Available online at www.sciencedirect.com
ScienceDirect
Computers and Composition 25 (2008) 165–181
Computers and Composition
www.elsevier.com/locate/compcom

A Study of Voice-Recognition Software as a Tool for Teacher Response

Thomas Batt [a,*], Sandip Wilson [b]

[a] *Department of Arts and Sciences, Maine Maritime Academy, Castine, ME 04420, USA*
[b] *Department of Education, Husson College, Bangor, ME 04401, USA*

Abstract

Voice-recognition technology (VRT) promises ease of use in responding to student writing, but its impact on writing processes and the quality of teacher commentary is unclear. This article details the results of a study undertaken to evaluate the effectiveness of VRT in responding to student writing. Over the course of two semesters, in response to texts written by students in his first-year college composition course, one of the authors composed 58 end comments, alternating between two methods of composition: typing on a keyboard and dictating directly into text by means of VRT. While his writing processes in the respective modalities differed dramatically, particularly in terms of revision, the quality of the resulting texts appeared roughly the same: A detailed content analysis of the comments, using Richard Straub and Ronald Lunsford's (1995) typology, revealed significant variation in only 1 of 25 variables, measured as a proportion of total words. Meanwhile, students surveyed indicated they found few differences between the typed and dictated comments in terms of their usefulness, clarity, and tone. These findings, along with a comparison of time on task and user impressions of the two modalities, indicate that VRT represented a valuable tool for producing end comments that the user was able to dictate fluently, but that the technology was ineffective for the limited editing and revising attempted within the design of the study.

Keywords: Voice recognition; Speech recognition; VRT; Teacher response; Dictation

1. Introduction

Available since the early 1970s, voice-recognition technology (VRT) has proved to be a boon to individuals with physical disabilities that prevent them from typing with ease (Stodden

* Corresponding author.
Email addresses: thomas.batt@mma.edu (T. Batt), wilsonsa@husson.edu (S. Wilson).

8755-4615/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.compcom.2008.01.004

& Roberts, 2005; Wade, Petheram, & Cain, 2001). By converting speech directly into text, VRT has also helped writers with learning disabilities who can bypass the difficulties of text production and concentrate on higher order concerns (De La Paz, 1999; Litten, 1999; Wetzel, 1996). Translation errors remain a problem, but with accuracy rates steadily improving—the latest iterations claim rates of 95% or higher—the technology appears poised to be the next "killer app": Bill Gates has proclaimed VRT to be "the future of computing itself" (qtd. in Honeycutt, 2003, p. 77). Facing stacks of student papers several times a semester, each paper demanding a detailed response, instructors of composition are understandably interested in a technology that purports to translate speech into text at rates up to 160 words per minute. The instructor end comment on student texts (i.e., remarks appearing at the end of or attached to the texts) appears particularly well suited to VRT: The technology translates fluent, uninterrupted speech most accurately, and the formulaic, transactional nature of the end comment (Smith, 1997) is conducive to fully formed phrases and sentences.

But still unclear is how VRT affects the composing processes of teachers, and in turn, the quality of their dictated responses. In his review of the literature on traditional dictation and VRT, Lee Honeycutt (2003) noted recent advances in voice-recognition technology and called for studies that "gather realistic data about the promise and limitation of VRT today" (p. 82), particularly in relation to writing processes and products. The present study answers this call by comparing two sets of instructor responses to student texts: the first set composed by means of VRT (Dragon NaturallySpeaking 7) and the second via keyboard. Also addressed are the processes by which the instructor composed these comments. More specifically, we sought to answer the following questions: What is the nature of his composing processes with VRT, including planning, drafting, revising, correcting and time on task? How does this process compare to silent writing via the keyboard? Is the content of the instructor's comments produced with the VRT—their focus and mode—different from comments he composed with the keyboard? Do they tend to assert a different degree of control over the student text? Do the two sets of comments differ quantitatively? Do students receiving the comments rate them differently in terms of helpfulness and clarity? How do students describe the tone of the comments? Finally, what can be said about the nature and implications of these various differences if and where they appear?

2. Background

In his review of research related to VRT, Honeycutt (2003) observed that studies that have compared silent writing to dictation have mostly involved children who have not yet reached the developmental stage where their speech and writing differ significantly. The one study that has compared the processes and products of mature writers using silent writing and dictation—John Gould (1978)—found few qualitative differences between dictated business memos and letters and those written in longhand, whether by writers experienced in dictation or novices. Composing via dictation was found to be considerably faster than typing. But it is important to note that, unlike users of VRT, the participants in this study did not see their dictated texts as they composed them. The ability to instantly view and revise text as one

vocalizes it is unique to VRT and certainly affects one's writing processes; the Gould study must be understood with this distinction in mind.[1]

Only one published study that we are aware of (Hartley, Sotto, & Pennebaker, 2003) addressed how VRT influences writing processes and products. This study compared academic correspondence produced by means of typing and VRT (Dragon NaturallySpeaking Version 5). The researchers found "large" differences in writing processes but very little in quality in terms of readability and frequency of surface errors. Dictated letters, sent as email, did tend to contain significantly shorter sentences than typed letters and fewer very long sentences (more than 50 words), but the mean lengths of dictated and typed letters (there were 14 of each type) were quite close. The author using VRT found dictating to be much easier and "more pleasant" than typing on a keyboard, in part because it felt "more like chatting." But he said that altering the structure of a sentence in "mid-flight" was much more difficult than when typing, as was constructing long and complex sentences (p. 9). It is not clear how frequently or extensively the writer sought to revise his prose, nor is the effect of genre on his writing processes discussed in depth. This is an important omission because VRT may be well suited to some genres but not others. In particular, VRT may not be an efficient modality for genres that require carefully crafted prose, such as formal academic writing. Honeycutt (2003) argued that VRT is best used to dictate first drafts and that authors should reserve their editing for the keyboard: "As most VRT users can attest, moving phrases, sentences, and paragraphs around the screen with oral instructions is incredibly slow when compared to keyboarding in such changes. There is little to suggest that continued product research or advancing technology will change this fact in the future; keyboard editing is simply faster and easer than voice editing" (p. 83).

