Human–Computer Interaction – INTERACT '90
D. Diaper et al. (Editors)
Elsevier Science Publishers B.V. (North-Holland)

THE ROLE OF VISUAL AND KINESTHETIC FEEDBACK IN THE PREVENTION OF MODE ERRORS

Abigail J. SELLEN†, Gordon P. KURTENBACH, and William A. S. BUXTON

Computer Systems Research Institute, University of Toronto, Toronto, Ontario, Canada, M5S 1A1*

The use of visual and kinesthetic feedback in preventing *mode errors* was investigated. Mode errors were defined in the context of text editing as attempting to issue navigation commands while in insert mode, or attempting to insert text while in command mode. Twelve novices and twelve expert users of the Unix-based text editor *vi* performed a simple text editing task in conjunction with a distractor task in four different conditions. These conditions consisted of comparing the use of keyboard versus foot pedal for changing mode, crossed with the presence or absence of visual feedback to indicate mode. Both visual and kinesthetic feedback were effective in reducing mode errors, although for experts visual feedback was redundant given that they were using a foot pedal. Other measures of system usability indicate the superiority of the use of a foot pedal over visual feedback in delivering system state information for this type of task.

1. INTRODUCTION

Mode errors as originally defined by Norman (1981) occur when a situation is misclassified resulting in actions which are appropriate for the analysis of the situation but inappropriate for the true situation. Mode errors in text editing are very common. Users attempt to issue commands when the system is actually in "text insert mode" or attempt to enter text while actually in "command mode". While mode errors frequently occur with computers, examples from diary studies of action slips (Norman, 1981; Reason and Mycielska, 1982; and Sellen, 1990) reveal mode errors occur in many other aspects of everyday experience. Examples such as trying to fast forward the tape in the VCR when in "record mode", or turning the key in the ignition when the car engine is already running could both be called mode errors.

In the context of computers, any given action can have very different effects depending on the state of the system. Fortunately the consequences of most mode errors are only minor inconveniences, and in well designed systems, are usually reversible. However, such errors in poorly designed interfaces or in highly complex systems such as aircraft and nuclear power plants can result in far more serious outcomes. In such situations, the importance of preventing such errors, or at least absorbing their effects, is critical.

Errors are not the only metric with which to measure users' problems with mode identification, however. In some cases, the user may diagnose the correct mode, but only after experiencing confusion or uncertainty. In such cases, the appropriate measure is in terms of the cognitive effort or decision time required to deduce the system state. Increased cognitive effort may in turn be reflected in users' opinions of the usability of the system.

Why not just do away with modes? This was the opinion voiced strongly by Tesler (1981). But almost everything we do involves modes in one way or another, including working with so-called "modeless" computer systems such as the Apple Macintosh. Whenever dialog boxes appear, or whenever the cursor changes from an arrow to an "I-beam" depending on its location on the screen, one is in a mode. What is actually meant by a "modeless" system in this context is design in which contextual information is provided to minimize mode errors, and where modes can easily be entered and exited. Further, while the number of elemental actions available to interact with systems remains relatively constant, the number of functions within an application is growing. One has only to look at applications such as Hypercard to see how modes are used to support rich functionality.

It is not clear that we can ever hope to completely eliminate the problems associated with modes, but it certainly seems possible to reduce them. One obvious solution seems to be to give users more salient feedback[1] on system state. Apart from the practical importance for system designers, this raises some interesting theoretical questions: What kind of feedback is most salient to the user? Through what perceptual modality is the feedback best delivered? At what point does feedback become obtrusive? One objective of our research is to shed some light these issues.

There is little directly relevant literature. The exception is Monk (1986) who investigated the use of auditory feedback in preventing mode errors. In this study, Monk demonstrated that mode errors could be reduced by a third by using a key-contingent sound change depending on the mode of the system. Monk argued that sound is a good choice for system feedback in that users do not constantly look at the display while working.

