The Potential for Use of Voice Recognition Software in Appraisal of Oral History Tapes

Presented at the Association for Recorded Sound Collections (ARSC) Annual Conference, May 4, 2007

Sonia Yaco
2343 W. Lawn Avenue
Madison, WI 53711
(608) 231-1318
yaco@wisc.edu

Last year, a potential donor spoke with me about whether the Wisconsin Historical Society Archives would be interested in receiving a donation of a series of oral history tapes. Although the material was valuable, this was a thorny question because working with spoken word material is extremely time-consuming. Everyone involved in using audio materials, from the archivists who appraise, access, process and describe them to the researchers who use them, can only access the content by listening to them in real-time. Unlike manuscripts, it is impossible to skim through a stack of audio recordings. The potential donor had been told that processing her tapes would be prohibitively costly. I explained to her why an institution might be reluctant to accept tapes. Current appraisal is done by archivists either at a macro level, whole collections at a time, or at a micro level, series by series.[1] Either way, if tapes come into an archive without a description of their content, archivists must take the time to listen to at least a sampling of them. Archivists, or their staff, typically create a written description of a tape's contents during the appraisal process that will be used in the collection's catalog record, if the tapes are retained.[2] Some institutions' simply log the creator, subject and date for each tape. Others abstract the contents of the tape at regular time intervals or create a narrative summary of the entire tape.[3] The donor's query set me to thinking about how to make not only her historically valuable tapes, but also those of others, more usable by archivists and researchers.

As a regular user of software that transcribes my spoken words into written text, I thought it might be feasible to use this software to transcribe oral histories as well. The software, known as voice-recognition technology or automatic speech recognition, lets me speak into a headset attached to my computer and have those words appear in almost any computer application. As a reference archivist, I use voice recognition software to create container lists in Word, enter data into Microsoft Access databases, browse the Web and answer patrons' e-mail requests. Sound tapes, which typically come into the Archives on analog cassette or reel-to-reel tape, could be converted to digital format, and then automatically transcribed with this software. While voice recognition software is often used by people like me who are disabled, it is increasingly used in other situations. Two common examples are Directory assistance and help line call routing software.[4]

Oral history tapes are cumbersome for archivists and researchers to appraise and use. Having computers automatically create transcripts could

greatly increase the speed at which archivists could process tapes, increase the richness of their description and make inexpensive transcripts available to researchers.

The use of voice recognition software to create transcripts could also solve three problems that reduce the utilization of spoken word material. The first problem is that because sound recordings are time intensive for archivists to work with there are delays in making tapes available to researchers. For the same reason, descriptions of the contents of spoken word tapes tend to be sparse, making collections relatively inaccessible to researchers. If computers rather than humans could create transcripts of spoken material, these transcriptions could be included in EAD and MARC catalog descriptions of collections.[5] This would increase the visibility and accessibility to researchers on and off-site. Thirdly, researchers prefer reading oral history transcripts over listening to the actual tape itself. Untranscribed tapes are rarely used in archival collections, which means some potentially valuable material is not available. Even when the description of an institutions' spoken material is adequate, researchers prefer written transcripts and are very reluctant to listen to oral history tapes. According to Dr. Carolyn Mattern of the Wisconsin Historical Society Archives, only "the most dogged researcher" will actually listen to untranscribed tapes.[6] With input from my colleagues, university librarians and archivists, I conducted a simple pilot study on two types of voice recognition software. These explorations give some insight into some of the potential benefits and some of the problems in using voice recognition software for oral history tapes.

The history of automatic speech recognition

The terms automatic speech recognition, voice-recognition technology and voice recognition software are used synonymously. "Transcription" refers to turning spoken words into written text. The two software packages I evaluated both transcribe, but the software behind automatic speech recognition in call routing, does not - it merely recognizes speech to make branching decisions.

