THE EFFECT OF FEEDBACK ON COMPOSITION RATE USING A SIMULATED LISTENING TYPEWRITER

K.E.P. Carter; S. Cookson; A.F. Newell; J.L. Arnott; R. Dye.
Microcomputer Centre, University of Dundee, Scotland.

ABSTRACT
Simulation Experiments have been conducted on text composition by voice input. Two experiments were conducted to investigate whether composition rate varied with different feedback strategies. No significant variation in composition rate could be attributed to the various feedback strategies used.

INTRODUCTION
Feedback has been defined "as any information provided by an automatic speech recognition system the user can utilise to determine whether an utterance has been recognised correctly and/or whether the required action will be carried out by the system" (4). Feedback is required in order that recognizer and human errors can be detected. Feedback in the "listening typewriter" application could be presented in a number of ways. Word-by-word feedback has been proposed, which would in addition force users to speak in an isolated word fashion (5). Whereas others have argued that, with a connected speech recognizer, feedback should be provided at the sentence level (6). It is generally acknowledged that such information is important. It is unclear however, how feedback should be presented so that it does not interfere with the user's task.

Previous experiments have found that composition rates by users of a simulated "listening typewriter" were consistently low with average composition rates not exceeding 12 wpm (1). It was hypothesized that the psychological effect of seeing speech as it is dictated may have hindered their dictation ability. It is with regard to this problem that the following two experiments described were conducted.

EXPERIMENT 1

METHOD

DESIGN
The experiment had a within subjects design. The independent variable was the type of feedback that the subject was presented with in the process of text generation. These were: (A) Normal Feedback; the text on the screen was automatically and continuously updated on the screen as fast as possible; (B) Syntactic Marker; the text displayed on the screen was updated only whenever the subjects inserted a syntactic unit (e.g. comma, fullstop, etc.) in the course of text composition; (C) Feedback on request; the text was updated on the screen whenever the subject specifically requested feedback by saying the keyword "DISPLAY"; (D) Mixed Feedback; the text was automatically updated on the screen whenever the subject inserted a syntactic marker (e.g. comma, fullstop, etc.) and whenever the subject specifically requested feedback by saying the keyword "DISPLAY".

The dependent variable was the subjects' composition rate, that is, the total number of words in the final document divided by the time to produce it. Measures of composition efficiency were also made; that is, the total number of words in the final document divided by the total number of words spoken. It was expected that the nature of the feedback would influence the subjects' composition rate and efficiency.

SUBJECTS
These were 10 male postgraduate students at the University of Dundee. The subjects varied in age from 22 to 28 years with a mean of 24.7 years. Subjects were recruited through advertisements and were paid for participating.

APPARATUS AND MATERIALS
A large vocabulary "listening Typewriter" was simulated using a skilled stenotypist and a commercially available Palantype Shorthand Machine (2), (3). Words spoken by the subject were picked up by a free standing microphone which was linked to the stenotypist's headphones. All input was vocal with editing commands spoken within a restricted natural language syntax. The stenotypist transcribed the subject's utterances which, after being converted to standard orthography appeared on a workstation display mounted immediately in front of the subject. The subjects were required to compose a number of letters based on outlines prepared by the experimenter.

PROCEDURE
The experiment was conducted in an "experimental office" over 5 consecutive days. In the first session the subjects learnt the restricted syntax which was necessary to operate the editor with which they would edit and format documents. In the following three sessions the subjects were trained to compose and edit documents with either (A) Normal Feedback, (B) Syntactic Marker Feedback or (C) Feedback on request operating. (These feedback levels did not operate when subjects switched into an editing mode; when
editing commands were given the system reverted to presenting feedback as fast as the computer could process the information.

In the final session the subjects composed one practice document and one letter with each of the following feedback levels: (A) Normal Feedback, (B) Syntactic Marker Feedback and (C) Feedback on Request. The order of testing was randomized. At the conclusion of the experiment the subjects were required to compose a practice document and a letter using the (D) mixed feedback condition to which they had not previously been exposed. All the subjects were aware of the simulation.