† Also at the Institute for Cognitive Science, University of California, San Diego, La Jolla, CA. 92093

* This research was supported by the Natural Sciences and Engineering Research Council of Canada, Digital Equipment Corporation, and Xerox PARC.

A variation on the Monk experiment could have been to present the feedback in the form of a sustained tone whose timbre (sound quality) depended on the current mode. In contrast to the action-contingent sound used in the actual experiment, subjects in this instance could determine the current mode before initiating a possibly erroneous action. This kind of feedback might be called "action-independent".

These examples illustrate that feedback can be characterized along several dimensions, including:

- Modality of delivery (visual, auditory, kinesthetic)
  Through what sensory modality is the information delivered?
- Action-contingent versus action-independent delivery
  Does the feedback depend on an action being executed?
- Transient versus sustained delivery
  How long does the feedback last?
- Demanding versus avoidable feedback
  Can the user choose not to monitor the feedback?

These are not all necessarily orthogonal. For example, feedback using the visual channel is generally avoidable: one can easily choose not to monitor visual information. Kinesthetic and audio feedback however, are more inherently demanding and inescapable.

Choosing the best way of delivering system state information clearly must be dependent on the task. All else being equal though, it seems reasonable that the more salient the feedback, the more effective it will be in preventing mode errors. Presumably feedback which is sustained is more salient than transient feedback (with the qualification that sustained feedback may become habituated to over time). Feedback which is demanding is presumably more salient than feedback which is avoidable.

This experiment compares the effectiveness of two kinds of sustained, action-independent feedback in the context of a simple text editing task. Visual feedback was delivered by changing the colour of the screen, while kinesthetic feedback was delivered through the use of a foot pedal to change modes. The prediction was that since kinesthetic feedback is inherently more demanding than visual feedback, it would be a more effective way of preventing mode errors.

2. METHOD

2.1 Subjects

Twelve expert and twelve novice subjects were recruited from the University of Toronto and paid for their participation. An *expert* subject had extensive experience in using *vi*, a Unix-based text editing system. A *novice* subject was one who had never used *vi*, but had experience in using a computer mouse. Eleven of the experts and seven of the novices were touch typists.

2.2 Tasks

The primary task consisted of navigating through and inserting text into a pre-existing document on a Sun workstation. Subjects were instructed to insert the string "errorerror"2 following any word in the document that was printed all in capital letters. They were instructed to complete this task as quickly as possible, only correcting typing errors if they detected them within a word before leaving insert mode. Each block of text contained approximately 190 words and a total of 75 capitalized words.

A simulated *vi* text editor was created in which only a small subset of the commands were available. In order to navigate, the keys h, j, k, and l moved the cursor left, up, down, and right, respectively. In addition, the space bar was available to move the cursor right. For keyboard conditions, in order to insert text, subjects were instructed to position the cursor over the point at which the word was to be inserted, and to press the 'i' key. Once in "insert mode", the text could then be entered. After typing the text to be inserted, the escape key returned the user to "navigation mode". For foot pedal conditions, inserting text was accomplished by positioning the cursor over the insertion point, depressing the foot pedal, and keeping the pedal depressed while typing the text. Releasing the foot pedal returned the subject to navigation mode.

In addition to the primary task, subjects were also required to perform a concurrent distractor task on a Macintosh computer positioned adjacent to the Sun workstation. Thirty seconds after the editing task was begun, after some random interval of time, beeps from the Macintosh signalled the presentation of a digit between 1 and 6 on its screen. Below the digit, 6 buttons numbered 1 to 6 appeared in a random order. The subjects' task was to use the Macintosh mouse to click on the button corresponding to the presented digit. Subjects were instructed to service this distractor task as quickly as possible. In order to encourage them to do so, the beeping would increase in frequency as time passed. The intervals between digit presentation were distributed according to a uniform distribution with an average interval between digits of 4.5 seconds and a range of 3 to 6 seconds.