According to Charles Lowe (2001), the composing stage best suited for VRT may be invention because the technology potentially allows the user to tap into the fluencies of oral speech. Proposing an oral version of Pat Belanoff, Peter Elbow, & Sheryl Fontaine's (1991) freewriting, Lowe suggested that "freespeaking" can allow users to generate content quickly and effortlessly while avoiding the anxiety that encumbers traditional invention strategies. Susan De La Paz (1999) made a similar argument in recommending VRT for students with learning disabilities. The emphasis on speech, she wrote, would help such students bypass problems with text production, such as handwriting, spelling, and punctuation, to concentrate on their meaning. The literature on oral instructor feedback is extensive (e.g., Anson, 1997; Sommers, 1989; Stratton, 1975). However, as Honeycutt (2003) argued, dictation that is transcribed and read on the page falls closer to silent writing than to speech since "[m]ost dictating authors are keenly aware that their readers, spatially and temporally distanced, use the same review criteria used by readers of silent writing. Some aspects of written 'voice' may be influenced by the act of speech dictation, but for the most part, dictation is a distinct form of writing and not a form of speech, at least for mature writers" (p. 79). Furthermore, neither Lowe nor De La Paz addressed how even occasional mistranslations can distract an author dictating with VRT. Turning off the computer monitor, as Peter Elbow (1987) suggested as a way to minimize

[1] Reece and Cumming (1996) studied the impact on a voice-recognition system on the writing process of children and some adults. The system involved dictating to a hidden individual who instantly transcribed the dictated speech onto a screen visible to the children.

visual distractions while freewriting (on a keyboard), is one solution. But unless corrected immediately, the baffling mistranslations VRT often produces when one speaks unclearly can result in indecipherable text.

In any event, despite recent challenges to traditional teacher response (e.g., Straub, 2000; Fife & O'Neill, 2001; Prior, 2004), many composition instructors will agree that composing responses to student texts generally involves neither a great deal of invention nor extensive revision. Summer Smith (1997) showed that the end comment genre features sixteen primary genres characterized by "relatively stable content, style, and structure" (p. 267). Patterns of use varied little: For instance, 88% of the end comments in her study began with a positive evaluation. The formulaic nature of the end comment is abetted by its brevity and the large number of comments the average composition instructor must generate in a given response session, which also discourages heavy revision. If the typical end comment follows familiar patterns and requires little more than light editing, it would appear that the end comment genre is well suited to VRT, since the accuracy of the software depends largely upon providing it sufficient context to distinguish, for instance, between "very" and "vary."

For this reason VRT would not appear to be an ideal means of composing marginal and interlinear responses to student texts, since most instructors are not accustomed to writing out full sentences in text as they would have to in order to achieve maximum voice-recognition accuracy. On the other hand, the speed of VRT makes longer in-text remarks feasible, and many teachers would gladly round out their comments if they could do so quickly. This practice would also be in keeping with the recommendation of researchers, such as Nancy Sommers (1982) and Clyde Moneyhun (2002), who urge teachers to focus their in-text comments on student meaning rather than error, an emphasis that requires that comments be framed in complete sentences. In this respect, VRT may have the unintended effect of improving teacher response.[2] Also facilitating in-text commentary is the ability of VRT users to insert macros—prewritten words or phrases the author uses frequently—by a single-word command. Weighing against embedding commentary digitally in the student text are the logistical difficulties involved with exchanging digital texts, including incompatible programs, lost disks, and viruses. However, as teachers and students grow more facile with ever-improving technology, these problems should recede.

3. Method

3.1. Composing processes and time on task

The site of the study was Husson College, a small four-year college in Bangor, Maine, offering undergraduate degrees in majors related to business, health and (recently) liberal arts. The first-year composition course from which the student texts were drawn is required of all students. In each of four sections of the course, two in the Fall semester, two in the Spring of the 2004–2005 academic year, one of the authors, Author A, composed end comments on a set of student essays by means of two response modalities: silent writing (typing on a computer

[2] Thank you to an anonymous reviewer of this manuscript for this insight.

keyboard) and VRT. In order to improve reliability, he alternated the response modality every four or five comments in all four sections so that he would be using the two modalities under similar conditions. He recorded the time he spent composing each comment, including time spent revising and correcting his text and training the software to recognize his pronunciation of a given word or phrase (the initial training period is not included in these values).[3] He also recorded his ongoing impressions of the VRT and how it appeared to affect his writing and writing processes. Although Author A wrote marginal and interlineal comments on student papers, this study focused only on his end comments. We made this choice both to limit the scope of the study and to avoid the logistical difficulties of collecting and returning digital versions of student texts. The texts were the "final" drafts of a four-draft essay that also included a rough draft, a mid-process draft (which he and the students discussed in conference) and what were called the concluding revisions. Author A's comments were accompanied by a grade, but he encouraged students to revise still further for a higher grade; thus most of his comments included both formative and summative[4] evaluations. Students chose their own essay topics within the genre the class was studying at the time: a narrative responding to a published text and a source-supported research paper.

