DOCUMENT RESUME

ED 441 854 TM 030 889

AUTHOR Fogg, Terry; Wightman, Colin W.
TITLE Improving Transcription of Qualitative Research Interviews with Speech Recognition Technology.
PUB DATE 2000-04-00
NOTE 8p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Opinion Papers (120) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Software; Educational Research; *Interviews; *Qualitative Research
IDENTIFIERS *Speech Recognition; *Transcription; Voice Recognition

ABSTRACT

The recent development of high-quality voice recognition software greatly facilitates the production of transcriptions for research and allows for objective and full transcription as well as annotated interpretation. Commercial speech recognition programs that are appropriate for generating transcriptions are available from a number of vendors, with varying degrees of difficulty in use. There are two fundamental approaches to using speech recognition to produce transcriptions: (1) real-time; and (2) batch. The real-time approach uses speech recognition while the interview is in progress; the batch approach relies on an audio recording to make it possible to process several interviews in a batch. The use of speech transcription still requires the use of a human transcriptionist, and the best that can be achieved for transcription speed is a mere doubling of the interview time. However, when the user is not a skilled typist, considerable savings of time can be achieved. If a researcher finds voice recognition software to be superior to conventional typing approaches for transcribing interviews, he or she is likely to find it useful for other tasks as well. (Contains 21 references.) (SLD)

Improving Transcription of Qualitative Research Interviews with Speech Recognition Technology

Terry Fogg and Colin W. Wightman

This paper is prepared for the:
Annual Meeting of the American Educational Research Association in New Orleans
April 2000

Improving Transcription of Qualitative Research Interviews with Speech Recognition Technology

Dr. Terry Fogg, Educational Foundations
Dr. Colin W. Wightman, Computer & Information Sciences

Minnesota State University, Mankato
Mankato Minnesota 56001

The Role of Transcriptions In Qualitative Research

Many approaches to qualitative research employ interview techniques. Morse (1994) named participant observation, phenomenology, ethnography, grounded theory, and ethno science as examples. Speaking at an early stage in the rise of qualitative methodology for educational research, Erickson (1986) stated that fully half of the literature produced concerning qualitative research dealt with data interpretation. However, opinions remain divided as to whether interview notes need to be transcribed as a part of the interpretation process (see Poland, 1995; Huberman & Miles, 1994; and Patton, 1990 for supporting arguments; Mischler, 1991 for opposing arguments; and Seidman, 1998 for alternative suggestions).

Although transcription remains a necessary fundamental element for methodologies involving content analysis (see Richards & Richards, 1994), we note that the majority of recent writers concerned with qualitative interpretation have moved away from generating full transcriptions. Without stepping into the argument of the "myth of objective transcription" (Green, Franquiz, & Dixon, 1997; see also Lapadat & Lindsay, 1998; Psathas & Anderson, 1990; and Denzin, 1989), this paper argues that the recent development of high-quality voice recognition software greatly facilitates the production of transcriptions and allows for both objective, full transcription as well as annotated interpretation.

Authors discussing the interpretation of transcripts describe the generation process as "labor intensive" (Patton, 1990; see also Seidman, 1998; Kvale, 1996; Huberman & Miles, 1994; and Miles & Huberman, 1994). Estimates of the time needed to generate transcripts range from a factor of 2.5 to 4 times the length of the recording being transcribed (Seidman, 1998 and Patton, 1990) up to 4 to 8 times, depending on the fineness of detail and familiarity with content (Miles & Huberman, 1994). After acknowledging the difficulty of generating transcriptions, the authors are then silent as to how the process should occur (although Seidman, 1998 suggests hiring another person). This paper postulates that one of the unacknowledged reasons for the shift away from transcription generation is the large investment in time and energy that they represent.

Nevertheless, electronic transcriptions remain the most reusable resource for qualitative research: transcribed interviews can be searched, re-interpreted, collected and shared with other researchers in a more meaningful way and with much greater efficiency than any other medium. Moreover, the availability of transcriptions can be seen to more readily allow audit trails (Janesick, 1994; Lincoln & Guba, 1985), member checks (Lincoln & Guba, 1985), and support of alternative explanations (Patton, 1990), all techniques that are seen as useful in establishing interpretive validity (Altheide & Johnson, 1994; Atkinson, 1992).