2.3 Design and Procedure

Each subject performed in each of the four conditions depicted in Figure 1. *Insertion method* refers to the method by which insert mode was entered and exited. Keyboard insertion means using the 'i' and 'escape' keys, while foot pedal insertion means holding down the foot pedal to insert text. In the *visual feedback* conditions, while in insert mode, the screen changed from white to pink. The order of the conditions for each subject was counterbalanced according to a digram-balanced Latin square.

All subjects were given a practice run on the editing task using the keyboard insertion method immediately prior to performing the first keyboard condition, as well as a different practice run using the foot pedal insertion method immedi-

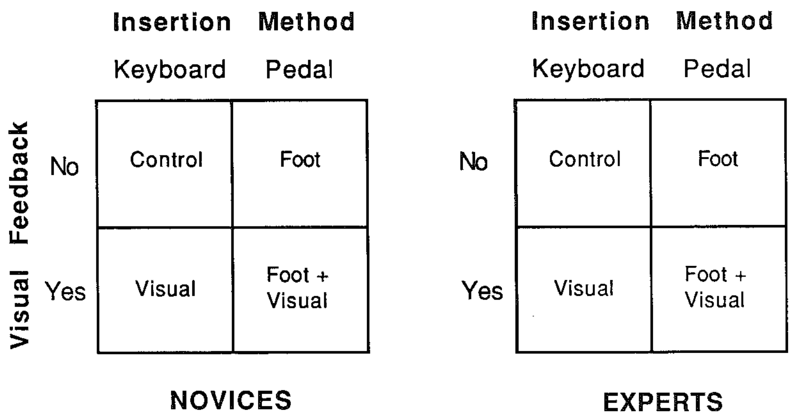

Figure 1. *Schematic diagram of the experimental design. Insertion method refers to the method of switching to insert mode, while visual feedback refers to the presence or absence of a pink screen colour while in insert mode.*

ately preceding the first foot pedal condition. Each practice run consisted of 28 insertions into a pre-existing block of text.

At the end of the experiment, subjects were asked to rank order the conditions in terms of preference and to provide comments on the comparative usability of each "system" for text editing. The entire experiment lasted approximately an hour for expert subjects and an hour and a half for novices including a five to ten minute break halfway through.

3. RESULTS

3.1 Mode Errors

Mode errors were operationally defined in the context of the task as follows, where <NAV> indicates switching to navigation mode and <INS> indicates switching to insert mode, by whichever method:

> A *navigation mode error* was defined as trying to navigate while in insert mode. Operationally, this meant the appearance of h, j, k, l, or spacebar characters while in insert mode and included any unexpected characters which could be construed as aiming errors around those keys, depending on the context. The presence of the "i" command when already in insert mode was also counted as a navigation mode error.
>
> e.g. <INS>errorerrorllk<NAV>...
>
> An *insertion mode error* was defined as trying to insert while in navigation mode. This meant the appearance of any portion of the string "errorerror" while in navigation mode and also included anything which might be an aiming error around those keys. In addition, the appearance of the "escape" character when already in navigation mode was also counted as an insertion mode error since it indicates, at the very least, uncertainty about the state of the system if not the belief that the system is in insert mode.
>
> e.g. llljjjjjlerr<INS>errorerror...

In addition to mode errors, a class of errors we called *synchronization errors* occurred in the foot pedal conditions. A synchronization error looked very similar to a mode error in that a navigation command would sometimes precede the release of the foot pedal, or the letter "e" would sometimes precede depression of the pedal. It was clear, though, that these errors were different from mode errors in that the time between the erroneous keystroke and the response of the pedal was very short (less than 200 msec). Thus these errors arose because of problems in synchronizing the action of the pedal with the keystrokes. Errors with times less than 200 msec. were therefore classified as synchronization errors. If there was any doubt about whether an error was a synchronization error or a mode error, it was classified as a mode error.

The mean number of mode errors of both kinds for novices and experts is shown in Figure 2. Experts made more errors than did novices ($F(1, 11) = 6.23$, $p < .030$). For both the novices and the experts, the pedal method of insertion resulted in significantly fewer mode errors than the keyboard ($F(1, 11) = 16.72$, $p < .002$). In addition, there were significantly fewer mode errors in conditions with visual feedback than those without for both novices and experts ($F(1, 11) = 6.65$, $p < .026$).