3.2. The voice-recognition software

The VRT we used in this study was Dragon NaturallySpeaking 7. The software can recognize single words in isolation but relies on the context of the user's speech to achieve advertised accuracy rates of up to 99%. The user "trains" the software initially by reading several passages aloud and correcting errors and improves accuracy further by correcting errors in subsequent sessions. Upon installation, the software scans the user's saved documents to learn user-specific sentence patterns and vocabulary, such as personal names and place names. As the user dictates (into a headset microphone), the software automatically supplies capitalization, apostrophes, and spacing between words, but most punctuation must be inserted by voice command (e.g., "I see what you mean period"). A full range of editing and navigation voice commands is available; however, as Honeycutt (2003) noted, manipulating text with voice commands is "incredibly slow" compared to doing so with keyboard and mouse (p. 83). In order to make the study as realistic as possible, then—that is, to approximate how most users would behave—we decided that Author A would use the keyboard and mouse to navigate around the text, copy and paste text, and correct errors except when the VRT repeatedly mistranslated a word or phrase, in which case he trained it on his pronunciation (an 8–10 second process). At the same time, we did not want to presume all revision to be impossible with VRT without investigating whether or not this is the case. Therefore, when dictating, Author A refrained from using the keyboard to *add* language to a given draft; all additions longer than one or two words were

[3] The training period consisted of four sessions, roughly 45 minutes each over a period of two weeks. Author A was familiar with the VRT, having worked with it three years previous, so the sessions involved refreshing his familiarity with the software, learning new features and training the software to recognize peculiarities of his speech (such as a tendency to slur certain consonants.).

[4] Formative comments are intended to guide the student's revisions to her next draft; a summative evaluation assumes that the student will not further revise the given text.

dictated. We realize we are drawing an arbitrary line, but we believe this protocol provided the best data for our purposes.

It is important to note that Author A was familiar with the software before the study began, having used an earlier iteration for a short while two years previous. Also, for several weeks before the study he used the VRT to compose emails and letters and to dictate passages from printed texts. So while he was not an expert in Dragon NaturallySpeaking, neither was he a complete novice. For instance, he had learned that speaking in a slightly exaggerated cadence, as if reading aloud to an audience, helped him to keep track of his syntax over long sentences. He also learned to enunciate much more clearly than he does in normal speech, moving his lips and jaws pronouncedly to avoid mumbling. Precise enunciation is crucial because VRT translates any disfluencies, slurred speech, or garbled pronunciations into correctly spelled but often bizarrely unrelated words and phrases that can be very difficult to spot (and decipher) later. Even carefully enunciated words are occasionally mistranslated, especially if they are inserted out of context. For instance, "A tough issue" becomes "A tough and should." As a result, texts produced by VRT must be proofread carefully, which can add considerable time to the composing process, especially during the initial training period when the software learns the user's speech patterns. As an advanced beginner, Author A made fewer errors than would most rank novices but more than truly proficient users. The data should be read with this caveat in mind.

3.3. Student survey

At the time he returned their essays, Author A gave students time to read his handwritten marginal remarks on their texts and the attached endnote. He then asked them to re-read the endnote before completing a brief survey. Intended to measure whether students perceived a difference between comments produced via VRT and comments typed on a keyboard, the survey consisted of three questions: "How helpful did you find the instructor comment?" "How clear did you find the comment?" and "What words would you use to describe the *tone* of the comment?" Students were to assume they were going to revise their essay. They had been informed several weeks earlier that the study involved VRT and teacher response, which may have influenced their responses somewhat. Students did not know which modality Author A had used to compose the endnote on their particular paper.

3.4. Content analysis

Author A's endnotes on 58 student compositions comprised the data set for measuring the occurrence of Straub and Lunsford (1995) categories of response to student writing. Of the available classification schemes, we found Straub and Lunsford's to be best suited for our purposes because it distinguishes between two elements of instructor comments we wanted to see reflected in our data: the *focus* of the comments and their *mode*. The focus has to do with the specific elements of the student text the instructor attends to: for example, the structure of an essay or its diction. The mode describes how or in what way the instructor presents his response: via open-ended questions, for instance, or reader-response observations. The second

Table 1
Straub & Lunsford's (1995) categories of instructor response

Focus	Mode
Global	*Corrections*
Ideas, development, structure	*Imperatives:* firm and soft directives
	Evaluations: qualified negative, praise, firm negative
	Advice
Local	*Indirect requests*
Wording, correctness, structure	*Closed questions*: problem posing, heuristic
	Open questions: problem posing, heuristic
Extra-textual comments	*Reflective statements*: interpretation, explanatory, experience, remarks, response, reaction

author, Author B, classified each comment into one of three broad categories under focus and eight categories under mode (Table 1).

Under the category of focus, global comments address larger concerns such as ideas, assertions, arguments, explanations and support as well as organizational issues: thesis, arrangement of paragraphs, overall focus, unity and coherence. Local comments deal with wording, structure (within a sentence and within paragraphs), and correctness: matters of grammar, punctuation, spelling, and mechanics. Extra-textual comments focus on concerns beyond the student text, such as the assignment, the classroom context, the writer's purposes, writing processes, or progress as a writer. Also included in this category are comments related to the essay as a whole, the student's work on the writing, and suggestions to seek help during office hours or at the writing center.

The category of mode of comment contains multiple subcategories. Corrections describe instances where the instructor actually corrects the student's text. Imperatives include firm and soft directives. Under evaluative comments fall firm and qualified negative evaluation and praise. Advisory comments suggest changes but do so in a tempered manner that falls short of imperatives. An example of an indirect request might be, "Can you start with this paragraph?" Questions can be either open ("Is this your main point?") or closed ("Is that the only option?") and can be posed as problem posing ("Is there a way to develop this argument?") or heuristic ("Do you see this material as your key point?"). Finally, reflective statements include "descriptive, interpretive, explanatory, reader response, and hortatorical statements that are not evaluative, directive, or advisory" (Straub & Lunsford, 1995, p. 170). A comment in reader-response mode, for example, might be, "I'm a bit lost here."