Voice recognition technology has a much longer history than most people are aware. While it is beyond the scope of this paper to describe the full history of voice recognition technology, providing a synopsis of its development will be useful for understanding this pilot study[7]. It is arguable that Alexander Graham Bell's 1876 patent for the telephone was the first step towards voice-recognition software. After all, the telephone is a device that translates sound waves from the human voice into modulated electronic current. However, it may be more

accurate to consider the telephone as a necessary precursor to voice recognition technology. The modern conceptual groundwork for voice recognition technology was laid by Harvey Fletcher and Homer Dudley of Bell Labs in the 1930s. The World's Fair of 1939 featured Dudley's voice synthesizer called the Voder, (Voice Operating Demonstrator). By the 1950s, scientists have begun to figure out how to recognize the sound waves made for single digits by a single speaker. They created graphs of the sound waves produced by the vocalization of the digit '9', for instance. Next, single syllable words spoken by a single speaker were tackled. In 1959, Fry and Denes came up with a statistical approach for recognizing speech. Statistical methodology makes it easier to recognize words in a sentence than individual words. In the 1970s, the concept of pattern recognition, "the automated identification of shapes or forms or patterns of speech" [8] became important. This is also when the Department of Defense became interested in voice recognition technology. AT&T and IBM began two different branches of research, reflecting their respective commercial interests and customer base. AT&T, a telephone company, developed research on software that would be able to recognize multiple speakers, known as *speaker independent*. IBM, a computer and typewriter manufacturer, started work on a "voice activated typewriter" (VAT). Designed to recognize the voice of a specific speaker, it is a type of software known as *speaker dependent*. The VAT project featured a large vocabulary and emphasized the structure of language i.e. the statistical probability of a particular word following another word.

The statistical methodology was refined in the 1980s with a new emphasis on reduction of recognition error rates. Another important advance made during this time was software that had the ability to recognize continuous speech rather than requiring the user to pause briefly between each word. By the mid-1990s, some software was able to distinguish between speech that should be transcribed and spoken commands such as, "Turn italics on," and "Close File."

The two components necessary for automatic voice recognition are the ability to identify acoustic phonemes[9], and the use of statistical probability to predict the most likely word. Although these basic components had been developed by the 1970s, it was not until the 1990s with the increase in computing ability and hard disk size of desktop computers that voice-recognition software on personal computers became feasible.

Despite the 70 year history of this technology there is little literature in the library or archives world on automatic voice-recognition.[10] Zumalt evaluated two types of voice-recognition software for use in cataloging[11]. Bertuca described

OCLC's use of voice recognition software in cataloging tasks to avoid repetitive stress injuries[12].

How does it work?

The mechanism for using voice recognition software on a personal computer begins with a user speaking into the microphone attached to the computer. The computer's sound card translates the analog sound to digital. The voice-recognition software compares the acoustic signal, now in digital form, to a vocabulary lookup table created by the software developers that maps sounds to words. The lookup table for instance, would equate the vocalization 'bat' with the word 'bat', 'caset' with 'cassette'. Large vocabulary tables have been built for specialized uses such as switchboards, radiology practices and customer call centers.

The first component of automatic speech recognition is acoustic phoneme recognition, that is, recognizing spoken sounds as words. Speaker dependent software adjusts its acoustical recognition by being calibrated or "trained" for an individual user's voice. Dragon NaturallySpeaking, until the most recent version, had to be trained to a user's voice by reading set texts aloud for five to fifteen minutes so that the software can be calibrated to that user's voice. [13]As Dragon NaturallySpeaking is used, its accuracy increases if the user utilizes its "Correct That" feature when the software incorrectly transcribes a word.

The second component of voice recognition technology is statistical language modeling methodology, a tool which weighs the probability that any given word will appear in a sentence. With this tool, algorithms, or formulas, allow voice recognition technology to select among similar sounding words for the most statistically likely word. This comes into play when the computer must choose between two homonyms such as 'eight' and 'ate'. When I dictate the sentence, "I ate the apple." Dragon NaturallySpeaking chooses 'ate' from that pair of homonyms because there is a higher probability that a verb will appear after the word 'I' than a number. Dragon NaturallySpeaking's built-in language model can also be supplemented by the Accuracy Optimizer feature, which examines the writing style in the Word documents and e-mails on the user's computer hard disk.

There is a trade-off between speed and accuracy with voice recognition software. Dragon NaturallySpeaking prioritizes speed of recognition for real-time users over accuracy in software.[14] Conversely, AudioMining, the second software I evaluated, which is used in batch mode, is slower but likely to be more accurate. AudioMining is a sophisticated voice recognition product that is

designed for use in legal, medical, financial and technology fields. A typical application would be a medical practice with multiple physicians, in multiple locations, dictating patient charting information. Physicians create audio files of their dictation and save them to a network drive. Then on some regular schedule, AudioMining transcribes groups of these audio files. The accuracy of the transcription is much more important than is speed in this case.

