Voice dictation in the secondary school classroom

Michael F. McTear[1]
Eamonn A. O'Hare[2]

[1]University of Ulster
[2]St Mark's High School, Warrenpoint, N. Ireland

ABSTRACT

This paper reports on an exploratory study in which a group of second year secondary school pupils with reading ages ranging from 8.3 to 12.9 performed a set of tasks using the IBM VoiceType dictation package in order to determine the benefits of voice dictation for classroom use. The results showed that pupils with varying reading ages could dictate at comparable speeds and often with similar degrees of accuracy. Homophones were almost never a source of error in the texts produced with voice dictation, as compared with the children's handwritten texts. The implications of these findings for the use of dictation software in the classroom and for further studies of the potential of voice dictation for improving children's spelling and composition skills are discussed.

1. INTRODUCTION

This study is concerned with the use of dictation software in an educational context. A group of second year secondary school pupils with reading ages ranging from 8.3 to 12.9 performed a set of tasks using the IBM VoiceType 3.0 dictation package to determine if the dictation software enabled the children to input text more quickly than when using the keyboard, and whether the resulting texts were more accurate in terms of spelling. A further aspect of the study was to investigate the usability of the dictation software in a school context. The paper is structured as follows. The next section provides a brief review of previous studies involving the use of dictation software with school-aged children. Section 3 describes the study and presents the results, which are discussed in section 4 along with some more general issues concerning usability. The concluding section reviews the findings in terms of their educational implications and suggests some directions for future research.

2. RELATED WORK

Most studies of the use of voice dictation have been carried out by researchers in the Human Factors community who have reported, among other things, that input using voice is faster than typing for users who are not professional typists [1]. There have been few published studies to date that have been concerned with the use of voice dictation software by school-aged children, and most of the information that is currently available is to be found on pages on the World Wide Web (see, for example, [2]). A pilot study conducted by the Devon Local Education Authority in England highlighted the effects of using voice dictation software with secondary dyslexic pupils, including improvements in reading and spelling ages and a general improvement in the standard of written work [3]. The Speaking to Write project, a joint initiative between the Education Development Center, Boston and the Communication Enhancement Center at the Children's Hospital, Boston, is concerned with the use of speech recognition software by secondary students who have significant difficulty with writing due to physical and/or learning disabilities. Although no major results are available as yet from this project, some case studies are presented that illustrate the successful deployment of the software by learning-disabled children [4]. The Trillium Speech-to-text Project is investigating speech recognition for severely learning disabled students with ages ranging from 11 to 18 years. A manual has been produced for the use of the Dragon Dictate voice dictation software with these students that includes recommendations for instructors and parents [5].

3. THE PRESENT STUDY

The system used in the current study was IBM's VoiceType 3.0, which was a state-of-the-art system at the time of the study. The study involved two tasks. The first task investigated the pupils' attempts at producing a pre-prepared text using different means of input. This task was undertaken by a group of 10 second year, secondary school children, 3 males and 7 females, whose chronological ages were between 12 and 13 years of age. The selection of the children was based on their reading age at the time of the study and their grading level based on a 3 band ability level grading system used in their secondary school. The second task examined how the dictation software could handle spontaneously produced texts. Five of the pupils who had taken part in the first task were chosen at random for the second task.

3.1. Task 1

The first task consisted of three sub-tasks. The first sub-task involved hand-written dictation in which the experimenter dictated a text to the group of children and the number of spelling errors made by the children was recorded. In the second sub-task the pupils typed the same text into the computer. All the pupils had previously completed a course in keyboard skills. The time taken to type in the text was recorded. The main purpose of this sub-task was to provide a comparison for speed of input between the keyboard and the dictation software. In the third sub-task pupils dictated the text to the computer using the dictation software. Before completing this sub-task, all of the pupils followed the enrolment procedure provided with the dictation software,

which created individual voice models for each pupil and instructed the children on how to dictate to the computer. The pupils were allowed four readings of the text in order to investigate whether accuracy improved, as would be expected as the pupils gained more confidence in their use of the software, and, more crucially, as the software adapted to the pupils' voice patterns. The number of errors was recorded along with the time taken to read in the text. These measures provided comparisons for speed of input with keyboard input and for accuracy of spelling with the hand-written text. The text for the hand-written task and an example of one of the children's first and fourth voice dictated texts are shown in Appendix A.