We chose Straub and Lunsford's taxonomy in part because it provides a means to gauge the degree of control exerted by the instructor over the student text, one more way to compare the quality of dictated and typed responses. Using the same criteria in a 1996 article, Richard Straub outlined a method for gauging the degree of teacher control on a continuum ranging from directive to facilitative, or most to least controlling, depending on the focuses and modes of response the instructor tends to exhibit. For instance, according to Straub, instructors who respond as readers and who focus on global or larger rhetorical concerns, tend to be less directive and more facilitative than instructors who issue imperatives and address mostly local matters, such as sentence style and paragraph development (1996, p. 234). Longer comments tend to be directive while a greater emphasis on extra-textual matters generally indicates a

more facilitative response style. We should point out that because this study is confined to endnotes, the data are more likely to portray Author A's style as facilitative than if we had included his marginal and interlineal remarks, which by their nature are more likely to be local and corrective. But our concern here is the *relative* quality of comments composed via the two modalities, and the data should be read with this focus in mind.

After completing an inter-rater reliability analysis,[5] Author B used a total of 29 variables to analyze Author A's comments, those included in Table 1, plus gender, grade (as a numerical value), the number of words, and the modality, entered as a binary value. Author B did not know at any point which comment sets she analyzed were typed and which was dictated. Our sample size gave us a statistical power of 0.75 with a specified effect of 0.5. Author B coded the 58 comment sets for focus and mode and generated frequencies and descriptive statistics for all variables. Independent sample *t*-tests were conducted on all variables to compare the two modalities. (Our null hypothesis was that the treatments are not different.) The independent tests allowed us to compare, for example, the average number of extra-textual comments generated via voice command to the average number of such comments composed via keyboard. We chose the independent *t*-test because it best suited the conditions of our study: one teacher and multiple students. Ideally, Author A would have responded to each student essay using both modalities. In this case, a repeated-measures *t*-test would have been appropriate. However, Author A could not read the same essay a second time and pretend not to have read it before. For this reason, we divided the student papers into two independent groups and assumed that Author A's abilities as a responder remained the same from one student paper to the next.

Variables with extreme outliers were retested without the outlier(s) to discern the extent to which they influenced the outcome. In a further analysis, each variable was computed as a *proportion* of the total number of words in the comment set and then the averages of the proportions between each modality were compared using the independent *t*-test. In addition, descriptive statistics and box plots were computed to help present the results. We thought such a proportional analysis would give a finer picture of the differences in the two modalities. One would expect more words of comments on the global development of ideas, for instance, in a comment set that is long, but looking at the proportion of such comments to the total words in the comment set gives a picture of a given variable relative to other variables in the set. No adjustments were made for multiple comparisons. Statistical significance was assessed at the 0.05 level. Finally, we used the method described by Straub (1996) to determine the degree of control asserted by Author A in comments composed using the two modalities.

Our efforts to minimize bias notwithstanding, we have no positivist assumptions about the objectivity of our approach and methods, which were inevitably shaped by our unique personal histories and cultural perspectives, among other factors. The quantitative nature of our data notwithstanding, we make no claims that our results can be generalized to any larger population, such as teachers of writing. Given these limitations, we do believe that the study sheds light on what may emerge as an important modality for teacher response.

[5] In the first reliability study the proportion of agreement was .67 for focus, .67 for mode; the second was .82 for focus and .73 for mode. Together, the studies gave us an agreement of 74 for focus and .70 for mode.

4. Results

Dictated comments took roughly the same amount of time to produce on average as typed comments; however, the rate of the fastest VRT times exceeded the fastest silent-writing times by a considerable margin. The average length of dictated comments was almost a third shorter than typed comments. In two important respects Author A's composing process when he used the VRT differed from his process when typing: He composed full sentences in his head before committing them to text, and he was less likely to edit and revise the resulting text to his satisfaction. (See section 4.1 for results on time on task and writing processes.) At the time he composed them, Author A believed his dictated comments were less cohesive and precise than his typed comments but this impression is belied by the student survey, which indicates students found the dictated comments to be slightly more clear than, and equally helpful as, typed comments. (See section 4.2 for survey results.) An analysis of the focus and mode of the comments showed significant difference between the two treatments in 5 of 25 variables; however, as a proportion of total words, only one variable—extra-textual comments—was significantly different: A lower proportion of extra-textual comments appeared in dictated comment sets. (See section 4.3, which also contains data on word production for each modality.) In this respect, dictated comments could be construed as less controlling than typed comments.

4.1. Time on task and composing process

The VRT did not save a significant amount of time on average in terms of words per minute.[6] The rate at which Author A composed comments via dictation was only slightly faster than the typing rate: 27.0 wpm and 26.4 wpm, respectively. But while the slowest times of the respective modalities were comparable (five values under 20 wpm for each modality), the fastest times of the VRT were considerably swifter than the fastest times of the typing modality: upwards of 55 wpm vs. 40 wpm, respectively. The mean length of dictated comments was 32% shorter than that of typed comments (165 words vs. 219 words).