In the ten years since the article, "Dictating This Article to My Computer: Automatic Speech-Recognition Is Coming of Age"[15], voice recognition technology has made huge strides. Anecdotal evidence comes from my experience in 1998 using Dragon NaturallySpeaking to transcribe a diary written by a high school educated woman versus my experience in 2006 with transcribing folder titles written by a Ph.D. researcher. In 1998, Dragon NaturallySpeaking did a passable job with the common words used in the diary. In 2006, Dragon NaturallySpeaking did an excellent job recognizing the specialized words used on the folders.

Quantitative evidence of the advances Dragon NaturallySpeaking has made comes from tests done by Joseph Zumalt at the University of Illinois. When users trained their voices with the minimum allowable training time on three successive versions of Dragon NaturallySpeaking (6.0, 7.3, 8.0), Zumalt found the errors for reading passages dropped with each version. The most substantial decrease in error rate among his sample texts was for the Gettysburg address, which dropped from twenty-six errors to five errors. Zumalt further found that with additional training, Dragon NaturallySpeaking's error rate on the Gettysburg address dropped from five errors to two errors.[16]

Methodology

The spoken material I used as test data in this study was a series of nine cassette tapes containing seventeen interviews conducted by University of Wisconsin-Madison emerita Professor June Weisberger in 2004. Professor Weisberger interviewed eight people about their role in the creation of the Marital Property Reform Bill that passed in 1984. The nine speakers on the tapes are college educated Caucasians ranging in age from 60 to 80. Eight of the speakers are women and one is a man. They were raised in the Midwestern or eastern part of the United States. For each of these speakers, English is their first language.

Professor Weisberger recorded these interviews on a high-quality Marantz cassette tape recorder that created clear recordings with little background noise. I

converted the analog cassette tapes to digital wave files. To do this I connected the Marantz tape recorder to a Windows XP machine with Y stereo cables. I used the Line Out port on the tape recorder and the Line In port on the computer. I used a trial version of the software RipEditBurn Plus using the Mono, 16-bit settings, with a sampling rate of 44100 HZ.

I evaluated two types of voice recognition software – speaker dependent, designed for use by a singe user, and speaker independent, intended for multiple users. I first tested speaker dependent software. There are several such packages currently available for consumers including Dragon NaturallySpeaking, IBM's ViaVoice and Microsoft Office's Speech Recognition. I chose Dragon NaturallySpeaking because of its reputation for high accuracy and my familiarity with it.

Since Dragon Naturally Speaking requires calibration or training, I trained it with my voice. I did not have Professor Weisberger calibrate the software with her voice. This would have increased the accuracy of transcribing her interview questions, but not the interviewee's answers. As appraisal of spoken material typically happens years after it is recorded, it would seldom be possible to have any of the speakers calibrate software.

Dragon NaturallySpeaking is typically utilized in real-time by a user speaking into a microphone directly attached to a computer. To facilitate the use of portable recording devices, Dragon NaturallySpeaking has a feature, "Transcribe Recording," that allows digital sound files created with an external device to be processed in batch mode. I utilized this feature. Once I converted the original seventeen interviews to digital files, I used this batch mode feature to transcribe each of the files into separate Microsoft Word documents.

I reviewed archives and library literature looking for protocols for evaluating the voice-recognition software used in past projects. I found nothing on this aspect of voice-recognition software prior to the beginning of my initial evaluation so I developed my own protocol. First, I read the contents of the transcripts for each of the tapes. Secondly, I asked a sound archivist[17] to examine a several Dragon NaturallySpeaking produced transcripts. I had her evaluate their usefulness for appraising the contents. Thirdly, I selected the first 60-second segment of a tape, manually produced a transcript and compared it to the transcript created by Dragon NaturallySpeaking.

Several months after I completed my evaluation of Dragon NaturallySpeaking, I began my evaluation of the second type of voice-recognition software -- speaker independent. Like speaker dependent software, it

recognizes words based on acoustic phoneme recognition and uses statistical probability to equate vocalizations with words. This type of software does not rely on calibration for accuracy. It is designed for multiple users so is potentially more appropriate for creating transcripts from oral history tapes than speaker dependent software.

