Teacher Commentary Via Speech Recognition

W. Keith Duffy, The Pennsylvania State University

W. Keith Duffy, Ph.D., is Associate Professor of English at Penn State.

Abstract
Providing written feedback to students on drafts for purposes of revision is the single most time-consuming aspect of a composition teacher's job. To lighten the load, many teachers attempt to *speak* feedback to students instead, either in conferences or by using various analog or digital recording devices. Exploring the evolution of this practice, the author anecdotally investigates the possibilities of providing students with textual feedback using speech recognition technology (SRT), transcribing spoken teacher commentary directly onto students' drafts.

Introduction
In April of 2013, the nonprofit educational group EdX (www.edx.org), founded jointly by Harvard University and the Massachusetts Institute of Technology, released a newly developed set of applications that "uses artificial intelligence to grade student essays" (Strauss, 2013). The software, which is being offered to many secondary and postsecondary schools for free, has ignited a debate between traditionalists and progressivists.

On one side, Anant Agarwal, an electrical engineer who is president of EdX, and his colleagues argue the software will allow students to receive instantaneous feedback on their essays for purposes of revision (Markoff, 2013). No longer will students have to wait to get their papers back from their instructors. On the other side are rhetoricians who argue that writing—a fundamentally human activity--can only be assessed by those it was meant for: other humans. Doug Hesse (2013) in "Grading Writing: The Art and Science — and Why Computers Can't Do It," makes an important distinction between *grading* writing and *responding to* writing: "It's one thing to say whether your writing is any good; it's quite another to explain to you helpfully why" (para. 4).

While there is no decisive winner in this debate, the entire argument highlights a single truth: Responding to student writing is the *single most time-consuming aspect of any writing teachers' job.* Indeed, EdX's grading software could be considered more of a time management tool than anything else. According to studies conducted by Richard Haswell (2008), postsecondary

composition instructors can spend an average of seven minutes per page commenting on student writing; in a classroom filled with pupils who may be each generating five-page research essays every two weeks (a typical scenario), the time expenditure increases exponentially. In fact, the sheer amount of labor required to provide individualized commentary to students has driven many colleges and universities to limit enrollments in writing courses. Such limits are well documented in highly visible policies like the "Statement of Principles and Standards for the Postsecondary Teaching of Writing," formulated and published by the National Council for Teachers of English (NCTE, 2012).

Investigating solutions to this dilemma, the following paper examines the history of writing teachers providing *spoken*—rather than written--feedback to students on drafts of their writing. The newest iteration of this practice involves the use of speech recognition technology (SRT) to transcribe and insert teacher feedback directly into electronic drafts of students' essays. In the process discussed here, students' essays are collected electronically as email attachments, they are read and commented on via SRT on screen, and returned to students in the same manner. By comparing traditional handwritten comments and SRT-generated comments, the following questions are examined: 1) Does utilizing SRT save writing teachers time and energy in the feedback process? 2) In what way does using SRT change the nature of the comments being provided to students? An anecdotal analysis shows that the use of SRT may take more time, but may paradoxically require less energy than writing or typing comments manually. Additionally, commentary provided via SRT has a tendency to be more narrative in structure, making feedback appear to be more formative, and less directive, in nature.

Speaking Up
The history of postsecondary writing instruction clearly illustrates one fact: *Speaking* feedback to students—in whatever form--is considerably easier than handwriting or typing comments on drafts. This has undoubtedly resulted in the widespread pedagogical use of conferencing in writing instruction, where teachers and students meet individually, or sometimes in groups, to discuss drafts face to face. However, while personal conversation can often ease the burden of writing comments to students (since typically students themselves are tasked with jotting down relevant points raised during a conference), this practice remains immensely time consuming.

Searching, then, for a technological solution to this quandary, as far back as 25 years ago the idea of vocalizing feedback to students on their drafts through the medium of audiotape appeared (Mellen & Sommers, 2003). In this asynchronous learning environment, teachers would make only minimal or no marks on a student's draft and instead speak feedback onto a cassette tape; when the paper was returned, it was accompanied by the spoken commentary. Neither the pedagogical effectiveness nor the widespread use of this approach was ever clearly determined. Nevertheless, as technology evolved from magnetic tape and into the realm of digital recording, the practice of vocalizing feedback to students continued with the use of MP3 players, WAV files sent over email, and podcasting (Still, 2006; Rotheram, 2007; Merry & Orsmond, 2008).