Finally, there was a significant interaction present between insertion method and visual feedback ($F(1, 11) = 6.77$, $p < .025$). In order to understand the source of this better, separate analyses were run on the expert and novice groups. The result was a significant insertion method by visual feedback interaction for experts ($F(1, 11) = 9.34$, $p < .011$) but not for novices. This indicates that for experts, while visual feedback was effective in reducing mode errors when the method of insertion was the keyboard, visual feedback was redundant in the case of the foot pedal.

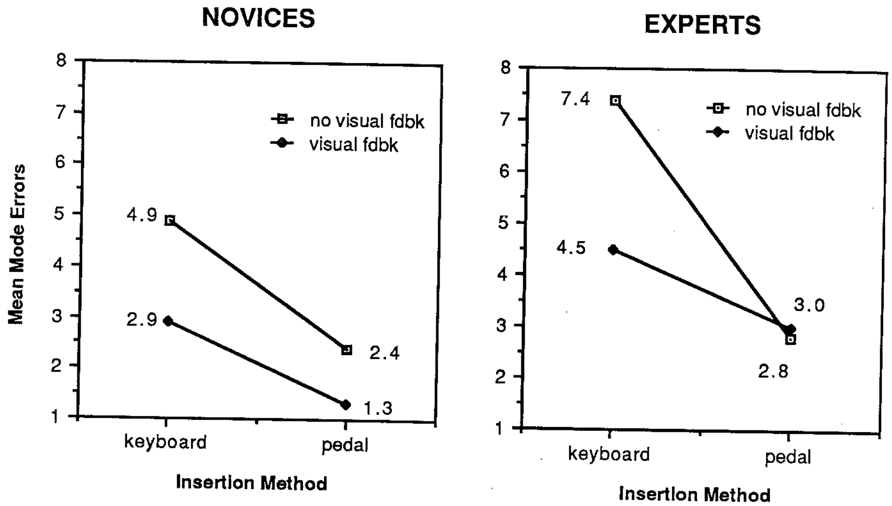

Figure 2. *Mean number of mode errors for novices and experts plotted as a function of method of insertion (keyboard versus foot pedal) and visual feedback (present versus absent).*

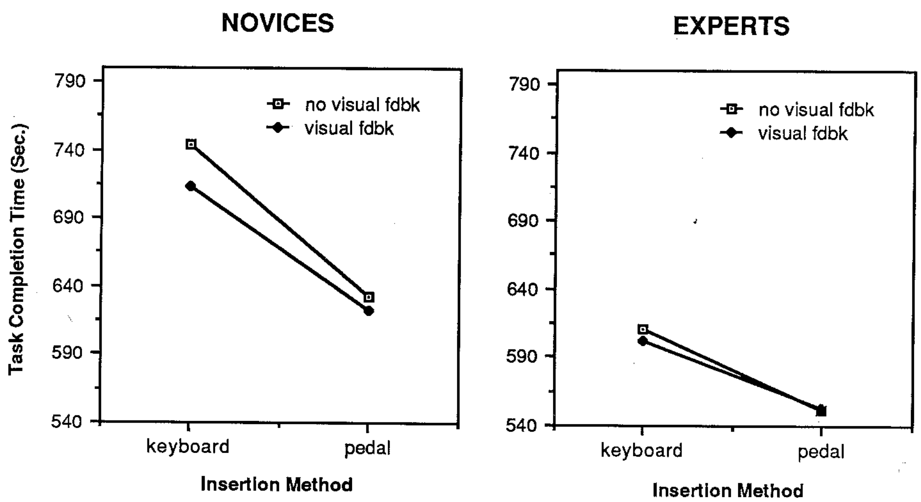

Figure 3. *Mean task completion times for novices and experts plotted as a function of method of insertion (keyboard versus foot pedal) and visual feedback (present versus absent).*

3.2 Task Completion Time

The total time to complete the task in each condition is shown in Figure 3. Experts were significantly faster than novices ($F(1, 11) = 7.35, p < .020$). The only other significant result was a main effect of insertion method, with the foot pedal being faster than the keyboard ($F(1, 11) = 18.42, p < .001$).