Table 1 shows the results from Task 1 in terms of the time taken, in minutes and seconds, to type in the text provided and to dictate the text using VoiceType. The dictation times are averaged across the four sessions as there was little variation between sessions for these times.

Name	Reading Age	Typing Time Mins Secs	Voice Type Dictation Average Time Mins Secs
Melissa	8.3	18:00	3:30
Linda	8.3	20:40	3:20
Gary	8.3	41:00	3:30
Gregory	8.6	25:10	3:40
Siobhan	9.3	21:40	3:30
Claire	11.6	10:05	3:10
Caroline	11.6	17:20	3:35
John	12.3	20:40	3:35
Ann Marie	12.3	17:50	3:20
Eimear	12.9	21:40	3:30

Table 1: Times for typing and voice dictation

The results show that the children took roughly 20 minutes to type in the text, except for Gary, who took 41 minutes and Claire, who took only 10 minutes. It is interesting to note that there is only a weak correspondence with reading age, so that a child's reading age would seem to have little bearing on the speed with which they can input a text using a word processor. As far as voice dictation was concerned, the pupils were able to dictate the text in less than 4 minutes, confirming the results of earlier studies that input by dictation is faster than input by the keyboard. The differences were particularly striking for these children whose keyboard experience was likely to be inferior to that of the adults investigated in earlier studies. Thus as far as time is concerned, the use of voice dictation would appear to be an extremely efficient method for text input for school-aged children. Furthermore, it is interesting to note that there is very little difference in each of the pupils' times for VoiceType input compared to the wide range in keyboard input, which suggests that the skill level for dictating is more consistent across the selected pupils, compared to their keyboard skills.

The second aspect of efficiency is accuracy, as there is little to be gained with faster input if a large number of errors are created in the process. The results for accuracy are presented in Table 2. Here comparisons are made between the first sub-task (hand-written dictation) and voice dictation of the same text using VoiceType for each of the four readings. As can be seen from Table 2, the children produced a fairly large number of errors in the hand-written dictation, with fewer errors in general produced by the children who had a higher reading age. The number of errors when using VoiceType tends to decrease with each reading. What is also interesting is that some of the pupils produced good results using VoiceType when comparisons are made with their performance in the hand-written exercise. Gary, who made 33 errors when the experimenter dictated the text to him, produced a text using VoiceType, which by the fourth reading had only 11 recognition errors. On the other hand, some of the children, particularly those with higher reading ages, produced many more errors using VoiceType. It may be that voice dictation is less suited for some children. Further research would be necessary to establish the factors that have a bearing on this unexpected outcome.

Name	Reading Age	Hand-written dictation	Readings			
			1	2	3	4
Melissa	8.3	16	46	16	14	11
Linda	8.3	34	51	21	23	34
Gary	8.3	33	41	10	20	11
Gregory	8.6	20	62	47	43	28
Siobhan	9.3	21	50	69	37	42
Claire	11.6	5	37	19	11	19
Caroline	11.6	18	47	23	13	10
John	12.3	18	31	19	12	7
Ann Marie	12.3	4	89	65	70	51
Eimear	12.9	3	54	19	31	24

Table 2: Number of errors for hand-written dictation and when using VoiceType

3.2 Task 2

One limiting factor of the study as presented so far is that the exercises that the children carried out were rather artificial, involving writing out a text dictated by the experimenter and using voice dictation to read a pre-prepared text into the computer. A more realistic scenario would involve dictating a text spontaneously into the computer. For this reason, in the second task the children were presented with a set of pictures representing a story about a camping trip. A period of ten minutes was allowed for the children to prepare the stories. The children then performed two sub-tasks. In sub-task 1 the children dictated their story spontaneously to the computer using the dictation software. In the second sub-task the children typed in their story using the keyboard. The time taken to compose the stories and the number of spelling errors are shown in Table 3. The final column shows the number of words per minute when the children entered their texts using the keyboard.

As can be seen from Table 3, there is a close correspondence between reading age and the two measures of accuracy and speed. Given the small sample it is not possible to draw firm conclusions from these results. However, it is possible that the ability to produce spontaneous text corresponds more closely to a child's reading age than the exercises carried out in Task 1. Further research would be necessary to substantiate this hypothesis. What is clear, however, from these results is that the children were able to produce texts with a fairly high degree of accuracy and at a much greater speed than when they used the keyboard and mouse for input. Thus the benefits of voice dictation for school children may also apply to the task of free composition.

