

# Effect of artificial voice feedback delay in typing and its implications for the blind user

JOSEPH R. DiLORENZO

*Vision Research Laboratory, The Lighthouse, New York, New York  
and Pace University, Pleasantville, New York*

and

ARIES ARDITI

*Vision Research Laboratory, The Lighthouse, New York, New York  
and New York University, New York, New York*

We examined the effect of voice-feedback delay on learning to type nonsense strings on a reduced computer terminal keyboard. The results suggest that long delays (1,642 msec) affect learning of the keyboard before mastery, whereas shorter delays (841 msec) do not. With mastery, subjects learn to ignore the disruptive effects of long feedback delays. Feedback delay may nevertheless be an important human-factors issue for blind consumers who are learning to type with a voice-based system, and for hardware and software designers of systems for the blind and visually impaired.

In recent years, artificial voice output systems have become the method of choice by which many individuals with severe visual impairment or blindness access computers. In such configurations, the voice device presents aurally to the blind user what would be visible on a CRT screen to the sighted individual. Ideally, a system would make accessible to the blind user all software that is available to the sighted. Human factors and other design issues associated with voice systems are discussed by Foulke (1987) and by Arditi and Gillman (1986). A general area of concern is how blind users use computer terminal keyboards to enter data or text.

When a blind user enters text into a computer, the voice output system can offer the verification that a sighted user would obtain either from the screen or by attending visually to the keyboard. Accomplished blind users can opt to hear each character or word echoed as it is typed, or may use the voice system to verify typed text. However, the blind user who is less familiar with the keyboard is wise to verify each character as it is typed, both to maintain a relatively low typing error rate and to reinforce the psychological association between the keyboard position and the individual characters. In fact, this character-by-character verification mode is a handy tool for learning to type with the terminal keyboard, even with no prior experience, and is the basis of some recent software packages designed to teach typing to sighted and blind alike.

It may also possibly be of aid to more accomplished typists, providing rapid verification of characters typed.

Some computer and voice system combinations, however, may have a significant delay between the time a key is pressed and the voice feedback, because of inability of the hardware and software to support rapid communications rates between the computer and the device, and because of the processing time needed for text-to-speech conversion. Indeed, this delay may be an important consideration in the design of voice-based microcomputers for the blind and visually impaired, and in the consumer's choice of such systems.

We conducted a short experiment in which we examined the effect of voice-feedback delay on learning to type on a reduced keyboard with only the voice output system providing feedback.

## METHOD

### Subjects

Twelve adult volunteers served as subjects. All were normally sighted.

### Apparatus

The experiment was conducted on an IBM PC XT microcomputer equipped with a Telesensory Systems Professional VERT voice synthesizer system. The VERT is normally used with PC VERT software that provides a broad range of screen review and editing capabilities; however, in order to control the feedback delay, we simply sent the individual characters typed to the VERT through a serial communications port. Our software obtained each character as it was typed at the keyboard and placed it on a queue, releasing it to the VERT a mini-

---

The authors extend thanks to Kenneth Knoblauch, Jeffrey Jackson, and Janet Szlyk for their helpful comments and suggestions on this work. The work was partially supported by NIH Grant AG06551 to Aries Arditi, and by an equipment grant from the Hyde and Watson Foundation. Address correspondence to Aries Arditi, The Lighthouse, 111 East 59th Street, New York, NY 10022.

imum of one delay period later. The VERT is the fastest device we have used: we estimate that the maximum delay imposed by processing time and the serial communications link between the PC and the VERT (at 9600 baud) was less than 5 msec. This represents the delay in our no-delay condition when there were no characters on the queue. Character output time for each character was variable, with characters A and W taking 180 and 290 msec, respectively, for output at the VERT speed setting we used. If other previously typed characters were in the queue, they were sent to the VERT at the rate of one per delay period plus character output time. This represents the maximum rate of speech output in our experiment.

### Stimuli

The subject's task was to learn to type nonsense words, with eyes closed, on a keyboard with new character-position assignments. The new assignments could be learned only by listening to the voice feedback; thus, initial familiarization with the keyboard layout had to be accomplished by trial and error. The same eight keys of the keyboard were used during the entire experiment: the subject's left hand used the A, S, D, and F keys and the right hand used the H, J, K, and L keys of the standard QWERTY keyboard. For each delay condition, the keyboard was rescrambled so that each of the eight keys was associated with a new, randomly chosen character. In order to avoid multiple errors arising from a single misplacement of the hand, the subjects were told explicitly that all the characters to be typed could be found on one of these eight keys, and that no hand movement was necessary.