Recently, the possibilities presented by speech recognition software been considered in generating interview transcripts (Fogg & Dzuik, 1999; Levitt, 1998) leading to more efficient generation of interview transcripts, and representing a savings in time and effort. Previously, most discussions on computer uses for qualitative interpretation have focused on storing and sorting evidence (Richards & Richards, 1995; Huberman & Miles, 1994; Tesch, 1990). In this paper we outline the logistical difficulties of producing interview transcripts and explore several ways in which speech recognition software can address them. In addition, we briefly examine several other ways in such software can facilitate qualitative research and identify areas in which speech recognition technology needs to be improved.

Characteristics and Capabilities of Speech Recognition

Commercially speech recognition programs that are appropriate for generating transcriptions are available from a number of vendors such as Lernout and Hauspie, IBM, and Dragon Systems. These programs are designed to recognize continuous speech (normal, fluent speech without requiring pauses between words) with large vocabularies (tens of thousands of words). The performance of these programs is highly variable and critically dependent upon several variables: (1) the capabilities of the computer on which they are installed, (2) the microphone quality, (3) the background noise environment, and (4) the degree to which the models used by the software reflect the speech being recognized. When these variables are properly controlled, recognition accuracies of better than 95% can be obtained.

Computer Hardware Requirements

Each speech recognition vendor specifies minimal requirements for the computer to be used with their product. It must be recognized that these are typically the absolute minimums and that there are good reasons to significantly exceed these requirements. Because speech recognition requires searching through very large data structures, the speed of the recognition will be directly affected not only by the processor speed, but also by the amount of memory that is available. Typically, doubling the recommended amount of memory will result in very noticeably faster responses. A leading VR vendor, for example, in 1999 recommended a 200 MHz processor speed with 64 MB of memory; simultaneously, a leading commercial trainer for VR software was finding a 333 MHz processor with 96 MB RAM necessary to achieve desirable performance with this same software (Fogg & Dzuik, 1999).

The other area in which "more is better" is disk space. Digitized speech can consume upwards of a megabyte per minute. That is, a couple of hours of recordings can easily occupy several hundred megabytes of disk space. Compression techniques, commonly used to greatly reduce the size of audio files sent on the web (such as MP3) create distortions that, even when inaudible to human ears, can increase the recognition error rate by an order of magnitude.

Microphone Quality and Background Noise Environment

The quality of the microphone is extremely important. Because any distortion in the speech signal can introduce recognition errors, it is essential that the microphone be of high quality. In some cases, even the headset microphone provided by the vendor to accompany the software is seen to be inadequate for optimal use (Fogg & Dzuik, 1999). Moreover, speech recognition software has only the most elemental capability to separate speech from background noises so it is essential that the microphone capture as little background noise as possible. The best recognition results are obtained when using a headset-mounted, noise-canceling microphone in a quiet room. Recordings made with a cheap microphone placed on a table between two speakers (as is typical in an interview situation) are likely to produce recognition error rates in excess of 50%.

Model Training

Mathematical models drive speech recognition software. The best recognition accuracy results when these models are closely tuned to the speaker's voice and word usage patterns. The potential user must recognize the necessity for training a model for each speaker's voice and customizing the vocabulary and word usage patterns for the topics being discussed. Conventional estimates of the time needed to train software for the voice model have centered around 10 hours. Newer versions of recognition software promote an enrollment period of only 5 minutes for initial voice recognition by the software, but this low estimate, as might be expected, is accompanied by a high error rate. The user is encouraged to have patience through this learning process; the system will become more efficient and accurate in recognizing voice patterns after about five hours of use.

Handling Punctuation And Disfluencies

In the view of the authors, the single greatest impediment to widespread use of speech recognition for transcription is the inability of the software to automatically generate punctuation. In all current speech recognition products, it is necessary to speak the punctuation characters ("period", "comma", "question mark") if they are to appear in the output text. This necessitates training the system to recognize the punctuation commands and for the user to become practiced in speaking punctuation.

Disfluencies such as word fragments and filled pauses (*eg.*,"umm", "err") will generally cause recognition errors: the software will attempt to recognize them as legitimate dictionary words. Consequently, users of speech recognition systems need to be trained not to use filled pauses and to formulate their thoughts so that they can speak fluently once they start to speak. Silence is reliably recognized so it is far better to dictate in short, fluent bursts separated by large silences than to fill the silences with speech that shouldn't be transcribed. This is very different from the way in which we speak to other people, and some practice is required to learn this discipline.