Author A's writing process was substantively different when he worked with the VRT than when he composed via the keyboard. When writing silently, his thoughts usually ran ahead of his typing and he used this lag time to reflect and to form sentences. He paused often but usually only at the end of complete sentences or ideas. Because the VRT, by contrast, would translate his words into text within seconds of his speaking them, Author A found that he would often momentarily run out of things to say in mid-sentence, a problem exacerbated by the necessity to speak at an even pace, enunciate carefully and avoid disfluencies such as "um" and "er." The dictation proceeded more smoothly when he composed complete sentences in his head before speaking them, a significant departure from his composing process when typing. Like Elbow (1987), Author A found that closing his eyes while he spoke helped him to concentrate on his meaning. But even when he knew just what to say and how to say it, Author A was careful not to dictate more than two or three sentences at a time: Unlike most of his typewritten

[6] Author A later revised the last sentence of the comment, increasing the time to 2.5 minutes. Because he corrected errors as he composed, he calculated the words per minute rate by dividing number of words by time and did not account for any remaining errors.

errors, which tend to be misspelled and so underlined in red (in Word), dictation errors often bore no resemblance whatsoever to his original meaning. So he found it prudent to correct the translation errors before he forget what he had meant to say. For instance, here are two versions of the same comment; the first is the corrected version (i.e., what Author A meant to write), the second is how the VRT translated his words:

This is a well put together essay, William. You convey a lot of information in a dry, knowing voice that somehow avoids deteriorating into cynicism. I appreciate how well structured the paragraphs are, and the ease of your transitions. You seem very comfortable with the subject. As a reader, I wanted more specific information, which would have been in keeping with this assignment—your source support is a bit light. For instance, I noticed in your photocopied sources some of the science behind corked bats and their performance; it would be nice to see some of that in the essay. You could also document some of the bad press that Sosa and others got—that kind of thing. But this essay is a bit of a breakthrough for you: you really seem comfortable as a writer for the first time this semester, and I wouldn't be surprised if you enjoyed writing the essay, at least a little bit.

This is a well put together essay, William. You convey a lot of information dry, knowing voice that somehow avoids deteriorating into cynicism. I appreciate how well structured the paragraphs are, and these few transitions. You seem very comfortable the subject. As a reader, I wanted more specific information, which would have been in keeping with the assignment—your source support is a bit light. For instance, I noticed in your photocopied sources some of the science behind corked bats and their performance; they be nice to see some and your essay. You could also document some of the bad press that Sosa and Scott—that convey. But this essay is a bit of a breakthrough for you: you really seem comfortable as a writer for first-time semester, it would be surprised if you enjoyed writing essay, at least help.

Some of the many errors in the passage would be easy to spot and correct at any time, such as the missing preposition and article in sentence two. But unless he proofread immediately, Author A would have been baffled by "these few transitions" in sentence two and the "Sosa and Scott—that convey" mistranslation in the seventh sentence. Constantly switching modes in this way between drafting and correcting impeded a fluent flow of ideas, he found, and seemed to make his comments less cohesive than usual. To regain his meaning and sense of direction, he would have to re-read from the top of the comment. To be sure, Author A also constantly corrected when composing with a keyboard; but perhaps as a result of long practice, the correcting has become an integral part of his drafting process and in the study he did not find it as distracting as having to stop drafting to hunt errors. Of course, for the same reason, experienced VRT users may find correcting with VRT to be similarly routine and seamless.

Author A found editing with the VRT to be difficult, although using the mouse (to navigate around the text and select language) and the keyboard (to delete text or add up to three words) facilitated the process. Because the software is less accurate translating words and phrases out of context, which editing often demanded, correcting the proliferation of errors was often time-consuming. When Author A tried to reword ideas in his head before dictating (in order to give the software sufficient context to translate his words accurately), he often struggled: Again, it could be no more than the product of long habit, but he found problem solving by keyboard, with its tactile dimension, to be much more efficient than problem solving by means of silent reflection and oral speech. These problems tended to discourage Author A

from honing his meaning quite to the degree he was accustomed to, and he had the impression that his dictated prose was somewhat less precise than the texts he refined on the keyboard.

Given these hurdles, it is not surprising that Author A also struggled to revise with the VRT. (We define "revision" as manipulation of global elements such as structure, focus, and voice.) When revising with the keyboard, Author A proceeded haltingly, often stalling on a phrase or passage while rehearsing and editing alternative language. When revising via VRT, Author A tended to hesitate frequently as he considered his meaning, to speak unclearly or inaudibly and to vary his cadence, all of which led to a profusion of translation errors that needed to be corrected, interrupting his ideation. Author A revised only five comment sets out of the 30 he dictated, about the same percentage as the typed comments; but, as was the case when he dictated local changes, he did not believe that his voice revisions as closely matched his intended meaning as did revisions he made with the keyboard.

4.2. Student survey

Students were asked to rate the helpfulness (1 = very helpful; 4 = not helpful at all) and clarity (1 = very clear; 4 = not clear at all) of the endnote (or "comment set") attached to their paper. They were also asked to describe in their own words the tone of the comment. Students rated typed and dictated comments about the same in terms of helpfulness: 1.43 mean for typed ($n = 23$) versus 1.41 for VRT ($n = 27$); in terms of clarity, typed comments were rated somewhat lower than comments composed with the VRT: 1.65 mean for typed ($n = 23$) versus 1.41 for dictated ($n = 27$). In describing the tone of the comments, students provided 45 descriptors overall, most of them occurring with roughly equal frequency in regard to the two sets of comments (typed and dictated). Most common were variations on "helpful/responsive" and "encouraging/motivating." The descriptor "concise" or a close variant appears three times in reference to the typed comments and once to a dictated comment. One student described a dictated comment as "less brief" than comments he or she had received on previous occasions. Only three students in all remarked on the voice of the comment they received, two who received typed comments and one who received a dictated comment. Of the first group, one student wrote (referring to the instructor), "I can hear him speaking to me, yet I feel it is very impersonal. The typing makes it too formal." The second wrote, "The tone of the comments was 'live.' It was as if you were talking to me face to face." The student receiving a dictated comment made a similar observation: "The tone was as if you were just talking to me about the way I should better my paper. As if we were face to face." Again, the students were aware that the survey involved a study of VRT, which may have influenced their responses. However, it should also be noted that Author A had met in conference individually with many of these students the previous week, which may also have inadvertently shaped their responses.