A product called AudioMining was released in 2005 by Nuance, the same company that makes Dragon NaturallySpeaking. The software creates keyword indexes with timestamps from audio content. AudioMining has a different focus then Dragon NaturallySpeaking. It is intended for use by a team of people all working at the same firm, sharing a common vocabulary. It is designed for professions that have complicated specialized vocabulary, such as radiology, so it has a feature that allows specialized vocabularies to be constructed before voice files are processed. The product ships with sixteen different pre-built vocabularies. Given these characteristics, AudioMining should be well suited to recognizing multiple voices in a series of interviews on similar topics.

AudioMining output format has great potential for creating better intellectual and physical access for researchers to the content of audio files. In addition to creating a transcription of the voice files, AudioMining also creates an index of keywords and creates XML files that link those keywords to time stamps in the voice files. This means that the phrase "Marital Property Reform" could be linked to the exact location in the sound file that it appears.

The AudioMining purchase price reflects the fact that it is intended for commercial use.[18] The installation process requires a fair amount of technical expertise because it is intended to be installed on a server, not an individual workstation. Nuance granted me a one-month Standard Development Kit evaluation version of this software for the purposes of this study. My evaluation could have been more thorough with a longer testing period and if I had had access to the full AudioMining package and technical support.

After a somewhat complicated software installation process, the seventeen original digital interview files were run through AudioMining. This took several days, as the transcription process is much slower with Dragon NaturallySpeaking.

To evaluate the accuracy of the AudioMining transcription, I read the files AudioMining created. For tape 6, side A, I listened to the interview with the transcript in front of me and struck out any words that were incorrect. For tape 1, side B, I selected the first 60 seconds and compared the AudioMining transcript, the manual transcript I had created and the Dragon NaturallySpeaking

transcript. I asked a sound archivist to evaluate Audio Mining's transcription of this 60-second segment for usefulness in appraising its contents.[19]

As a further test of the accuracy of both Dragon NaturallySpeaking and AudioMining, I searched the transcript files for keywords that researchers might hope to find in these tapes. I developed a list of keywords from subject classifications that had been used for material related to the Marital Property Reform Bill in the Wisconsin Historical Society Archives Catalog (ARCAT) and University of Wisconsin – Madison Library Catalog (MADCAT). I added the names of the interviewer and interviewees captured on these tapes to the list of subject keywords.[20]

Findings

The accuracy of the transcriptions the Dragon NaturallySpeaking software produced was significantly less accurate than I anticipated. Based on ad hoc tests I had done having people use Dragon NaturallySpeaking on my computer, I had expected transcriptions that would be inaccurate but that would provide a good sense of the general discussion in the interviews. At minimum I had anticipated that enough keywords, including names and dates, would be recognized the transcription would be of value in showing an archivist appraising the tapes the general subjects that were covered.

What was produced instead was a transcript that had little relation to the recorded interview. Dragon NaturallySpeaking recognized few important keywords. 'Divorce', 'property' and 'commission' were correctly transcribed but none of the other keywords. It was unsuccessful in capturing the names of the interviewer or interviewee, or date of the interview. Many extraneous words were added to the transcript. The resulting transcript was of such poor quality that it would be of almost no use to either archivists or researchers[21].

AudioMining was much more accurate in transcribing Professor Weisberger's interviews. AudioMining did create a transcript that conveyed a general sense of the topics covered in the interviews. Many important keywords including 'divorce', 'commission on women', 'marital', 'property reform', 'Weisberger' and even 'Dreyfus' were correctly transcribed. Some extraneous phrases were included in the transcriptions.

Table 1 shows the transcription of the first sixty seconds of one of the tapes, by three different transcription methods -- manual, by Dragon and by AudioMining. There is significant variation between the three methods of transcription. Dragon NaturallySpeaking bears almost no resemblance to the

manual transcription. Only three phrases from the entire 60 second transcript were accurately transcribed, "where we were", "all of these resources" and "actually became one of our". None of these phrases would be helpful in determining the subject of the tape. There is much more correlation between AudioMining's transcription and the manual transcription. Many of the key concepts are accuracy transcribed including, "Divorce Reform Act."