3.3 Effects of Switching Between Tasks

Resume time was defined as the amount of time required to make the first keystroke in the editing task after servicing the distractor task. This was taken to be a measure of confusion about the mode in the editing task. The means are shown in Figure 4.

The pedal resulted in a significantly faster mean resume time than the keyboard ($F(1, 11) = 9.41, p < .011$). There were no significant effects of visual feedback, no differences between novices and experts, and no interactions found.

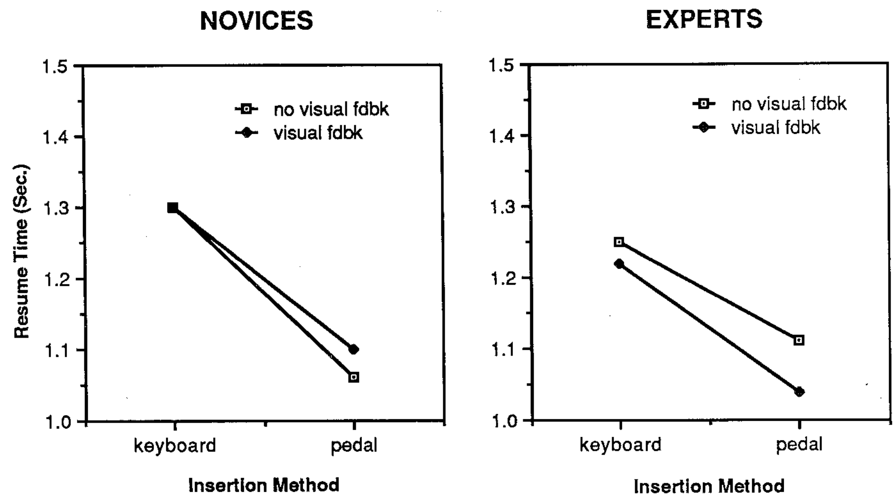

Figure 4. *Mean resume time for novices and experts plotted as a function of method of insertion (keyboard versus foot pedal) and visual feedback (present versus absent).*

Service time for the distractor task was also examined. This was defined as the time between the occurrence of an audio interruption by the distractor task and the mouse click cancelling the number on the Macintosh screen. There were no significant differences found between conditions or between novices and experts.

3.4 Ranking Data

At the end of the experiment subjects were asked to imagine that each condition represented a system that they might use to do text editing on a daily basis and to rank order each of the four "systems" according to their preference. It was clear that subjects fell into three main categories: those who preferred the foot pedal, those who preferred the keyboard, and those who preferred the "systems" with visual feedback, regardless of insertion method. Operationally, they were classified as pedal-oriented, keyboard-oriented, or visual-oriented according to which conditions they chose as their first and second preferences versus their third and fourth choices.

One expert and one novice failed to complete the ranking task properly and could not be classified. Of the remaining eleven experts, five were keyboard-oriented, five were pedal-oriented, and one was visual-oriented. Of the eleven novices, eight preferred the foot pedal systems, two preferred the visual feedback systems, and one was unclassifiable. This last subject preferred either visual feedback, or the foot pedal, but not both.

4. DISCUSSION

This experiment has shown the effectiveness of both visual and kinesthetic feedback in preventing the occurrence of mode errors regardless of whether or not the subjects were experienced users of a system with no explicit mode indicator. Thus, even though many of the expert subjects commented that they were used to keeping track of the mode "in their head", feedback significantly reduced their mode errors nonetheless.