4.3. Analysis of instructor comments

The analysis of Author A's comments composed by means of keyboard and VRT respectively showed the two treatments to be significantly different in 5 out of 25 variables. However, the mean number of words for the VRT comments (1 5 5) was less than the mean total word count for the typed comments (2 1 3); looking at the variables as a *proportion* of total words in each

Table 2
Results of independent sample *t*-tests in which statistical significance was discerned

	Means of words for typed and dictated comments and		Sig. (two-tailed)	Sig. (two-tailed) with outliers selected
	Typed	Dictated		
Grade	7.1	8.7	0.012	
Global development of ideas (focus)	42.57	35.0	0.021	
Extra-textual comments (focus)	70.	40.4	0.007	0.013
Qualified negative comments (mode)	35.0	19.3	0.067	
Reflective interpretive (mode)	42.3	23.4	0.011	0.009
Reader remarks (mode)	26.5	13.5	0.005	0.008
	Mean proportion of typed and dictated comments to total words in comment sets.			
Proportion of extra-textual comments to total words (focus)	0.3360	0.2348	0.020	0.008

modality, then, only one variable—extra-textual comments—emerges as significantly different in the two modalities: Typed comments had a higher proportion of extra-textual comments than comments composed with VRT. (Again, these are comments referring to contexts beyond the immediate student text, such as the student's writing processes, her progress as a writer, conversations the instructor may have had about the text at hand, and so on.) In Table 1, results of independent sample *t*-tests that disclosed statistically significant differences are presented with equal variances NOT assumed.

Table 2 shows the results of independent sample *t*-tests in which statistical significance was discerned. Refer to box plot below (Figure 1) for apparent outliers. When both outliers are selected, rather than the outlier for modality one (typed comments), two-tailed significance increases to 0.003 whether or not equal variances are assumed.

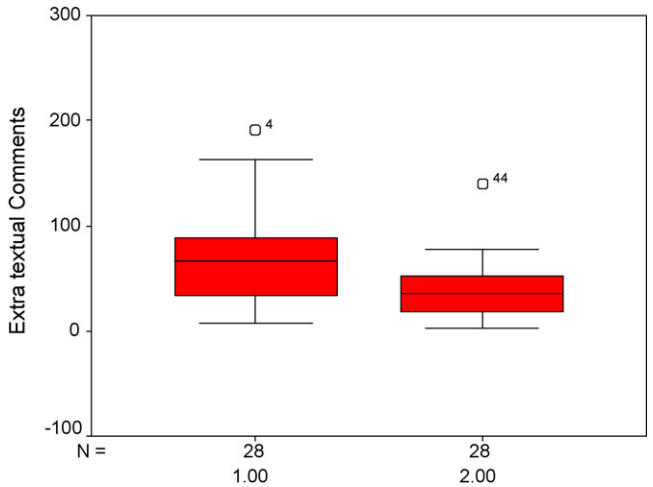

Fig. 1. Typed comments (1.00) and dictated comments (2.00).

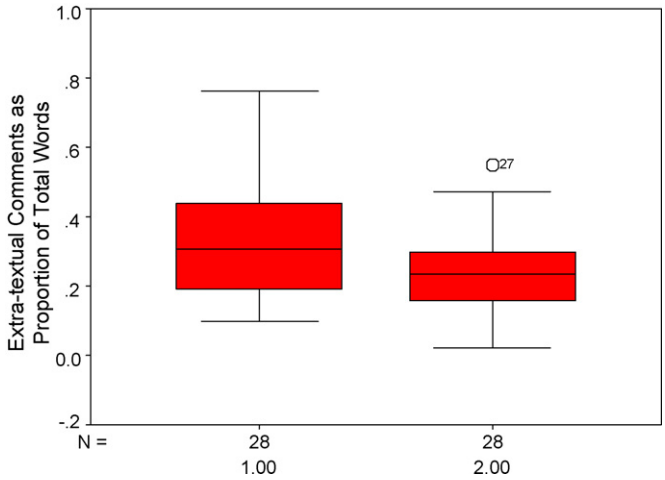

Fig. 2. Typed comments (1.00) and dictated comments (2.00).

Figures 1 and 2 show differences in the one variable that was statistically significant as a proportion of words to total words in the comment set. The outliers that appeared in the box plots were the ones removed for the repeat *t*-tests as a way to investigate their influence on the significance of the difference. Figure 1 shows the number and range of extra-textual comments, with outliers (modality one, typed; modality two, VRT): The vertical lines represent the range of scores for extra-textual comments in each modality, with 28 cases in each modality and a minimum of seven and a maximum of 163 with a possible outlier at 191 words when comments were typed, and a minimum of three and a maximum of 77 when they were composed via VRT with an outlier at 139. The mean of scores (40.4) and median score (36) are lower in the dictated comment compared to the mean (69) and median score (66) of typed extra-textual comments. The two cases that appeared to be outliers in each treatment (#4 in typed comments and #44 in dictated comments) were removed and the *t*-tests repeated to investigate their influence on the results ($p = .007$ with the outlier with the outlier; $p = .013$ without the outlier). There is a statistically significant difference between the two modalities with regards to extra-textual comments regardless of the two outliers. Box plots were computed for all the variables as a proportion of total words in the comment sets. Figure 2 shows the box plot for the proportion of extra-textual comments in the two modalities.

Twenty-eight cases appear in each modality with a high proportion of 0.76 of total comments for modality one (typed comments) versus 0.47 in modality two (dictated comments) with a possible outlier of 0.55. The mean proportions are 0.34 and 0.23 respectively in the two modalities ($p = .02$ and .008 with outliers removed). The results of the independent sample *t*-tests show five variables and one variable as a proportion to total words out of 29 variables, plus 29 proportional variables, as being statistically significant. A caveat: In the measurement for a number of variables the standard deviation was so large that significant differences were impossible given our sample size, so caution needs to be used when considering our results. Also, we need to be cautious about discriminating between the importance of the variables and their statistically significant differences.