Transcription of Tape One Side B - first 60 seconds	
Transcription method	Content
Manual	[June Weisberger:] March 3, 2004 between June Weisberger and Linda Roberson. Okay, Lindy. [Linda Roberson:] Okay. Do you need to check that and make sure … [June Weisberger:] No, no. [Linda Roberson:] You can see that it's going. I was talking about what Mary Lou brought to the table. And where we were at that time was that she had just very successfully united bench and bar behind the Divorce Reform Act, which was quite revolutionary in its way and so that was her work there with pretty much done. She had all of its goodwill and all of these resources that she could bring to bear. An it's just d she just finished a tremendous job on a major piece of legislation itself and so could we just slide in another major piece of legislation and move her from protecting women at divorce to protecting people during the ongoing marriage, which actually became one of our big arguments as I know you recall. [June Weisberger:] One of the best arguments. [Linda Roberson:] Yeah.
Dragon NaturallySpeaking	As or a clean July 3 a in a any as you know you and I on and I am rocking in and where we were a time that he is very fast lane eight at him are behind format with evolutionary flat on said who worked in the last time she had all been with all of these resources the heated rings in their sieges if any action is tiny and flashes of who defy a knee in the last angle or for a minute for everything he owns or any there was actually became one of our in our is unknown after you

AudioMining	To pitch their 2004 between giveaways. And at a list of the need to check many sure and I give you a call, and I may move off the table and where we were at a time with which he just very successfully is not a bench and bar behind the divorce Reform Act, which was quite evolutionary and it's why on and so that was her work there with pretty much done. She had all of its goodwill and all of these resources that she could bring to bear. And she just finished a tremendous job on a major piece of legislation itself can be defined in another age in the Atlanta life said any move for protecting women at work to protect the people during the ongoing marriage, which actually became one of our big arguments as I know you all -- started out because like

Table 1

Discussion

While the quality of AudioMining was superior to that of Dragon NaturallySpeaking, neither software produced a truly usable transcript.

Why was the accuracy so low?

From my every day usage of Dragon NaturallySpeaking, I know that voice recognition software can be very inaccurate, even after months of building a personal sound recognition file. Variations in my speech -- having a cold, being tired, using a different microphone, speaking more quietly or loudly than usual, changing my position from sitting to laying down -- all can negatively affect the accuracy of Dragon NaturallySpeaking. This problem is compounded when we ask the software to recognize speech from multiple speakers. If the sound quality is poor, there is background noise, the speaker is elderly, soft spoken or a non-native English speaker the accuracy of recognition of recorded sound can be further compromised.

One hundred percent accurate rendition of spoken material is not possible even under the best of circumstances for several reasons. Voice recognition software can only transcribe words, not sounds, so it cannot convey important semantic but non-word sounds such as grunts, moans or laughter. While it can detect and record changes in intonation such as questions, it cannot detect nuances such as sarcasm or emotional speech. It can be taught to recognize different accents, but not to note that the speaker has a German accent.

Similarly, it can be taught to recognize multiple speakers but not, at this point, provide a paragraph break or other indication that the speaker is now

Person A instead of Person C. This can make it difficult to make sense of transcripts.

No matter how accurate voice recognition software becomes, it cannot learn new words on its own. If a speaker uses a word that is not in the software's vocabulary table, it will simply use a word it knows. It cannot reason out when an utterance might be a legitimate new word. This is particularly problematic if we are hoping that voice recognition software will detect family or place names, which are important keywords in many oral histories.

One workaround is preparing the software for specialized words that might be spoken. This feature exists in many voice recognition software products. The Accuracy Optimizer in Dragon NaturallySpeaking and the Vocabulary Builder feature in AudioMining provide this capability. The customized software developed by IBM for the Multilingual Access to Large Spoken Archives (MALACH) project described below, created a vocabulary table with 60,000 such entries because they realized that researchers would need access to names and places.[22]

The possible uses of voice recognition software

Commercially available voice recognition software can currently only provide cursory subject level access to spoken word material. However, intellectual access to oral history tapes could still be provided for researchers by providing key word identification. The Multilingual Access to Large Spoken Archives (MALACH) project has shown the potential for providing keyword access to spoken material.[23] This project is a large-scale digital oral history transcription project, involving 52,000 witnesses to the Nazi Holocaust consisting of 116,000 hours of videotaped and digitized interviews. The project, begun in the mid-1990s, has vast resources behind it -- IBM, the Johns Hopkins University, the University of Maryland and Survivors of the Shoah Visual History Foundation, involving thousands of engineers, computer scientists, linguists and volunteers. It is much more complex than any project most archivists will ever have to appraise. It involves 32 languages, security and privacy issues, and massive storage requirements.