4.1 Kinesthetic versus Visual Feedback

Not all kinds of feedback are equal, however. The results make a particularly strong case for feedback delivered via the foot pedal as opposed to a visual mode indicator. Both visual and pedal feedback reduced mode errors. In the case of experts, though, visual feedback was redundant when pedal feedback was present. This is somewhat surprising given that all but one of the experts were touch typists and frequently monitored the screen. Conversely the beneficial effect of visual feedback for the novices was also surprising given that five of the twelve were not touch typists and constantly monitored the keyboard. One might therefore expect that visual cues would be less effective for this group. What may explain these results is that we frequently observed novices making deliberate visual checks to ascertain the mode when returning from the distractor task. It could be that experts were more likely to be looking at the screen but not necessarily for the purpose of making a visual check on the mode.

Resume time is perhaps a more sensitive measure than mode errors with which to compare kinesthetic to visual feedback. It seems reasonable to assume that the amount of time required to resume the editing task in part reflected decision time during which subjects were attempting to diagnose the state of the system. Such cognitive processes are effortful and increase the mental workload of the task, especially since they are likely to involve short term memory. Any differences in resume time among conditions must reflect a difference in cognitive operations since there are no differences in the physical actions required to switch tasks.

Use of a foot pedal led to a significantly faster resume time than the keyboard while the presence of visual feedback made no difference. Further, these results are independent of level of skill, since novices and experts both benefitted from the foot pedal and not from visual feedback. On the basis of this result, we claim that pedal feedback reduces the cognitive load of the system, at least with respect to confusion about system mode. Visual feedback does not achieve this effect.

Why might the foot pedal be a better way of reducing the cognitive load imposed by confusion about modes? There are at least three possible explanations:

1. Subjects were simply not monitoring the visual feedback. That is, subjects could simply forget to monitor the screen and therefore fail to benefit from visual cues. Visual feedback is *avoidable* and therefore is not as effective.

2. Information delivered through the visual channel is simply not as *salient* as information delivered kinesthetically. This may be the case even though the visual cues in this study involved changing the entire screen area pink. If this is true, this has important implications for systems which rely on more subtle visual cues such as changing the shape of the cursor or the colour of the menu bar.

3. Visual feedback *competes* with the visual nature of the editing task. In other words, a user who is text editing has as his or her main goal the task of searching the screen or monitoring the outcome of their keystrokes. The colour of the screen may therefore compete for attentional resources required for this visual task. Thus it may be that using a different "channel" for indicating mode is more effective since it does not compete with task-specific resources.

These issues cannot be directly addressed in the context of this experiment, but they are good candidates for further research.

4.2 Other Issues of Usability

What about other aspects of the usability of the different systems? One clear difference was the faster speed with which the task was completed for foot pedal versus keyboard conditions. This was true not only for novices but also for experts (most of whom had many years of experience with *vi*). The fact that keyboard insertion caused more mode errors and therefore may have incurred more cost in terms of error recovery time, probably contributed to this difference.

A more fundamental difference between keyboard and pedal was that assigning the mode changing task to the foot pedal meant that this could be accomplished without interfering with the other tasks of navigating and typing. Both subjects who could and could not touch type commented that having to alternate between "i" and "escape" and the navigation keys meant having to constantly re-position the fingers on the keyboard. Many of them felt that this led to more errors in typing. Many of them also said that using "i" and "escape" meant they had to spend more time searching the keyboard, which made the task more effortful.Whatever the contributing factors, subjects (both novices and experts) commented that they liked the increased speed with which they could edit using the foot pedal.

There were different problems associated with the foot pedal. Most notably, from time to time there were synchronization errors where subjects would either depress or release the pedal a fraction of a second too early or too late. These were fairly infrequent though, averaging less than one error per subject during the entire experiment. In addition, some subjects commented that they thought that eventually their foot would become tired. The ergonomics of the design of a foot pedal would therefore have to be an important consideration.

