Speech Technologies to Reduce Faculty Workload in Online Instruction

Presentation to The WNY Consortium of Higher Education, October 1, 2010

Dennis G. Mike, Ph.D.
Associate Professor
Canisius College, Buffalo, NY, USA
miked@canisius.edu

Session Abstract

It is commonly acknowledged that the provision of quality online instruction is more time consuming than face-to-face instruction. For this reason, faculty workload has emerged as an important concern in online higher education. This paper will focus specifically on the use of "speech-driven" technologies to reduce faculty workload in online instruction without sacrificing quality. Voice recognition software (e.g., Dragon Naturally Speaking 10, Preferred Edition) allows faculty to more easily participate in discussion boards and respond to student email. In addition, the "voice embedding" feature of certain software (e.g., Adobe Acrobat Extended Pro 9, Camtasia) allows the instructor to "speak" his/her feedback to student writing and to provide additional elaboration on instructions to students. Application of these programs specific to tasks associated with online instruction will be described, as well as the technical considerations of program use. Last, equity and access issues relating to the use of such technology will be addressed.

Session Outline

INTRODUCTION
- Online faculty workload and the importance of reducing it whenever practical
- Brief description of my own use of “speech-driven” technologies

VOICE RECOGNITION SOFTWARE
- Overview
- Demonstration

SPEECH “EMBEDDING” SOFTWARE
- Overview
- Demonstration

SCREEN CAPTURE SOFTWARE
- Overview
- Demonstration

TEXT-TO-SPEECH SOFTWARE
- Overview
- Demonstration

TECHNICAL CONSIDERATIONS
- Memory, speed and storage capacity of computer
- Microphones
- PC vs. Mac

ACCESSIBILITY CONSIDERATIONS
- The need to (among other things) provide a transcript of instructor speech in online courses
    - how speech-driven technology can help

CONCLUSION / Q&A

Speech Technologies to Reduce Faculty Workload in Online Instruction

Dennis G. Mike, Ph.D.
Associate Professor
Canisius College, Buffalo, NY, USA
miked@canisius.edu

Introduction

It has been clearly established that, for faculty, providing quality online post-secondary instruction is far more time-consuming than is the case with face-to-face instruction (Cavanaugh, 2005; Hartman, Dziuban, & Moskal, 2000). Specifically, it has been found that responding to student work and answering questions requires a greater time commitment (Sellani & Harrington, 2002), possibly because such activity is highly writing intensive for the instructor. Because all communication with students is conducted in an electronic environment, the role of text in online instruction has assumed greater import than would be the case with face-to-face classes, where much information can be conveyed verbally. The increased time commitment has been found to be a major barrier in attracting faculty to online instruction and keeping them involved (Schifter, 2000; Berge et. al., 2002; O'Quinn & Corry, 2002).

Because greater faculty workload in online instruction is problematic, any option that would have the effect of reducing it, while still maintaining quality, is worth noting. The purpose of this paper is to suggest that "speech driven" technologies can save the instructor time and effort by substituting spoken language for textual communication for certain tasks. The four technologies that will be described below are 1) voice recognition software, 2) speech "embedding" software, 3) screen capture software, and 4) text-to-speech software. For each of these, a personal example of use is provided, emphasizing the online task(s) that each program addresses. Last, technical considerations and the quasi-legal concerns of equity and access are addressed.

Examples are derived from online courses taught by the author at Canisius College, Buffalo, NY, USA. It is beyond the scope of this article to convey a full description of each genre of software discussed. However, it is hoped that description of specific programs cited below may generalize for the reader to other software within the same genre.

Voice Recognition Software

Voice recognition software essentially allows the user to "speak" to the computer. The program converts the user's speech to text, inputting it as if it were typed. Such software has existed for more than 20 years. However, advances in both hardware (e.g., memory capacity, storage capacity and processor speed) and software design have led to this genre of software "coming of age" relatively recently.