The similarity between the typed and dictated texts means that they were also similar in terms of the level of control they asserted over the student texts. The one variable in which the VRT modality differed significantly from silent writing as proportion of total words—the lower rate of extra-textual remarks in dictated comment sets—suggests a less controlling, more facilitative approach, according to Straub (1996): "The more a teacher attends to the student's writing processes and the larger contexts of writing, and gears his comments to the student behind the text and her ongoing work as a writer, the less likely he is to point to specific changes or to assume control over the student writing" (p. 233–234).

5. Discussion

Perhaps the most important finding of this study is that, with the exception discussed below, comments dictated by means of VRT appeared to be of similar quality to comments typed on the keyboard. This result should not surprise us, for several reasons: First, as Honeycutt (2003) observed, dictation falls much closer to writing than to speech because the dictating author knows that his or her words will be read as text and held to the same standards; most authors using VRT, then, will likely strive to match the quality of their typed prose, as Gould (1978) study of traditional dictation suggested would be the case. Second, like writers composing on a keyboard, VRT users can see—and revise—their prose instantly, which gives them at least the opportunity to achieve their normal standards for text quality. Finally, the formulaic nature of the end comment genre, with its familiar repertoire of response patterns, would probably incline most instructors to produce comments of similar quality no matter what modality they use. Yet our findings do surprise us because, at the time he composed them, Author A believed his dictated comments to be less cohesive and less precise than his typed comments. This impression was based on how frequently he had to stop to proofread when drafting with the VRT (which interrupted his drafting process), the difficulty composing orally, and his struggles to edit and revise while correcting frequent translation errors. It may be that our instruments were not sufficiently precise to measure this perceived quality gap; further studies are needed to determine the nature of the disparity, if in fact it exists. But an unambiguous finding of our study is that translation error had a negative effect on Author A's composing processes, interrupting his ideas and discouraging him from refining his meaning.

Translation error also appears to be linked to three other effects we measured: the short mean length of dictated comments compared to that of the typed comments; the slow overall performance of the VRT as a response tool, compared to its peak rate under optimal conditions[7]; and—in the one significant qualitative difference we found between dictated and silent writing—the low frequency of extra-textual remarks in the VRT-produced comments. The shorter length of Author A's dictated responses might be related to his frustration editing and revising his prose orally: He may have truncated his remarks rather than struggling further to achieve the effect he desired. The continual pauses to proofread dictated sentences certainly slowed his composing process when dictating, particularly when he had to correct errors; when

[7] The Nuance website ⟨http://www.nuance.com/naturallyspeaking⟩ claims that users of Dragon NaturallySpeaking can dictate texts up to 120 words per minutes.

Author A spoke confidently and fluently, as per VRT instructions, the resulting prose emerged relatively quickly and accurately. On these occasions, the VRT was considerably faster than silent writing. Otherwise, the software was no faster—and in many cases slower—than typing. Author A's struggle with translation error may explain the one significant difference we found between the typed and dictated texts: the relative paucity of extra-textual comments in the dictated endnotes. Watching carefully for mistakes in his texts as he dictated, Author A may have been less mindful than he usually is of the broader contexts related to the student text he was commenting on, leading him to omit at times the kinds of extra-textual comments (such as remarks about the assignment, the student's writing history or the need to seek extra help) that occur to him when he is free from this distraction. This missing dimension may be the root of Author A's sense that his dictated comments were less precise than his typed responses.

In sum, the VRT was not an effective tool overall for composing instructor end comments under the conditions of this study. For the most part, the software was frustrating to use and no faster than silent writing, especially when Author A attempted to revise. Moreover, Author A was not satisfied with the quality of his dictated comments, although the student survey and content analysis showed the comment sets to be similar. Where these instruments showed the sets to be different—the lower proportion of extra-textual remarks in the dictated set—we see another indication that the difficulty of using the VRT impeded what Author A wanted to say to his students. The software was fast and relatively accurate only when Author A dictated comments that required little or no editing or revision.

As accuracy improves with new iterations of VRT, future studies could investigate whether oral editing and revision is more efficient than we found, or whether these tasks should be restricted to the keyboard. Future inquiries might also address how instructors with disparate response styles and composing processes fare with VRT; for instance, VRT appears well suited for transcribing texts, such as passages from student writing. Author A did not make use of the VRT for this purpose, but most tests of VRT that yield results of 95% or great accuracy involve reading prepared texts (For example, see Pogue, 2006). Our study involved a culturally homogeneous population of students at a career-oriented school in Maine. The situatedness of participants, researchers, locale, study design, and method necessarily limits the conclusions one can draw from this study. Studies at other kinds of institutions, involving different contexts, populations, and research methods would complement the present inquiry. Finally, this study focused on instructor end comments, yet teachers could also use VRT to compose marginal and interlinear remarks. A study of the efficacy of VRT for this purpose would also be useful.

For instructors interested in experimenting with VRT, Dragon NaturallySpeaking comes in several versions, including those that support medical and legal lexicons. At $200, the Preferred version is about twice as expensive as the Standard version but supports macros (pre-composed texts) and allows the user to play back recorded passages. Both versions come with a microphone and work in all Windows-based applications. To work smoothly, the program requires 512 of RAM and at least one gigabyte of free hard-drive space. Dragon NaturallySpeaking has been the dominant player in the market in recent years; IBM ViaVoice, marketed by the same company (Nuance Communications), has not been updated in four years and, according to some reviewers, may be phased out soon. Dragon NaturallySpeaking works well with Mac computers. An alternative for Mac users is iListen, but this software is less accurate and sophis-

ticated than Dragon NaturallySpeaking according to David Pogue (2006), who speculates that the strongest competition for Dragon NaturallySpeaking may end up being the dictation utility embedded in Windows Vista.