Using Speech Recognition Software For Transcription

Fundamentally, there are two approaches to using speech recognition to produce transcriptions of an interview: real-time and batch. The real-time approach seeks to use speech recognition while the interview is in progress to produce a transcript. This approach would be similar to the use of a stenographer and does not require the creation of an audio recording. The batch approach relies on an audio recording of the interview and attempts to create a transcript after-the-fact. The term "batch" comes from the fact that several interview recordings can be processed together in a batch.

Real-Time Transcription

Real-time transcription would occur if a qualitative researcher could carry on a conversation with one or more interviewees and an accurate electronic transcription could be generated at that same time. This scenario would clearly have advantages for immediate member-checking and faster analysis of content. However, the present state-of-the-art is not capable of supporting this application. Current software is designed for recognizing dictations produced by a single speaker, not dialogues. Add to this the need for very quiet background environments, the intrusiveness of requiring the interviewee to wear a headset-mounted microphone, the recognition errors created by word fragments and filled pauses, and the need for each speaker to complete several hours of enrollment training to tune the recognition models and the difficulties *become* far greater than can be justified. Better to concentrate on simply obtaining a good audio recording.

Batch Mode Transcription

In this mode, the transcript is to be produced from a recording of the interview. This can be done in one of two ways: either the speech recognition software can be used to process the recording itself with a human transcriptionist then editing the resulting output, or the transcriptionist can use the speech recognition software directly and re-dictate the interview in a process known as "ghosting". Attempting to use the recognition software to directly process the audio recording is subject to all of the same limitations associated with real-time transcription. Nevertheless, doing so can accomplish some transcription that, despite a very high error rate, can show the range of the interview and may be of some value in helping to prepare a more accurate transcription.

The most reliable method for generating a transcript is for the transcriptionist to simply re-dictate a recorded interview. The transcriptionist listens to the recorded interview and then repeats the sentences, including punctuation, into a high quality microphone. Because the only voice being recognized is that of the transcriptionist (who can go through the full enrollment and model training process), and the re-dictation can be done in a quiet environment, the recognition accuracy can be quite high. Typically, the transcriptionist listens to the tape for one full sentence or thought, and then dictates it into the program. Errors (either in dictation or in recognition) are typically edited using voice-activated commands at the end of each conversation passage (these roughly correspond to one paragraph of written text). This approach makes it easier to check the written text against the tape-recorded segment.

VR Transcription vs. Conventional Transcription

Given that the use *of* speech recognition to produce interview transcripts still requires the use of a human transcriptionist, one must question whether any real advantage accrues from its use. Even given optimum computer processing speed and optimum cognitive processing (in this case, the user's ability to listen to a tape-recorded speech segment, understand it fully, and reliably repeat it with correct punctuation and disfluencies faithfully represented), the best that can be achieved for transcription speed is a mere doubling of the recorded interview time. That is, a recorded sentence that took ten seconds for the interviewee to utter takes ten seconds for the VR transcriptionist to listen to, and about ten seconds for the VR transcriptionist to repeat into the headset. While a transcription rate of twice the recording time would represent a significant gain in productivity, the VR transcriptionist will still need to make editing changes to the text to correct recognition and re-dictation errors. The length of time to generate a voice recognition transcription can thus increase to the point where it would exceed the length of time needed for an experienced conventional transcriptionist to produce a transcript.

For most researchers, the primary concern with regard to producing transcripts is cost. Cost is driven by two factors: the productivity and the cost of the person making the transcription. It is the authors' experience that the existing speech recognition dictation products do not yield significant productivity gains in comparison to skilled transcriptionists who are experienced with the field and terminology that is being discussed. However, when the transcriptionist is not a skilled typist (such as the authors of this paper, or many of their research assistants) significant timesavings can be obtained. Moreover, if the transcriptionist was involved in the interview, or has intimate knowledge of the field, he or she can often gain additional timesavings by being able to interpret the recording more easily. Thus, with the use of speech recognition software and ghosting, it is possible for most researchers and their assistants to achieve productivity comparable to that of a professional transcriptionist. Producing one's own transcripts, or having them produced by an assistant, can not only yield considerable dollar savings, but also aids the researcher in reviewing the content of the interviews and can result in additional time savings.

Applications for Speech Recognition In Qualitative Research

If a researcher has found VR software to be superior to conventional typing approaches for transcribing interviews, then he or she would quite likely benefit from its use in other areas as well, thus gaining further improvements in productivity, maintaining dictation skills, and continuing to improve the recognition software models. In addition, use of voice recognition software would seem to promise relief from repetitive stress problems associated with prolonged periods of typing.