The program used by the author for online instruction is Dragon Naturally Speaking Preferred Edition, version 10. Initial preparation (i.e., "training") of this program for takes approximately 15 minutes. In training, the user reads text from the computer screen, allowing the software to become familiar with the individual's speech mannerisms. An additional aspect of training is to direct the software to directories where the user has stored previously written text. By "washing through" these files (similar to a spell checker), the software learns the manner in which the user engages in discourse, as well as the specialized vocabulary of his/her field of study. Training the software is relatively simple. Far more problematic is acclimating the user to the conventions of voice recognition software. More precise enunciation is required than is the case with conversational speech. In addition, punctuation must be "spoken out loud." A reasonable expectation is that the program can be used with over 98% accuracy within a week of first use. While, as first glance, 98% appears impressive (and, considering the difficulty of the technical task, it is), the prospective user should keep in mind that this means that two words or phrases out of 100 will be inaccurately processed. By comparison, after several months of regular use, my

personal accuracy measure is 99.7%. While this is certainly good enough to justify continued use of the program, when I dictate something that requires precise accuracy, a great deal of care must be taken to proofread for errors made by the program. However, with respect to the tasks associated with online instruction, precise accuracy is *not* always required.

There are two tasks associated with online instruction for which I use voice recognition software: communication with my students via email and participation in online discussion forums. To prepare my students, I inform them ahead of time that I use Dragon Naturally Speaking for these two purposes. By the same token, I have prepared a Camtasia video(see below) to demonstrates my use of the software. The primary reason for acquainting my students with Dragon is to prepare them for the eventuality that occasionally my messages may not make total sense due to errors in processing. I've informed my students that I do not carefully proofread email or discussion forum posts and that, should a message of mine should require clarification, they are to contact me. This is rarely necessary.

In the Fall 2009 semester, for one course alone, I initiated 523 email messages. This figure does not approach the total number of email messages that were sent and received throughout the semester, as it fails to account for my replies to messages which I did not initiate or for more involved email threads. I type at approximately 70 words per minute. I can dictate text through Dragon at between 150 and 175 words per minute. The savings in time, for both email and discussion board activity, are evident.

Speech "Embedding" Software

By "speech embedding" software (my own expression – not referenced elsewhere), I am referring to that software which allows users to input their voice directly into a computer file (most typically a document). Here I make the distinction between this type of software (which essentially records the user's speech as actual sound) and voice recognition software (which digitally converts the user's speech to text). To clarify this distinction further, with speech embedded software, the person receiving the file hears the actual sound of the instructor's voice. When voice recognition software is employed, the user only sees text.

Many software programs allow the user to embed sound, including speech. However, for practical use in online instruction, it is essential that the recording process be as quick and simple as possible. In some cases, there are technical barriers to such use, even if the program allows for sound to be embedded. For example, the instructional authoring tool SoftChalk allows for sound to be embedded, but, in order for this to happen, the user must use *another* program (e.g., Audacity) to record speech and then save it as a separate file (one file per utterance). This file then needs to be "manually" added to the SoftChalk page. This process requires so many steps and added complexity (e.g., a large number of files with which to content) that it would not, in fact, save the instructor a significant amount of time.

In addition, there is a technical consideration regarding file size. Files must be kept reasonably small because students need to access them online. Large files take an inordinate amount of time to access or download. While most word processing programs (e.g., Microsoft Word) allow you to embed speech directly into a document, the result is often a huge file, problematic for students with slower than average Internet access.

The program I use for embedding speech is Adobe Acrobat Extended Pro, version 9. This program has an option that allows the user to record speech through only a single menu command. In addition, the resultant file size of the document, while larger than a file without embedded speech, is much smaller than would be the case with virtually any word processing program on the market.

There are two tasks associated with online instruction for which I use this Adobe Acrobat Extended Pro 9. The first is responding to student written work. Students submit their papers to me in Rich Text Format (RTF). I use Microsoft Word to open the file and convert it to Portable Document Format (PDF). In the conversion process, the file is automatically loaded into Adobe Acrobat Extended Pro 9. From that program, as I read through the student's work, if I have a comment to make, I access a single command and "speak" the comment into the computer. The comment is saved in the form of a blue speaker icon that appears in the margin. All of my feedback is entered in this fashion; there are no written comments whatsoever. When the student receives the file back, s/he opens it Adobe Reader version 9 (a free download) and then clicks on the speaker icons in order to

hear my feedback.