Finally, we had expected that there would be some differences in service time and in "chunking" behaviour across conditions. Chunking behaviour refers to the tendency to finish one sub-task before attending to another (see Buxton, 1986). In this case, chunking was defined as the tendency to delay servicing the distractor task until completion of a sub-task within the editing task. For example, subjects had a strong tendency to complete navigation to the next word, or to complete typing of the inserted word before attending to the distractor. They did this even though they were instructed to service the distractor task as quickly as possible. We predicted that with improved feedback subjects might feel secure enough to interrupt their primary task (text editing) mid-stream, in order to service the distractor. This was not the case, however, and perhaps speaks to the strength of the tendency to chunk in all conditions.

4.3 Experts versus Novices

There was some question as to whether the experts in this study were truly "experts" since subjects could use only a restricted set of commands in the editing task. Differences between experts and novices suggest that they were in fact drawn from distinct populations.

First, experts completed the task much faster than novices in *all* conditions. Note that this was the case even though they were as naive as novices with regard to the foot pedal. This suggests that experts had no trouble integrating the new device with their previously established skills in *vi*.

Second, experts exhibited a different pattern of behaviour with regard to mode errors. Not only did experts make more mode errors over all, but they also did not benefit from visual feedback in combination with the foot pedal. This might be explained by the fact that for experts, there is less overhead involved in correcting errors. Users of *vi* make errors all the time, and are highly skilled at recovering from them. The cost of an error for an expert is thus considerably lower than the cost of an error for a novice and may explain why experts made more mode errors overall. This increased cautiousness on the part of novices may also account for why they benefitted from visual feedback given that they were already receiving feedback from the foot pedal.

Increased cautiousness on the part of novices might cause them to exploit every available cue in an effort to avoid errors.

5. CONCLUSION

The research reported brings us to three main conclusions:

- That with appropriate design, a common class of error can be significantly reduced for both novices and experts.
- That the modality (sensory channel) used for feedback is an important design consideration.
- That designing to reduce errors can also lead to other improvements in system usability including faster performance times and lower cognitive load.

As the complexity and functionality of systems grow, we must learn to anticipate the errors users will make and to design interfaces to minimize their occurrence. In order to cope with this growing responsibility, we feel strongly that interface design will be served well by looking beyond the traditional "mouse-keyboard-display" configuration and investigating other channels and modalities of interaction. We believe that the work of Monk (1986) and the research reported in this paper support this view, and hope that it will stimulate additional research and activity in this direction.

ACKNOWLEDGEMENTS

The members of the Input Research Group at the University of Toronto provided the forum for the design and execution of this project. We especially thank Scott MacKenzie for his comments and assistance. In addition, we are grateful to our colleagues at the University of Toronto: Daniel Read and Ian Spence of the Department of Psychology for their advice in statistical matters, and Alison Lee and George Drettakis of the Dynamic Graphics Project for implementation advice. We also gratefully acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada, Digital Equipment Corporation, and Xerox PARC.

NOTES

[1]We define "feedback" in this context as information about system state received through any of the human sensory modalities.

[2]The string 'errorerror' was chosen so that mode errors consisting of attempts to insert these characters in navigation mode could be clearly distinguished from navigation commands or aiming errors around the navigation keys, and similarly for errors in attempting to navigate in insertion mode.

REFERENCES

Buxton, W. (1986). Chunking and phrasing and the design of human-computer dialogues. In H. J. Kugler (Ed.) *Information Processing '86*, Proceedings of the IFIP 10th World Computer Congress, Amsterdam: North Holland Publishers, pp. 475-480.

Monk, A. (1986). Mode errors: A user-centred analysis and some preventative measures using keying-contingent sound. *International Journal of Man-Machine Studies*, *24*, pp. 313-327.

Norman, D. A. (1981). Categorization of action slips. *Psychology Review*, *88 (1)*, pp. 1-15.

Reason, J., and Mycielska, K. (1982). *Absent-Minded? The Psychology of Mental Lapses and Everyday Errors*. Englewood Cliffs, N. J.: Prentice-Hall, Inc.

Sellen, A. J. (1990). *Human Error Detection Mechanisms*. Doctoral dissertation in progress. University of California, San Diego, La Jolla, CA.

Tesler, L. (1981). The Smalltalk environment. *Byte*, August, pp. 90-147.