Many writers find that speech recognition software benefits the writing process once the document is well outlined and a consistent voice in the text is defined. As an example, this paper was written and edited with a commercial dictation product after the ideas were organized in point form. When used to assist the composition of new texts, the productivity gains associated with the use of speech recognition are substantially greater than those observed for re-dictation. This is because the time required to listen to the recorded interview is eliminated.

Many users have been using speech recognition effectively with email correspondence. Subjectively, this technique allows increased productivity because the writer can accomplish other tasks while dictating (referencing reports, budgets, calendars, etc.). In those situations, the program is quick and easy to use because the insertion of punctuation is not difficult during the composition process. Similarly, VR has been used for generating written records of classroom lectures. In situations where the content is very familiar, the writer can quickly speak extemporaneously and insert punctuation as it occurs.

Summary and Conclusions

Because of the methodological requirements for interpretive validity, the need for member checking, and desirability of audit trails, we believe that more consistent production of data transcripts is needed for qualitative studies. These transcripts represent primary data that could and should be shared and pooled amongst qualitative researchers, once issues of confidentiality are resolved. Our strong contention is that the adoption of VR techniques will allow workers with good subject knowledge but poor typing skills (graduate students or qualitative researchers) to operate at the high ends of productivity achieved by conventional transcriptionists. Economic savings can thus be realized and researchers would benefit from a more intimate experience of their data.

Technical Challenges

Clearly, the biggest technical problem associated with speech recognition software is providing for punctuation. The inability of the software programs to interpret verbal punctuation such as pauses and inflections (collectively called "prosody"), and resulting requirement that all punctuation must be explicitly dictated, limits the utility of the software. Neither is software technology at the point where it can effectively use artificial intelligence heuristics to infer the correct punctuation of spoken conversation. This limitation, in combination with the requirements for high-quality recordings in quiet environments, effectively limits the use of speech recognition to re-dictation or ghosting.

Cognitive Aspects And Productivity Gains

Voice recognition software involves the cognitive aspects of switching modalities (hearing to speech vs. hearing to typing). The types of cognitive processes involved suggest the possible involvement or interaction with cognitive style (auditory or visual). The cognitive processes involved with voice recognition software indicate that reading abilities and memory capacities are involved, and we know that these abilities are found in different measure among users. One claim that needs to be validated is that examination of a printed transcript, with its opportunity for recursive reading, allows for more thorough and more insightful analysis of content. We also need to develop a systematic comparison of transcription speeds for professional typists versus speech recognition programs used by qualitative researchers.

Enhanced Productivity for Qualitative Research

Qualitative research is often derided for drawing conclusions from very limited sets of interviews. By recognizing interview transcripts as a primary data source, producing transcripts more consistently, and pooling them across groups of qualitative researchers, the authors hope for an enhanced public estimation of the value of qualitative research. By improving the productivity and reducing the cost associated with the production of transcripts, speech recognition should a valuable addition to the qualitative researcher's arsenal.

References

Altheide & Johnson (1994). Criteria for assessing interpretive validity in qualitative research. In N.K. Denzin and Y.S. Lincoln (Eds.) Handbook of Qualitative Research, 485-499. Thousand Oaks, CA: Sage.

Atkinson, P. (1992). Understanding ethnographic texts. Qualitative Research Methods Series, 25. Newbury Park, CA: Sage.

Denzin, N.K. (1989). Interpretive interactionism. Applied Social Research Methods Series, 16. Newbury Park, CA: Sage.

Erickson, F. (1986). Qualitative methods in research on teaching. In Handbook of Research on Teaching (3rd ed.) M.C.Wittrock (Ed). New York: MacMillan.

Fogg, T.L. & Dzuik, M. (1999). Using voice recognition software for generating interview transcripts. Paper presented at the Fifth Annual Midwest Qualitative Research Conference (Minneapolis, MN. May

24-25).

Green, J., Franquiz, M. & Dixon, C. (1997). Myth of the objective transcript: Transcribing as a situated act. TESOL Quarterly, 31 (1), 172-176.

Huberman & Miles (1994). Data management and analysis methods. In N.K. Denzin and Y.S. Lincoln (Eds.) Handbook of Qualitative Research, 428-444. Thousand Oaks, CA: Sage.