The second task associated with online instruction for which I use Adobe Acrobat Extended Pro 9 is elaborating upon written directions for assignments. As cited above, teaching in an online environment requires the instructor to anticipate and address as many sources of student confusion as possible. However, I've found that, if my written directions for assignments are too involved, the result document is longer than many students will take the time to carefully read. By keeping the written directions to a page or two, and supplementing them with spoken comments embedded in the margins, students are more likely to access all of the requisite instructions. In this way, I can provide the same informal elaboration that I would if I was engaging with a face-to-face class. In addition, students have stated in course evaluations that they feel more of a personal connection with me as an instructor because they regularly hear the sound of my voice.

Screen Capture Software

Screen capture software is a genre which allows the user to make videos of screen activity while narrating. Technically, this might include "still image" screen captures generated by hitting "Alt-Print Screen" on a PC or using programs such as TechSmith's SnagIt. However, for the purposes of this paper, I am referring to the use of software that allows the *dynamic* capture not only of a static computer screen, but of the computer screen *activity*. There are several companies that offer such software (e.g., Camtasia Studio, WM Capture 3.0, liteCam). TechSmith offers this capability in a free tool called Jing which saves the video file online where it can be readily accessed by users. (A slightly more capable version called Jing Pro is available at a nominal yearly cost.) However, Jing and Jing Pro limit the length of video to five minutes. To circumvent this limitation, the user must adopt a more full featured commercial program.

The software program that I use for capturing screen activity with speech narration is TechSmith's Camtasia Studio version 7. This program is the "big brother" of the aforementioned Jing and Jing Pro.

The task associated with online instruction for which I use this program is identical to a task cited above: the elaboration of instructions to students. I previously cited the use Adobe Acrobat Extended Pro 9 for this purpose, which I would employ to provide elaboration within a single document. However, if I wish to provide more global elaboration (e.g., to the structure of an entire online module within ANGEL), I have found Camtasia to be the more appropriate tool. Examples of Camtasia videos I've created for my classes can be found at:

http://www.screencast.com/users/DennisMike/folders/Default/media/ed3d4103-c904-43c1-86fa-22e6291642e2

http://www.screencast.com/users/DennisMike/folders/Default/media/9b820399-c998-46d7-beaf-ccae5c51b9e2

Text-to-Speech Software

Up until now, the software I have described has involved speech *input*. The opposite of this is speech *output*, one form of which is text-to-speech software. Such software uses algorithms to digitally convert text into computer synthesized speech. The result is robotic sounding to the ear, but understandable. Text and screen reader software for those with visual or learning disabilities are examples of this genre of software.

The software program that I use for text-to-speech is Texthelp's Read and Write Gold, version 9. This software is "transparent" in that it is not intended to be used as a standalone, but, instead, to be employed in conjunction with other software. For example, when used with an Internet browser, it can "read" the screen.

The task associated with online instruction for which I use this program is reading student work, but only when I do not have time to read, but do have time to listen. An example of this use (which is thankfully limited)

is an occasion a few months ago when I had to drive for 20 hours straight while during an online instructional module. The students had already submitted written work, but I didn't have time to read it. Instead, I loaded each file into Read and Write Gold and set the program "to read aloud." Using another piece of software (High Criteria's Total Recorder), I recorded the resultant synthesized speech, burning it to CD. Consequently, I was able to listen to the students' writing on CD while driving. Furthermore, when I had a comment to make in response to a student's work, I would pause the playback and record my comment into a handheld digital voice recorder using a high quality headset. When I returned home, I plugged this recorder into my computer and ran the files through Dragon Naturally Speaking 10 (Preferred Edition), the aforementioned voice recognition software. Dragon then converted the recording of my voice comments to text, as if I had typed them. Students received these comments as feedback to their work. This example is notable because it illustrates an important point: what would have normally been purely a *literacy* activity (i.e., reading and writing) had become, through the use of speech technologies, an *oralcy* activity (i.e., listening and speaking). I would suggest that the line between literacy and oralcy is blurring, largely due to advent of this kind of technology.