Name	RA	No. words	Accuracy		Speed	
			Errors	% correct	WPM dictated	WPM typed
Melissa	8.3	161	37	77	51	10
Linda	8.3	166	38	77	52	9
Gary	8.3	184	36	80	58	4
Claire	11.6	204	33	84	66	18
John	12.3	241	31	87	62	9

Table 3: Errors and times for Task 2. RA=reading age,WPM=words per minute.

3.3. Error analysis

In addition to the quantitative analysis presented for the two tasks, a qualitative analysis was made of the types of errors produced by the children in the hand-written and voice dictated texts. This analysis revealed that in the hand-written texts the children produced a large number of spelling errors which included misspellings (e.g. *pidgen* for *pigeon*), word boundary errors (e.g. *a round* for *around, infront* for *in front*), and errors involving consonant doubling (e.g. *bobing* for *bobbing*). There were also several errors involving homophones (e.g. *there/their*, *pail/pale*, *threw/through*, *where/were*), and some involving a failure to capitalise at the beginning of a sentence following a full stop. The errors produced when using the voice dictation software were different. As far as homophone confusion was concerned, there were only 7 errors across all forty dictated texts compared with 27 in the ten hand written texts, due to the use of a word-usage (or language) model that makes it possible for the speech engine to differentiate words that are acoustically identical. Successful capitalisation is easily achieved as the software can be programmed to output capitals in predictable contexts, for example, following a full stop. Other errors that occurred resulted from misrecognition of punctuation commands. For example, *full stop* was sometimes misrecognised as *VoiceType* or *doorstep* and as such reproduced as dictated text.

4. DISCUSSION

Bearing in mind the limitations of this small-scale study, such as the size of the sample and the artificiality of some of the exercises, it can be concluded that voice dictation is potentially a viable tool for the school classroom. All of the pupils were able to input text much more quickly using voice dictation (roughly 50 words per minute) as compared with typing (roughly 10 words per minute). The accuracy achieved was on average 82% for the first task and 81% for task 2. While these rates are much lower than the rates claimed for voice dictation products (90% - 95%), it should be noted that the users of the package were secondary schoolchildren as opposed to professional adult users and would be less accustomed to articulating clearly using discrete speech. Furthermore, the children in this study speak with a strong regional accent, which would initially cause some problems for a speech recognition system trained on standard British English pronunciation. However, it could be expected that accuracy rates would increase with use as the software adapted to the children's voice patterns and as the children became more accustomed to using a voice dictation package. One interesting finding of the experiments was that voice dictation appears to help reduce differences between children that are attributable to reading age. The study showed that pupils with varying reading ages could dictate at comparable speeds and often with similar degrees of accuracy. Thus voice dictation allows the weaker reader (and weaker writer) the opportunity to produce work which is of a high standard of presentation.

Despite these encouraging results, it is worth noting that voice dictation is likely to be beset with teething problems, especially in a school environment. In the present study some difficulties were encountered during the training sessions. Some pupils were unable to read some of the words that were to be used in the training process. The length of time taken to enrol varied from between 45 minutes to just over 2 hours. One major source of difficulty is that the package is designed for adult use and hence little attention has been given to the problems faced by young school pupils. As some of the pupils found difficulty in leaving pauses between words, a 20 minute choral session was arranged to provide the group the opportunity to practise isolated speaking of the text. Each of the pupils was given a copy of the 50 statements to practise reading in isolated speech at home. As the pupils became more experienced, discrete speech almost became a natural method of talking. One advantage of discrete speech is that it forces children to focus on individual words, thus enabling them to see more clearly the relationship between their pronunciation of the word and its graphical representation. It will be interesting to investigate whether more recent software permitting continuous speech input will be less suitable in this respect.

5. CONCLUDING REMARKS

The results from this study are sufficiently encouraging to warrant a larger scale study to substantiate the present findings. With a longitudinal study it would also be possible to investigate the extent to which children learn from the use of the software, in particular, whether the use of the software results in improvements in the children's spelling performance as well as in their ability to produce better written texts. A qualitative analysis of the errors made by the dictation software would also indicate which words were particularly

error-prone and whether the errors were a consequence of the children's regional accents. In summary, the potential for the use of voice dictation in the classroom has yet to be realised. The results of the present study are encouraging and provide a basis for more extensive research in the educational applications of speech technology.