Janesick, V.J. (1994). The dance of qualitative research design. In N.K. Denzin and Y.S. Lincoln (Eds.) Handbook of Qualitative Research, 209-219. Thousand Oaks, CA: Sage.

Kvale, S. (1996). InterViews: An introduction to qualitative research interviewing. Thousand Oaks, CA: Sage.

Lapadat, J.C. & Lindsay, A.C. (1998). Examining transcription: a theory-laden methodology. Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13 - 17, 1998).

Levitt, H. (1998). Automatic speech recognition and its applications. In Issues unresolved: New perspectives on language and deaf education. (A. Weisel, Ed.), 133-138. Washington, DC: Gallaudet Press.

Lincoln, Y.S. & Guba, E.G. (1985). Naturalistic inquiry. Beverly Hills, CA: Sage.

Miles, M.B. & Huberman, A.M. (1994). Qualitative data analysis: An expanded sourcebook (2nd ed.) Thousand Oaks, CA: Sage.

Mischler, E.G. (1991). Representing discourse: The rhetoric of transcription. Journal of Narrative and Life History, 1, 255-280.

Morse, J.M. (1994). Designing funded qualitative research. In N.K Denzin and Y.S. Lincoln (Eds.) Handbook of Qualitative Research, 220-235. Thousand Oaks, CA: Sage.

Patton, M.Q. (1990). Qualitative evaluation and research methods (2nd ed.). Newbury Park, CA: Sage.

Poland, B.D. (1995). Transcription quality as an aspect of rigor in qualitative research. Qualitative Inquiry, 1, 290-310.

Psathas, G. & Anderson, T. (1990). The "practices" of transcription in conversation analysis. Semiotica, 78, 75-99.

Richards, T.J. & Richards, L. (1994). Using computers in qualitative research. In N.K. Denzin and Y.S. Lincoln (Eds.) Handbook of Qualitative Research, 445-462. Thousand Oaks, CA: Sage.

Seidman, I. (1998). Interviewing as qualitative research: A guide for researchers in education and the social sciences. New York: Teachers College Press.

Tesch (1990). Qualitative research: Analysis types and software tools. London: Falmer.

AERA

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM030889

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Improving Transcription of Qualitative Research Interviews with Speech Recognition Technology	
Author(s): Dr. Terry Fogg & Dr. Colin W. Wightman	
Corporate Source: Minnesota State University, Mankato	Publication Date: 26 April 2000

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ☑	Level 2A ☐	Level 2B ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Colin W. Wightman Terry Fogg	Printed Name/Position/Title: Colin Wightman/Asst Prof Terry Fogg Asst. Professor	
Organization/Address: Minnesota State University Mankato, MN 56001	Telephone: 507 3896765	FAX: 507-389-1463
	E-Mail Address: colin.wightman@mankato.msus.edu	Date: 4/26/2000

(over)

ERIC®

Clearinghouse on Assessment and Evaluation

University of Maryland
1129 Shriver Laboratory
College Park, MD 20742-5701

Tel: (800) 464-3742
(301) 405-7449
FAX: (301) 405-8134
ericae@ericae.net
http://ericae.net

March 2000

Dear AERA Presenter,

Congratulations on being a presenter at AERA. The ERIC Clearinghouse on Assessment and Evaluation would like you to contribute to ERIC by providing us with a written copy of your presentation. Submitting your paper to ERIC ensures a wider audience by making it available to members of the education community who could not attend your session or this year's conference.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed, electronic, and internet versions of *RIE*. The paper will be available **full-text, on demand through the ERIC Document Reproduction Service** and through the microfiche collections housed at libraries around the world.

We are gathering all the papers from the AERA Conference. We will route your paper to the appropriate clearinghouse and you will be notified if your paper meets ERIC's criteria. Documents are reviewed for contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality. You can track our processing of your paper at **http://ericae.net.**

To disseminate your work through ERIC, you need to sign the reproduction release form on the back of this letter and include it with **two** copies of your paper. You can drop of the copies of your paper and reproduction release form at the ERIC booth (223) or mail to our attention at the address below. **If you have not submitted your 1999 Conference paper please send today or drop it off at the booth with a Reproduction Release Form.** Please feel free to copy the form for future or additional submissions.

Mail to: AERA 2000/ERIC Acquisitions
The University of Maryland
1129 Shriver Lab
College Park, MD 20742

Sincerely,

Lawrence M. Rudner

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

ERIC/AE is a project of the Department of Measurement, Statistics and Evaluation at the College of Education, University of Maryland.