Technical Considerations

Speech technologies work best on relatively powerful computers. Voice recognition software, in particular, takes a good deal of computing power in that the entire lexicon of the language has to be loaded into active memory. If, on the other hand, the program needed to retrieve components of the language from storage (i.e., the computer's hard drive) instead of memory, the technical process would be slowed to such an extent that accuracy would suffer and the program would "lag." Trying to use Dragon or Adobe Acrobat Extended Pro 9 on older computers would likely be a frustrating experience for any user.

By the same token, it is important to use the best quality microphone one can afford. Some versions of Dragon come with headsets bundled in the package. I would recommend against using such microphones – the result will be reduced accuracy in voice recognition and, with respect to recording of the instructor's speech in speech embedding software, sound quality that is less than optimal. A good quality microphone ($50-$100) is well worth the investment.

Accessibility Considerations

It is beyond the scope of this paper to fully address the laws governing the education of those with disabilities. Suffice to say that every student has a legal right to have instruction presented in such a way that it is readily accessible to him or her. Heavy use of the speech-driven technologies discussed in this paper could gravely disadvantage those students with hearing impairments. The solution is to provide a transcript for everything that presented aurally, including the instructor's speech.

I personally struggled with this consideration because attending to it was contrary to the reason that I had adopted these technologies in the first place: to saving time and reduce workload. The requirement of creating a written transcript of everything that I recorded would have been so time consuming as to be prohibitive. While this is not an issue with voice recognition software (because the output is textual), it *is* an issue with sound embedding software such as Adobe Acrobat Extended Pro 9 or Camtasia Studio. A solution to this dilemma was readily available and, I'm happy to say, cost effective in terms of time. After recording speech into either of the two aforementioned programs, I now play it back through the computer speakers with Dragon Naturally Speaking enabled. The voice recognition program then generates a text file from the sound which is, for all intents and purposes, a written transcript of my recorded speech. Admittedly, the resultant text does have to be proofread carefully for errors. However, the process is sufficiently accurate that it is well worth the time and trouble required.

Conclusion

The central premise of this article is that the use of speech-driven technologies (specifically the programs cited above) can reduce faculty workload in post-secondary online education while maintaining, perhaps even enhancing, quality of instruction. It stands to reason that online instructors should be the early adopters of such software and stay current with its use. Online instructors teach in an electronic environment, one that fosters effective and creative use of technology. By employing speech-driven technology in our own instruction, we not only save ourselves valuable time – we also model the appropriate mindset regarding instructional technology for our students.

If I can provide additional support to any reader considering the adoption of the technology described above, please feel free to contact me via email.

References

Berge, Z. L., Muilenburg L.Y., Haneghan, J. (2002, May 15). Barriers to distance education and training: Survey results. *eModerators*, Retrieved May 1, 2010 from http://www.emoderators.com/barriers/barriers2002.shtml

Cavanaugh, J. (2005). Teaching online: A time comparison. *Online Journal of Distance Learning Administration*, 7 (1). Retrieved May 1, 2010 from http://www.westga.edu/~distance/ojdla/spring81/cavanaugh81.htm

Hartman, J., Dziuban, C., & Moskal, P. (2000). Faculty Satisfaction in ALNs: A Dependent or Independent Variable?. *Journal of Asynchronous Learning Networks*, 4 (3), 155-179.

O'Quinn, L., & Corry, M. (2002). Factors that deter faculty from participating in distance education. *Online Journal of Distance Learning Administration*, 5 (4), 1-18. Retrieved May 1, 2010 from http://www.westga.edu/~distance/ojdla/winter54/Quinn54.htm

Schifter, C. C. (2000). Compensation models in distance education. *Online Journal of Distance Learning Administration*, 3 (1), 1-7. Retrieved May 1, 2010 from http://www.westga.edu/~distance/ojdla/spring31/schifter31.html

Sellani, R. J., & Harrington, W. (2002). Addressing administrator/faculty conflict in an academic online environment. *Internet and Higher Education*, 5, 131-145.