have not addressed this in any comprehensive manner. How long does the attention shift take? Does the user attend to the display after each key press, just once at the end of each word, or not at all? Are such behaviours different for novices than for experts? So, we have at least a three-pronged problem: identifying when attention events occur, quantifying the duration of such events, and developing a reasonable model of user behaviour that accommodates the range of users involved.

For the example word, it is interesting to chart the progress of the mobile phone's display as entry proceeds:

Key presses	Display
2	a
28	at
288	but
2888	butt
28886	butto
288866	button
288866*	autumn

Measuring the time to enter the word is straightforward; however, researchers in text entry are often interested in a priori analyses – viz. modeling – to explore interesting design possibilities. Modeling the motor component of input is simple; however, if the interaction includes perceptual or cognitive operations, all bets are off.

Most mobile phones beep or vibrate on incoming text messages, or when the battery is low; however, there is preciously little feedback that accompanies text entry. If during T9 text entry of a word, there is no dictionary entry for the current key sequence, there is visual feedback (my phone displays "SPELL?"), but there is no aural or tactile feedback. There is no point continuing, yet the user attending to the keypad will have no sense of this, and may just continue pressing keys. This point is a sidebar to the topic of this presentation, as an opportunity for a design modification, rather than as a problem or challenge in modeling the interaction. Yet, it seems reasonable to suggest that aural and/or tactile feedback could partner with the attention needs of an interface. This, at present, does not appear to be the case.

If the interface involves word completion, accommodating attention in the model is even more important because users must attend to the display to benefit from the design's most important asset: the candidate word list. For the example word, the following is one possibility of the display, assuming a five-word candidate list:

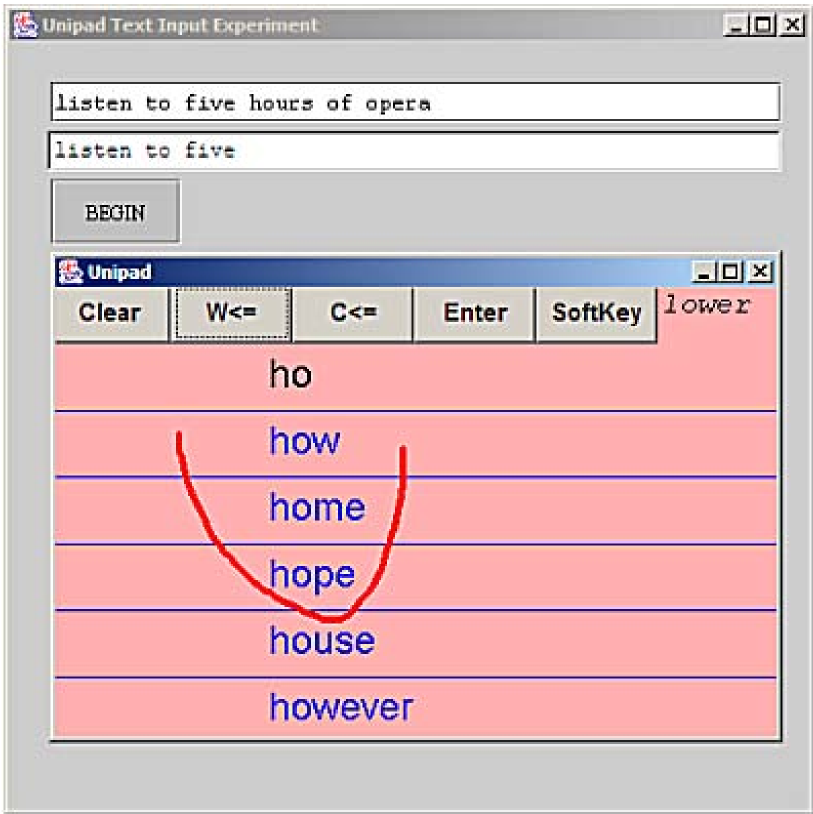
Input	Display
a	and a as at are
au	authority authorities august audience author
aut	authority authorities author autumn automatically
autu	autumn autumnal autumns
autum	autumn autumnal autumns
autumn	autumn autumnal autumns

If we assume a telephone keypad, the display might look as follows:

Key presses	Display
2	a and be by at
28	at bt but business available
288	but cut authority attention authorities
2888	autumn cutting butter button buttons
28886	autumn button buttons
288866	autumn button buttons

There are numerous design issues in designing a word completion interface – such as the number of key presses before the list appears, the minimum size of candidate words, or word suffix exclusion or completion rules – but these are not the primary focus here. We return to similar questions posed earlier. How often does the user attend to the display, and what is the performance cost of such operations? There is an additional component of the interaction because the user is scanning a list of size greater than one. Some form of the Hick-Hyman law no doubt is at work in the interaction [4, 5, 9].

It was noted as a sidebar that appropriate feedback might help reduce the performance cost of, or the need for, certain attention operations. Since "*efficient* text entry" is the topic for this workshop, opportunities such as this are perhaps more than a sidebar. For word completion interfaces, one possible way to reduce the performance cost of attending to list is to put the list in close proximity to the users primary attention point, thus reducing the gaze distance or bringing the candidate list in the proximal field. For stylus-based systems, putting the candidate list directly under the entry surface might help, as shown in the following screen shot of a prototype interface.



Here, the candidate list is directly under the display's digital ink. Words – i.e., the desired word – may catch the user's attention without the need for a gaze shift. This hypothesis remains to be tested.

References

1. Bouteruche, F., Deconde, G., Anquetil, E., and Jamet, E., Design and evaluation of handwritten input interfaces for small-size mobile devices, *HCI 2005 Workshop on Improving and Assessing Pen-Based Input Techniques*, (British HCI Society, 2005).
2. Card, S. K., Moran, T. P., and Newell, A., The keystroke-level model for user performance time with interactive systems, *Communications of the ACM*, *23*, 1980, 396-410.
3. Card, S. K., Moran, T. P., and Newell, A., *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum, 1983.
4. Hick, W. E., On the rate of gain of information, *Quarterly Journal of Experimental Psychology*, *4*, 1952, 11-36.
5. Hyman, R., Stimulus information as a determinant of reaction time, *Journal of Experimental Psychology*, *45*, 1953, 188-196.
6. MacKenzie, I. S. and Soukoreff, R. W., A model of two thumb text entry, *Proceedings of Graphics Interface 2002*, (Toronto: CIPS, 2002), 117-124.
7. MacKenzie, I. S. and Soukoreff, R. W., Text entry for mobile computing: Models and methods, theory and practice, *Human-Computer Interaction*, *17*, 2002, 147-198.

8. MacKenzie, I. S. and Zhang, S. X., The design and evaluation of a high-performance soft keyboard, *Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI '99*, (New York: ACM, 1999), 25-31.
9. MacKenzie, I. S. and Zhang, S. X., An empirical investigation of the novice experience with soft keyboards, *Behaviour & Information Technology*, *20*, 2001, 411-418.
10. Silfverberg, M., MacKenzie, I. S., and Korhonen, P., Predicting text entry speed on mobile phones, *Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI 2000*, (New York: ACM, 2000), 9-16.
11. Soukoreff, W. and MacKenzie, I. S., Theoretical upper and lower bounds on typing speeds using a stylus and soft keyboard, *Behaviour & Information Technology*, *14*, 1995, 370-379.
12. Zhai, S., Hunter, M., and Smith, B. A., Performance optimization of virtual keyboards, *Human-Computer Interaction*, *17*, 2002, 229-269.