6. Conclusion

This study sought to determine how VRT influenced the writing processes and writing quality of a college composition instructor composing endnotes in response to student texts. We found that composing with VRT involved dramatically different processes than composing on a keyboard, particularly in terms of editing and revision, which were difficult with the VRT. However, the impact of VRT on writing quality appeared to be minimal. Dictated comments did contain significantly fewer extra-textual remarks proportionately than typed comments, a result we believe is linked to Author A's continual proofreading of his dictated texts. When used to compose comments or elements of comments that the instructor was able to dictate fluently, the VRT was a faster modality than silent writing; used as a tool to edit or revise comments, or to compose comments that required recursive drafting methods, the VRT did not represent an efficient means of teacher response.

Thomas Batt is an associate professor at Maine Maritime Academy, where he teaches courses in composition and humanities. His research interests include composition pedagogy, teacher response, and writing center theory and practice.

Sandip Wilson is an assistant professor of Literacy and Teacher Education in the School of Education at Husson College in Bangor, Maine.

References

Anson Chris, M. (1997). In our own voices: Using recorded commentary to respond to writing. In Sorcinelli Mary Deane & Elbow Peter (Eds.), *Writing to learn: Strategies for assigning and responding to writing across the disciplines* (pp. 105–113). San Francisco: Jossey-Bass Publishers.

Belanoff, Pat, Elbow, Peter, & Fontaine, Sheryl. (Eds.). (1991). *Nothing begins with n: New investigations of freewriting*. Carbondale and Edwardsville: Southern Illinois University Press.

De La Paz, Susan. (1999). Composing via dictation and speech recognition systems: Compensatory technology for students with learning disabilities. *Learning Disabilities Quarterly*, *22*, 173–182.

Dragon NaturallySpeaking 7 [computer software] (2003). New York: Nuance Communications.

Elbow, Peter. (1987). Closing my eyes as I speak: An argument for ignoring audience. *College English*, *49*(1), 50–69.

Fife, Jane Mathison, & O'Neill, Peggy. (2001). Moving beyond the written comment: Narrowing the gap between response practice and research. *College Composition and Communication*, *53*, 300–321.

Gould, John D. (1978). How experts dictate. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 648–661.

Hartley, James, Sotto, Eric, & Pennebaker, James. (2003). Speaking versus typing: A case-study of the effects of using voice-recognition software on academic correspondence. *British Journal of Educational Technology*, *34*(1), 5–16.

Honeycutt, Lee. (2003). Researching the use of voice recognition writing software. *Computers and Composition*, *20*, 77–95.

Litten, Malcolm. (1999). Introducing voice recognition software to dyslexic students users in a special school. *Dyslexia*, *5*(2), 118–120.

Lowe, Charles. (2001). Speech recognition: Sci-Fi or composition? *Currents in Electronic Literacy 4*,. ⟨Retrieved May 15, 2006 from http://www.cwrl.utexas.edu/currents/spr01/lowe.html⟩

Moneyhun, Clyde. (2002). Less is more in response to student writing. In Veronica Pantoja Duane Roen, Susan K. Miller Lauren Yena, & Waggoner Eric (Eds.), *Strategies for teaching first-year composition* (pp. 238–326). Urbana, IL: NCTE.

Pogue, David. (2006, July 20). Like having a secretary in your pc. New York Times. ⟨Retrieved August 15, 2006, from http://www.nytimes.com/2006/07/20/technology/20pogue⟩.

Prior, Paul. (2004). Tracing processes: How texts come into being. In Charles Bazerman & Paul Prior (Eds.), *What writing does and how it does it: An introduction to analyzing texts and textual practices* (pp. 167–200). Mahwah, NJ: Lawrence Erlbaum.

Reece, John E., & Cumming, Geoff. (1996). Evaluating speech-based composition methods: Planning, dictation, and the listening word processor. In C. Levy Michael & E. Ransdell Sarah (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 361–380). Mahwah, NJ: Erlbaum.

Smith, Summer. (1997). The genre of the end comment. *College Composition and Communication*, *48*(2), 249–268.

Sommers, Jeffrey. (1989). The effects of tape-recorded commentary on student revision: A case study. *Journal of Teaching Writing*, *8*, 49–75.

Sommers, Nancy. (1982). Responding to student writing. *College Composition and Communication*, *33*(2), 148–156.

Stodden, Robert A., & Roberts, Kelly D. (2005). The use of voice recognition software as a compensatory strategy for postsecondary education students receiving services under the category of learning disabled. *Journal of Vocational Rehabilitation*, *22*(1), 49–64.

Stratton, Charles R. (1975). The electric report card: A follow-up on cassette grading. *Journal of Technical Writing and Communication*, *5*(1), 17–22.

Straub, Richard. (1996). The concept of control in teacher response: Defining the varieties of "directive" and "facilitative" commentary. *College Composition and Communication*, *47*(2), 223–251.

Straub, Richard. (2000). The student, the text, and the classroom context: A case study of teacher response. *Assessing Writing*, *7*(1), 23–55.

Straub, Richard, & Lunsford, Ronald. (1995). *Twelve readers reading: Responding to college student writing*. Cresskill, NJ: Hampton Press.

Wade, Julia, Petheram, Brian, & Cain, Robert. (2001). Voice recognition and aphasia: Can computers understand aphasic speech? *Disability and Rehabilitation*, *23*(14), 604–613.

Wetzel, Keith. (1996). Speech-recognizing computers: A written communication tool for students with learning disabilities? *Journal of Learning Disabilities*, *29*, 371–380.