RESULTS

Results were obtained from the experiment in terms of how the task was performed. The average speech rates, composition rates and composition efficiency are shown in Table 1.

<table>
<thead>
<tr>
<th>Feedback Condition</th>
<th>Speech Rate (WPM)</th>
<th>Composition Rate (WPM)</th>
<th>Composition Efficiency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>27.6</td>
<td>7.9</td>
<td>29.6</td>
</tr>
<tr>
<td>Syntactic M.</td>
<td>32.2</td>
<td>9.1</td>
<td>28.2</td>
</tr>
<tr>
<td>Request</td>
<td>31.4</td>
<td>7.1</td>
<td>22.5</td>
</tr>
<tr>
<td>Mixed</td>
<td>32.1</td>
<td>5.9</td>
<td>18.9</td>
</tr>
</tbody>
</table>

Analyses of Variance and T tests were carried out on the feedback condition. These showed reliable differences in speech rates and composition rates between the different feedback strategies. These are not reported here for reasons which are explained in the following discussion.

DISCUSSION

The analyses of variance and T tests seemed to provide partial confirmation for the hypothesis that the type of feedback subjects are given in the process of text generation has some effect on composition rate. However, the results were open to an alternative interpretation. The final documents produced by some of the subjects included extensive formatting of the text and this would have biased the composition rate. That is, a materials effect might have produced the results. This effect is also indicated by the reduced efficiency for some conditions. Furthermore, the subjects had not received very much practice with mixed feedback condition and it was always presented at the end of the session. This might account for results in Experiment 1. In addition we wished to investigate longer term effects of familiarity to a listening typewriter on composition rate thus a second experiment was designed to overcome these problems, and was timetabled to occur approximately one month after the first experiment.

EXPERIMENT 2

METHOD

DESIGN AND MATERIALS

The experiment had a within subjects design. The materials were four new outline letter drafts.

SUBJECTS

These were 10 male postgraduate students who had taken part in Experiment 1.

PROCEDURE AND APPARATUS

The apparatus was the same as that used in Experiment 1. The experiment was conducted over two experimental sessions. In session 1 the subjects refamiliarized themselves with the editing syntax and practiced composing text with each of the feedback conditions. In session 2 the subjects composed one practice document and one letter with each of the four feedback conditions. The materials and order of conditions was randomized for each subject.

RESULTS

The average speech, composition and efficiency rates produced during the course of the experiment are shown in Table 2. A one way analysis of variance (ANOVA) was carried out on the Speech Rate scores. The differences between the feedback conditions did not approach significance, (F=0.898; df=3,27; P>0.05). A one way ANOVA was also carried out on the composition rate data. It was found that the feedback conditions did not differ reliably, (F=0.302; df=3,27; P>0.05).

The efficiency of dictation was low; on average only 32% of the words dictated appeared in the final document.

DISCUSSION/CONCLUSION

The results of Experiment 1 suggested that the type of feedback users received in text generation had some effect on composition rates. However, an alternative explanation was considered, in that the results could be attributed to confounding variables. The results of Experiment 2 were consistent with this interpretation as no reliable differences in composition rates were found in Experiment 2.

In both experiments, composition rates were found not to exceed averages of 10 wpm which is consistent with previous research by this group. The average speech rates were also found not to
be faster than 33 wpm. The efficiency of dictation generally was low although the subjects in Experiment 2 had shown some improvement. Overall the results are consistent with our earlier findings that voice input for text composition has proved to be slow and not very efficient. The changes in feedback with which we experimented did not greatly affect composition rate and it may be concluded that increasing the speed with which users operate a listening typewriter is a non-trivial task, and changes to the ways in which feedback are offered is unlikely to have a major effect on this.

REFERENCES