The Advent of Speech Recognition

In recent years, one approach experiencing a surge in development is speech recognition technology (SRT), one function of which is to transcribe the spoken word into editable text on a computer screen. Known equally as voice recognition technology (VRT), it was originally developed by AT&T Bell Laboratories in the 1940s, but it took over 40 years for it to become fully functional in the 1980s ("History," 2010). It didn't reach a consumer-friendly status until more than ten years later when several developers released commercial software products like Dragon's "Dictate" and IBM's "Via-Voice," allowing users to generate on-screen text in a word-processing program by speaking continuously without having to pause between each word (Pinola, 2011). Over the last decade, the technology has been further refined; the predictive capabilities of the software have improved, recognition error rates (a perpetual problem for dictation technology) have decreased, and transcription speed has increased. Voice-enabled applications on mobile devices, such as iPhone's Siri, have exploded in popularity, and the prices on a variety of speech-recognition software titles have plummeted to less than $100. As the technology evolves, speech recognition has found its way into many academic and professional fields; research has been conducted to determine its usefulness regarding younger students with learning disabilities (Barbetta & Spears-Bunton, 2007; Sherman, 2008; Silió & Barbetta, 2010), in legal dictation (Honeycutt, 2004; O'Shaughnessy, 2008), and in medical transcription (Carter-Wesley, 2009; Alapetite, Andersen & Hertzum, 2009), to name only a few.

Considering this perfect storm of development and price cutting, it's not surprising that postsecondary writing teachers and scholars have investigated the efficacy of using speech recognition software to provide textual feedback to students on their drafts. Since this technology transcribes the spoken word, it seems the ideal solution for easing the writing teacher's workload and time commitment.

Concerning these issues, in a recent article titled "'Speech-Driven' Technologies to Reduce Faculty Workload in Online Higher Education," Mike (2010) examines the use of SRT in online adult education. Without making any definitive suggestions, he catalogues the various ways online instructors naturally use voice recognition software to participate in discussion boards and respond to multiple student emails, thus eliminating the time-consuming job of manual typing. But when it comes specifically to *writing instruction* at the postsecondary level, few studies, save one, have attempted to examine whether such a method works in a practical way, if it saves time, and if students respond appropriately. In their article "A Study of Voice-Recognition Software as a Tool for Teacher Response," Batt and Wilson (2008) conducted research where a first-year composition instructor provided end comments on electronic drafts of student writing for purposes of revision using both silent writing (with a keyboard) and voice recognition software (with Dragon Naturally Speaking). Results of the study showed that "dictated comments took roughly the same amount of time to produce on average as typed comments," and the authors conclude that this approach "did not save a significant amount of time" (173). Ultimately, the authors argue that "VRT was not an effective tool overall for composing instructor end comments" (179).

Implementing Speech Recognition

Regardless of these negative findings, in 2010 I implemented SRT in my own first-year, required writing courses to ameliorate the time-heavy burden of manually commenting on students' drafts. Instead of collecting students' essays on paper and writing on them by hand, I collected their electronic drafts and used speech recognition software to insert text-based comments into the essays. I then returned the drafts via email so the student writers could revise.

After trying this technique for two years, I wanted to anecdotally observe if, and how, the nature of my feedback had changed with the use of this technology. Although examining the ways students responded to this pedagogy is important, that question is beyond the scope of this article. Instead, my primary questions were simpler, less formal, and more teacher-focused: Were my comments more or less detailed when handwriting versus using speech recognition? Was my feedback to students longer or shorter in either form? Was my feedback more or less directive or facilitative depending on my approach? Did the nature of my evaluations change with one technique over the other? Additionally, on a personal level, would providing feedback via SRT consume less time than writing feedback by hand? Of course, my familiarity with the technology increased over the two-year time frame, which unquestionably resulted in improved performance; nevertheless, would it outpace my time-tested approach of providing feedback by hand?

To inform this casual investigation, I examined the interlineal and end comments I wrote for 40 student essays over the past two years—20 of the essays included handwritten comments, and 20 included typed comments inserted via the speech recognition software Dragon Naturally Speaking 11 (the most recent version of the application per this writing). All of the essays originated from required first-year writing classes, and the essays were all essentially similar in terms of requirements (a persuasive essay requiring multiple documented sources in Modern Language Association style). In general, the essays were 4 to 5 pages in length, double-spaced. Below are three aspects I examined and a short discussion of the differences that were apparent between the two techniques.

Marginal/Interlineal Comments
Interesting to note, the Batt and Wilson (2008) study discussed previously only considers the effect of speech recognition technology on the *end comment*—typically a summative closing statement-- that teachers provide to students. Alternatively, *interlineal or marginal comments*—feedback provided to students throughout the body of an essay—are quite different. While both types of feedback are usually used in concert by writing teachers, the often graphical and interactive nature of interlineal or marginal comments may seem poorly suited to text-only speech recognition. For example, when I handwrite marginal comments on a student's paper, I (like many writing teachers) regularly use circles, underlines, and arrows throughout the body of an essay to highlight issues for writing students to address.