Each nonsense word was a four-character string, with two characters chosen from the left-hand positions and two from the right-hand positions. Each successive word to be learned by a subject contained three of the same characters as the preceding word, and a new character from the set of eight defining the reduced keyboard made up the fourth character. There were three such reduced keyboards. It is easy for the uninitiated to confuse certain characters spoken by the VERT, such as P and T or V and B; therefore, care was taken to ensure that no keyboard contained both members of such a confusable pair.

### Procedure

Each subject's learning performance was tested on three delay conditions—no delay (< 5 msec), 841 msec, and 1,642 msec—in one of six possible delay-condition orders. Prior to testing with each reduced keyboard, each subject was familiarized with the particular pronunciation of the voice synthesizer to rule out recognition difficulty as a source of error. First, the subject's ability to recognize and identify each of the characters spoken by the artificial voice was tested with feedback from the experimenter. The experimenter typed keys from the reduced keyboard; if the subject identified the character incorrectly, the subject was told the correct character. This procedure continued until the experimenter was satisfied that the sub-

ject was able to identify perfectly the eight characters of the subset. Prior to testing with each character string, the subject repeated the character string aloud many times, until the experimenter was satisfied that failure to remember the string could also be ruled out as a source of typing errors.

Finally, the typing test was given, in which the subject's task was to type the character string correctly to a criterion of three consecutive times. The subject was told that speed, accuracy, and total number of characters typed were being assessed, and to type as accurately and quickly as possible while minimizing the total number of keystrokes. The experimenter measured with a stopwatch the time from the striking of the first key to the last keystroke of the third correct repetition. Additionally, the number of keystrokes to reach criterion was recorded.

The orders of delay condition and keyboard subset were independently randomized and counterbalanced for all subjects, with 2 subjects each being tested on each of the six possible delay-condition orders, and 2 subjects each tested on each of the six keyboard subset orders. The nonsense character-string order was randomized for each subject as well.

## RESULTS AND DISCUSSION

### Number of Characters Typed

Number of characters typed was a measure of how quickly each subject learned to type each nonsense string in informational units, rather than in units of time. Inspection of the individual cells showed variances to be quadratically related to the means. Therefore, after Winer (1971), the data were logarithmically transformed prior to subsequent analysis of variance (ANOVA), to homogenize the variances. Overall, the subject did better on later than on earlier trials, reflecting their learning of the reduced keyboard [ $F(3,33) = 6.824, p < .005$ ]. This learning was differentially affected depending on delay condition, as documented by a trial  $\times$  delay-condition interaction [ $F(6,66) = 2.256, p < .05$ ]. There was no significant main effect of delay ( $p > .18$ ).

Figure 1 plots the geometric means for each delay condition. Within the no-delay and 841-msec-delay conditions, there is little evidence for improvement in performance over the four trials. This is surprising, since only the nonsense string changed over these trials, while the reduced keyboard was kept constant. On the other hand, learning a keyboard, and not simply a nonsense string, probably involves rote learning of individual strings that, over many more trials than the four in our experiment, would generalize into learning of a keyboard map.

A more interesting feature of Figure 1 in the present context is the poorer performance of our subjects on the early than on the later trials of the longest delay condition. Apparently, our subjects quickly learned to adapt to long feedback delays, so that these delays soon had no effect on total number of characters required to perform the task to criterion.

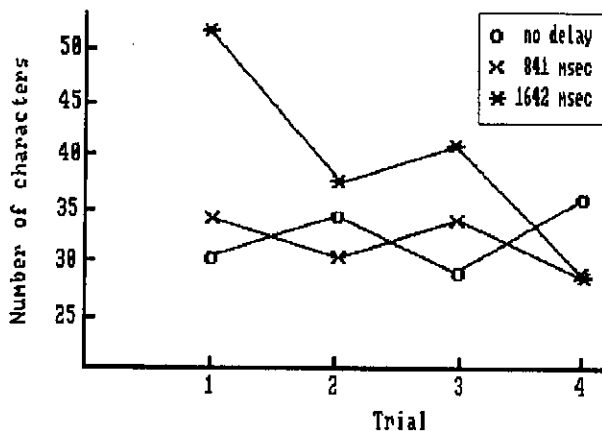


Figure 1. Geometric mean of the number of characters typed to criterion performance. Average standard error about each point is 17% of the mean.