IBM developed custom voice-recognition software for this project to build up databases of keywords from the oral history video soundtracks. The project explores ideas such as automatic cataloging and automatic creation of template driven summaries of entire oral history interviews. While the ultimate goal of the

project is full transcription of all of the interviews, the short-term goal is word and phrase identification.

Future Directions of Research

The results of the evaluation of Dragon NaturallySpeaking show a low accuracy rate when used with multiple speakers. While AudioMining is more accurate, it is expensive and requires a high level of technical expertise to install and process digital files with it. Consequently, it seems unlikely that either software will be utilized by archivists in their current versions. What degree of higher accuracy, lower price and simplification of technology would be required before archivists would be willing to use transcripts for appraisal is a matter for further research.

Hura makes the point that recognition accuracy is only part of the issue. Customer opinion of that accuracy is equally important.[24] Therefore, it would be important to have archivists test voice recognition software to determine if they could do appraisal with transcripts it creates. It may be that while the transcripts created are not useful for appraisal, the keyword indices created would be useful in augmenting the sampling technique sound archivists currently use to examine audiotapes. A functional examination of voice recognition software by archivists may lead to the creation of new appraisal techniques for spoken word material.

Conclusion

The automatic transcription of spoken material is a laudable goal. It could help us increase the speed of the appraisal of sound material and provide better intellectual access to potential researchers. However, the voice recognition software I evaluated in this pilot study produces results that are too inexact to provide substantial assistance for archivists or researchers. Neither of the software packages created traditionally formatted transcripts. This may never be possible. Without that ability, automatic speech recognition may not be useful to archivists in the appraisal process. However, it is increasingly possible for voice recognition software to provide keyword access to spoken word material that will be valuable to archivists doing subject cataloging and for researchers.

This is a period of tremendous growth in voice recognition technology products. In January, Sonic Foundry released Mediasite[25], which contains a voice recognition component with keyword searching capabilities for videotape recordings.[26] In February, the New York Times wrote about Blinkx, a web search engine that utilizes automatic speech recognition technology to crawl audio video content on multi-media web sites.[27] In March, Microsoft purchased the software company Tellme Networks in order to provide keyword access to cell phones using voice recognition technology.[28] In April, Google announced a beta version of a telephone directory assistance program run completely by voice recognition software.[29] In the very near future, voice recognition technology may be the key to much fuller access to the rich content available in spoken word material for archivists and researchers.

Appendix A

Subject Keywords

Divorce
Marital Property Reform
Property Reform
Marriage
legislation
Commission on Women
Governor Dreyfus
June Weisberger
Norma Briggs
Nancy Hunt
Mary Lou Munts
Midge Miller
Janice Baldwin
Mona Steele
Linda Roberson
Don Dyke

---

[1] The seminal work on this subject is Helen P. Harrison, *The Appraisal of Sound Recordings and Related Materials: a RAMP Study with Guidelines* (Paris: UNESCO, 1987). A review of appraisal theory for those new to audio material can be found in Christopher Ann Paton, "Appraisal of Sound Recordings for Textual Archivists," *Archival Issues* 22 (1997): 117-132.

[2] Conversation with Kyle Kruse, Wisconsin Historical Society, March 23, 2007.

[3] According to participants at the Society for American Archivists Oral History Section meeting Washington, DC, August 4, 2006.

[4] One measure of how mainstream this technology has become is that it is the subject of a Saturday Night Live skit in Episode 2 of Season 31, http://www.ling.ohio-state.edu/~park/384/6-dialogue-systems/demo/SpeechRecoDate.wmv, (accessed: April 8, 2007).

[5] Karen Baumann, Archives Cataloger, Wisconsin Historical Society, February 9, 2006.

[6] Discussion with Wisconsin Historical Society archivist Dr. Carolyn Mattern, August 9, 2006.