Obviously, these sorts of graphical annotations are not possible with text-only speech recognition software. Nevertheless, I wanted to test the feasibility of providing students with interlineal comments via speech recognition; I wondered what form those comments would take. I found that, essentially, in instances where a handwritten circle and arrow might do the

trick, I instead needed to provide guidance in wholly narrative terms when speaking my feedback aloud. I accomplished this by inserting text in all capital letters (for ease of identification) within the body of the paper itself explaining my suggestions in concrete terms. For example, instead of circling and drawing lines to indicate problems, I would insert text by speaking aloud a correction such as: "Your pronoun and antecedent do not match in number. Make them either both singular or plural."

Regardless that speech recognition technology has become surprisingly accurate even when the user speaks at a relatively fast pace, there is no question that this latter approach took more time than simply drawing a circle around the sentence and including a note—although I did prove to myself that useful, relevant interlineal commentary can be provided to students using speech recognition. In addition, I realized that this approach did not require the effort of handwriting and attention to legibility since the comments were typed directly into the essay with minimal errors. I'll reflect on these competing realities—more time spent but seemingly less effort--in the conclusion.

Length of Commentary
I was also interested to know if the length of my feedback—both interlineal and end comments—changed when handwriting versus speaking aloud. Calculating average word counts for each mode over the span of the same 40 essays, I found: Interlineal comments provided by hand averaged 11 words in length, with the smallest comment being 1 word long and the largest being 53 words long. End comments provided by hand averaged 44 words, with the smallest being 25 words in length and the largest being 67 words. Alternatively, interlineal comments typed into the drafts via speech recognition averaged 26 words in length, with the smallest block of text containing 4 words and the largest block of text containing 114 words. End comments provided this same way averaged 152 words over the span of 20 essays, with 102 words and 205 words being the smallest and largest blocks of text.

In rough summation, when I used speech recognition software, the interlineal feedback provided to students was slightly more than double the length than when commentary was handwritten. End comments provided via speech recognition were roughly more than three times as long on average as handwritten comments.

While these results are not entirely surprising to me, the numbers do not automatically mean "more is better." As mentioned previously, graphical annotations widely used in handwritten commentary can often effectively substitute for words, which most likely accounts for the disparity. In addition, the relative speed at which feedback can be spoken allows for much more commentary to be included; in comparison, handwriting seems woefully slow. Nevertheless, I will say that the non-graphical, text-based nature of using speech recognition technology to provide comments to students (in other words, requiring students to read through a teacher's text-based commentary, understand it, and transform it into better writing, rather than interpreting circles and arrows) is commensurate with the overall goals of a composition class—namely the explicit focus on written text.

Directive versus Facilitative Tone

Richard Straub's 1996 article "The Concept of Control in Teacher Response: Defining the Varieties of 'Directive' and 'Facilitative' Commentary" was seminal in coding the various types of written feedback that composition teachers provide to students. Straub found that "directive" feedback generally exerted more control over student writing and "facilitative" commentary was less controlling and more suggestive in nature. However, he argues that neither is fundamentally right or wrong; Straub maintains that these various types of commentary have some value--different forms of feedback are used in different situations and are governed by many forces.

I wanted to know how utilizing speech-recognition technology changed the directive or facilitative tone of my commentary. When handwriting comments, I often will make specific corrections or changes while only providing short (and sometimes no) explanations for these changes. In Straub's terms, this is decidedly directive. For example, I will directly insert missing punctuation, circle stylistic errors, and cross out problematic words—while only offering two- or three-word explanations at most. This is evidenced by the smaller word counts discussed earlier. However, when using speech-recognition, by necessity, most of my comments were conversational and highly explanatory in nature. Therefore, I found them to be more facilitative in tone, albeit lengthier. I believe some of this effect was due to the spoken nature of the comments. For example, when making revision suggestions verbally, I would often include phrases such as "Here you might think about restating your point," "I just want to throw out an alternative," or "The last half of this sentence gets difficult for me to read." In addition, I found myself often asking sometimes lengthy questions to the student: "Could you consider combining these last two sentences into one?" and "Do you think you've made this error elsewhere in your paper? Should you check it?" Handwriting or typing these longer, more facilitative responses would generally be out of the question since they would take an inordinate amount of time and energy.

A Concluding Paradox: More Time, Less Effort?

This investigation has left me with a paradox: Commensurate with previous findings (Batt & Wilson, 2008), using speech recognition software did not save me any time regarding the Herculean task of providing students with individualized feedback on their drafts. In fact, it may have slightly lengthened the time I spent with each essay, though it wasn't deleterious. Contrarily, the amount of text-based feedback provided to students increased two-fold. While this in itself is not an automatic benefit, providing this increased amount of feedback seemed to require less effort (physically, mentally), likely reflecting the relative ease of verbalizing feedback versus mechanically writing it down.