### Time to Type to Criterion

Time to type to criterion, of course, measured how quickly the subjects learned each reduced keyboard. Given the actual delay imposed after each character is typed, one expects a priori that longer delays will produce longer times to learn the keyboard. This is simply because the typist must wait at least one delay period to obtain the feedback upon which the learning is based. We sought originally to subtract the total amount of delay imposed upon the subject (i.e., the number of characters typed  $\times$  the delay duration) from the times recorded, to obtain a measure of learning time that could be compared across delay conditions and that would not reflect the delay imposed by the feedback. Our subjects, however, behaved in an unexpected way that was interesting in its own right, but which made it unfeasible for us to perform the planned adjustment for feedback delay.

As the subjects learned the positions of the characters required for each nonsense string, they began to ignore the feedback in the longer delay conditions. Thus, after developing a certain level of confidence in their knowledge of character positions, in the interest of speed, they usually typed their three required repetitions without waiting for feedback. The times we recorded reflected the total duration of typing, but these times did not include a delay period for the characters that were typed ahead at the end of a trial. The fact that nearly every subject behaved this way demonstrates the irrelevance of the feedback once mastery of the keyboard was achieved.

The results we obtained are not unlike those seen in a variety of other production tasks. For example, speaking in the presence of delayed auditory feedback induces stuttering and mispronunciation (Lee, 1951). Siegel, Fehst, Garber, and Pick (1980) studied this phenomenon in adults and in children 5 and 8 years of age, using several delays ranging from 0 to 625 msec. Although their measures of the effects of feedback tended to asymptote above some delay ranging from 250 to 500 msec, they found that the feedback was less disruptive with age and, presumably, language mastery. Although our subjects hardly had a

chance to master even the reduced keyboards we used, they too showed greater effects in the early trials, in which reliance on the feedback was of necessity.

Delayed sensory feedback has also been found to be disruptive to performance of complex sensorimotor skills such as manual tracking (Kao, 1977) and visual and auditory tracking (Wargo, 1967), although the relationship between mastery and the effects of feedback is not clearly established (e.g., Smith, 1966).

Our study confirms the general finding that as skill increases, external feedback that was useful in learning may become less relevant and may be ignored. Such feedback may be perceived as an annoyance, but seems to have little effect on typing performance. Other variables contributing to intelligibility and comprehension of synthesized speech (Nusbaum & Pisoni, 1985; Slowiaczek & Nusbaum, 1985) may, however, be important in combination with delayed feedback.

Since there was little effect of delayed feedback even in the 841-msec-delay condition, we did not find evidence to suggest that designers of computer systems for the blind need meet particularly stringent feedback-delay standards. Virtually all systems currently on the market already have shorter feedback delays than the shortest delay we used, even when connected with relatively slow communications links.

An interesting question for further study is whether other forms of feedback, specifically designed for the skilled typist, may be more useful than voice feedback and may actually enhance skilled typists' efficiency. For example, with increased speech speeds, if blind typists can learn to identify rapidly spoken characters, they may be able to obtain sufficiently rapid verification to make feedback relevant rather than disruptive. If sufficiently rapid speech is unintelligible, musical tones associated with each character on the keyboard might prove viable as an alternative auditory feedback medium.

### REFERENCES

- ARDITI, A., & GILLMAN, A. E. (1986). Computing for the blind user. *BYTE*, 11(3), 199-208.
- FOULKE, E. (1987). Overcoming the limitations of speech synthesizers. *Behavior Research Methods, Instruments, & Computers*, 19, 291-294.
- KAO, H. S. R. (1977). Effects of delay of feedback and intermittency of feedback in compensatory tracking. *Perceptual & Motor Skills*, 44, 1079-1085.
- LEE, B. S. (1951). Artificial stutter. *Journal of Speech & Hearing Disability*, 16, 53-55.
- NUSBAUM, H. C., & PISONI, D. B. (1985). Constraints on the perception of B synthetic speech generated by rule. *Behavior Research Methods, Instruments, & Computers*, 17, 235-242.
- SIEGEL, G. M., FEHST, C. A., GARBER, S. R., & PICK, H. L., JR. (1980). Delayed auditory feedback with children. *Journal of Speech & Hearing Research*, 23, 802-813.
- SLOWIACZEK, L. M., & NUSBAUM, H. C. (1985). Effects of speech rate and pitch contour on perception of synthetic speech. *Human Factors*, 27, 701-712.
- SMITH, K. U. (1966). Cybernetic theory and analysis of learning. In E. Bilodeau (Ed.), *Acquisition of skill* (pp. 425-480). New York: Academic Press.
- WARGO, M. J. (1967). Delayed sensory feedback in visual and auditory tracking. *Perceptual & Motor Skills*, 24, 55-62.
- WINER, B. J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.