[7] B.H. Juang and Lawrence R. Rabiner's paper, "Automatic Speech Recognition – A Brief History of the Technology Development," provides an excellent history of this technology,

http://www.caip.rutgers.edu/~lrr/lrr%20papers/353_LALI-ASRHistory-final-10-8.pdf (accessed: April 15, 2007).

[8] pattern recognition. Dictionary.com. Dictionary.com Unabridged (v 1.1). Random House, Inc. http://dictionary.reference.com/browse/pattern recognition (accessed: April 15, 2007).

[9] "The smallest phonetic unit in a language that is capable of conveying a distinction in meaning, as the m of mat and the b of bat in English." phoneme. Dictionary.com. *The American Heritage® Dictionary of the English Language, Fourth Edition*. Houghton Mifflin Company, 2004. http://dictionary.reference.com/browse/phoneme (accessed: April 15, 2007).

[10] The fields of electrical engineering and computer science do provide relevant research on some aspects of voice recognition use on oral history tapes that I discuss elsewhere in this paper. Terry Fogg and others in the education field have written on improving the quality of transcription for qualitative research interviews:
Terry Fogg and Colin W. Wightman "Improving Transcription of Qualitative Research Interviews with Speech Recognition Technology" (paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April 24-28, 2000).
Lynne M. MacLean, Mechthild Meyer and Alma Estable, "Improving Accuracy of Transcripts in Qualitative Research," *Qualitative Health Research* 14 (2004):113-123.
Susan A. Tilley, "'Challenging' Research Practices: Turning a Critical Lens on the Work of Transcription," Qualitative *Inquiry* 9 (2003): 750-773.

[11] Joseph R. Zumalt, "Voice Recognition Technology: Has It Come of Age?," *Information Technology And Libraries* 24, no. 4 (December 2005): 180-185.

[12] David Bertuca, "Voice Recognition Software And OCLC: Technology That Works," *OCLC Systems & Services* 16, no. 2 (2000) 69-75.

[13] Dragon NaturallySpeaking, version 9.0 does not require training although it suggests that accuracy will increase if the user will train the software to their voice.

[14] Dragon NaturallySpeaking allows users to set their "Speed vs. Accuracy" preference along a continuum ranging from Fastest Response to Most Accurate.

[15] S.R. Hedberg, "Dictating this article to my computer: automatic speech recognition is coming of age," *Expert IEEE* 12 (Nov/Dec 1997):9 – 11.

[16] Zumalt 2005, 183.

[17] Sally Jacobs, Accessioning Archivist, Wisconsin State Historical Society.

[18] The full retail purchase price of AudioMining July 2007 was $5,000.

---

[19] Sally Jacobs, Accessioning Archivist, Wisconsin Historical Society.

[20] See Appendix A.

[21] Discussion with Sally Jacobs, Accessioning Archivist, Wisconsin Historical Society, July 2006.

[22] Bhuvana Ramabhadran, Jing Huang, Upendra Chaudhari, Giridharan Iyengar, Harriet J. Nock, "Guess Who's Speaking: Audio Segmentation for the Automated Transcription of Large Spoken Archives," http://www.research.ibm.com/AVSTG/Audio_Segmentation_Eurospeech_2003.pdf (accessed: April 15, 2007).

[23] "MALACH: Multilingual Access to Large Spoken Archives," http://malach.umiacs.umd.edu/ (accessed: April 15, 2007).

[24] Susan Hura, "How Good Is Good Enough?," *Speech Technology* 12 (Jan/Feb 2007):42.

[25] "Mediasite.com," http://www.mediasite.com, (accessed: April 13, 2007).

[26] "New Sonic Foundry search engine finds words, phrases in video presentations," http://wistechnology.com/article.php?id=3663, (accessed: April 8, 2007).

[27] Jason Pontin, "Slipstream- Millions of Videos, And Now A Way to Search Inside Them," *New York Times*, February 25, 2007.

[28] "Microsoft Agrees to Buy Maker Of Voice-Recognition Software," *Bloomberg News*, March 15, 2007.

[29] "Google Tests Directory Assistance for Phones," http://www.eweek.com/article2/0,1895,2112181,00.asp, (accessed: April 8, 2007).