At first, I thought this feeling of ease, the feeling of this task being less burdensome, might be attributed to the novelty of the approach. However, after two years of implementation, the newness factor had completely disappeared, and I would still argue that using speech recognition felt easier, more natural, and more personal than I would have imagined. Also, as previously noted, handwriting of course requires a certain amount of precision and attention to legibility, even when only writing short phrases. Similarly, typing requires some level of

mechanical precision as well. None of that effort is exerted when simply speaking aloud, and I personally found the latest iteration of speech recognition to be surprisingly accurate. For the present, I will continue using this technique to provide interlineal and end comments to students on their drafts.

References

Alapetite, A., Andersen, B., & Hertzum, M. (2009, January). Acceptance of speech recognition by physicians: A survey of expectations, experiences, and social influence. *International Journal of Human-Computer Studies*, *67*(1), 36-49. doi:10.1016/j.ijhcs.2008.08.004.

Barbetta, P. M., & Spears-Bunton, L. A. (2007, March). Learning to write: Technology for students with disabilities in secondary inclusive classrooms. *English Journal*, *94*(4), 86-93.

Batt, T., & Wilson, S. (2008). A study of voice recognition software as a tool for teacher response [Electronic version]. *Computers and Composition*, *25*, 165-181. doi:10.1016/j.compcom.2008.01.004

Carter-Wesley, J. (2009, July). Voice recognition dictation for nurses [Electronic version]. *Journal of Nursing Administration*, *39*(7/8), 310-312.

EdX. 14 June 2013. https://www.edx.org.

Haswell, R. (2008). Teaching of writing in higher education. In Bazerman Charles (Ed.). Handbook of research on writing: History, society, school, individual, text (pp. 331-346). New York: Lawrence Erlbaum Associates.

History of speech recognition. LumenVox, 1 Jan. 2010. Web. 28 Jan. 2013. http://www.lumenvox.com/resources/tips/historyOfSpeechRecognition.aspx.

Hesse, D. (2013, May). Grading writing: The art and science — and why computers can't do it. *Washington Post.* Retrieved June 11, 2013, from http://www.washingtonpost.com.

Honeycutt, L. (2004). Literacy and the writing voice: The intersection of culture and technology in dictation. *Journal of Business and Technical Communication, 18*(3), 294-327.

Markoff, J. (2013, April). Essay-grading software offers professors a break. *New York Times.* Retrieved May 29, 2013, from http://www.nytimes.com.

Mellen, C., & Sommers, J. (2003). Audiotaped response and the two-year-campus writing classroom: The two-sided desk, the "guy with the ax," and the chirping birds. *Teaching English in the Two-Year College, 31(*1), 25-39.

Merry, S., & Orsmond, P. (2008). Students' attitudes to and usage of academic feedback provided via audio files. Available: http://www.bioscience.heacademy.ac.uk/journal/vol11/beej-11-3.aspx [Accessed: January 15, 2013].

Mike, D. (2010). "Speech-driven" technologies to reduce faculty workload in online higher education. In *Proceedings of World Conference on Educational Multimedia, Hypermedia, and Telecommunications 2010*, 3661-3665.

National Council for Teachers of English (2012). Statement of principles and standards for the postsecondary teaching of writing. In *CCCC Position Statement*. Retrieved January 28, 2013.

O'Shaughnessy, D. (2008, October). Automatic speech recognition: History, methods and challenges [Electronic version]. *Pattern Recognition*, *41*(10), 2965-2979. doi:10.1016/j.patcog.2008.05.008.

Pinola, M. (2011, November 2). Speech recognition through the decades: How we ended up with Siri. *PC World.* Retrieved January 28, 2013.

Roberts, S. (2008). Pod casting feedback to students: Students' perceptions of effectiveness. Available: http://www.jmu.ac.uk/lid/lid_docs/p44_CETL_Journal_No2.pdf [Accessed: January 15, 2013].

Rotheram, B. (2007). Using an MP3 recorder to give feedback on student assignments. *Educational Developments, 8*(2), 78-88.

Sherman, S. L. (2008). *Facing the dragon: Boy writers speak up using voice-recognition technology* (Doctoral dissertation, University of Colorado-Denver, Denver). Retrieved January 15, 2013, from ProQuest Dissertations and Theses.

Silió, M., & Barbetta P. (2010). The effects of word prediction and text-to-speech technologies on the narrative writing skills of Hispanic students with specific learning disabilities. *Journal of Special Education Technology, 25*(4), 17-32.

Still, B. (2006). Talking to students: Embedded voice commenting as a tool for critiquing student writing. *Journal of Business and technical Communication, 20*(4), 223-251.

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

Strauss, V. (2013, April). Can computers really grade essay tests? *Washington Post.* Retrieved May 25, 2013, from http://www.washingtonpost.com.