5. REFERENCES

1. Baber, C. and Noyes, J.M. (eds.), *Interactive Speech Technology : Human Factors Issues in the Application of Speech Input/Output to Computers*, Taylor & Francis, London, 1993.
2. Litterick, I. "How dictation systems benefit the education of students with writing difficulties", WWW document: http://www.dyslexic.com/dictben.htm ,1998.
3. Miles, M., Martin, D. and Owen, J., *A Pilot Study into the Effects of Using Voice Dictation Software with Secondary Dyslexic Pupils.* Devon LEA, Feb, 1998.
4. Follansbee, B. "Speaking to Write: Realising the Potential of Speech Recognition for Secondary Students with Disabilities", WWW document: http://www.edc.org/spk2wrt/, 1998.
5. Lubert, J. and Campbell, S. A Manual for Implementing Dragon Dictate. WWW document: http://snow.utoronto.ca/best/manual2.html, 1998.

APPENDIX A

1. = The text dictated to the children.
2. = Melissa (first reading): 46 errors, time: 3 mins.35 secs.
3. = Melissa (fourth reading): 11 errors, time: 3 mins 30 secs.

1. Pat shut the door of the warm kitchen behind her FULLSTOP Outside the air was like ice FULLSTOP It took her breath away for a moment FULLSTOP NEWPARAGRAPH
2. **Tat shot** the door of the warm kitchen behind her. **Aside** the **are** was like ice. It **Tucker for** breath away for a moment.
3. Pat **shot** the door of the warm kitchen behind **power**. Outside the air was like ice. It took her breath away for a moment.

1. All around her the white garden sparkled and shone in the pale winter sunshine FULLSTOP NEWPARAGRAPH
2. All around **for** the White **cabin spiders** and **shown** in the **real** winter sunshine.
3. All around her the white garden sparkled and shone in the pale winter sunshine.

1. Pat blew on her hands to warm them FULLSTOP Then she walked across the frosty grass to the pigeon cage that stood by the back fence FULLSTOP NEWPARAGRAPH
2. Pat **to hurt** hands to warm them **faster** then she **what** across the frosty **grouse** to the pigeon that stood by the back fence **bolt stamp**
3. Pat blew on her hands to warm them. Then she walked across the frosty grass to the pigeon cage that stood by the back fence.

1. She stood in front of the cage COMMA watching the pigeons FULLSTOP NEWPARAGRAPH
2. She stood in front of the cage, watching the pigeons.
3. She stood in front of the cage, watching the pigeons.

1. There were twelve pigeons FULLSTOP Some were grey with long black or white feathers in their wings FULLSTOP The others were brown COMMA speckled with white FULLSTOP NEWPARAGRAPH
2. There were 12 pigeons. Some **work** Gray **whiff** long black or white feathers in **the** wings **VoiceType** the others were **crowned, spiders** with white.
3. There were 12 pigeons **Voicetype** some were grey with long black or white feathers in their wings. The others were brown, speckled with white **VoiceType**

1. They turned their heads from side to side and looked at Pat COMMA first with one dark eye COMMA then with the other FULLSTOP NEWPARAGRAPH
2. **Their demand** their **hits** from side to side and **let out tact**, first **one or less daft Harry**, then **whisk** the other **wars tap**
3. **A torrent** their heads from side to side and looked at Pat, first **whiff** one **black** eye, then with the other.

1. The pale winter sun shone through the wire HYPHEN netting and over the nesting HYPHEN boxes FULLSTOP Some of the birds were sitting on their nests FULLSTOP The others walked up and down the cage COMMA cooing and bobbing their heads FULLSTOP NEWPARAGRAPH
2. The **tear San** shone through the wire-**18** and over the nesting- boxes **VoiceType** some of the birds were sitting on their nests. The others **worked** up and down the cage, **caring** and **dubbing** and **their hits**.
3. The pale winter sun shone through the wire- netting and over **death** nesting **hasten axes.** Some of the birds were sitting on their nests. The others walked up and down the cage, cooing and bobbing their heads.