

*Evaluating Voice Recognition Software:
An Analysis of Quality & Costs*

Sarah Ayres, University of Bristol & Ian Stafford, Cardiff University

SUMMARY

This paper provides an overview of a small scale evaluation of voice-recognition (VR) software compared to professional transcription services. The VR software was used to transcribe interview data as part of a two-year Economic and Social Research Council (ESRC) funded project and provides an insight into the feasibility of using VR software as part of a working research project. The paper is divided into four sections. Section one explores the rationale underpinning the decision to use VR software for transcribing qualitative field data. Section two provides an outline of the methodology used to evaluate the VR software and professional transcription services, while section three sets out the key findings of the evaluation. The paper concludes by reflecting on the quality of VR software and the validity of using it as opposed to professional transcription services.

INTRODUCTION

A common issue faced by researchers when conducting qualitative research is the complex question of whether to use professional transcription services or undertake their own transcription. Due to the time consuming and arduous nature of transcription, many researchers simply do not have the time or the resources to undertake their own transcription and instead use professional transcribers. Nonetheless, professional transcription is not without its problems. Professionally transcribed interviews are often criticised for being inaccessible in terms of format and are often littered with errors, which significantly affect the quality and accuracy of data. Given this, researchers often need to invest considerable time working through professional transcripts in order to ensure the accuracy of the scripts and presentational issues.

This paper explores the extent to which recent technological improvements within voice-recognition software (VR software) provides a potential solution to these problems. It provides an overview of an evaluation of VR software in comparison to traditional professional transcription services. The paper is divided into four sections. Section one explores the rationale underpinning the decision to use VR software for transcribing qualitative field data. Section two provides an outline of the methodology used to evaluate the VR software and professional transcription services, while section three sets out the key findings of the evaluation. The paper concludes by reflecting on the quality of VR software and the validity of using it as opposed to professional transcription services.

THE RATIONALE

In 2005 Ayres was in the process of writing a research bid to the ESRC as part of the First Grants Scheme. Aware that the ESRC were keen to promote methodological developments and the use of cutting edge technologies, Ayres conducted a web search to identify appropriate equipment to use for qualitative interviewing. The original search aimed to identify high quality digital recorders that would improve the sound quality of recordings, enable the electronic storage of digital files and make transcription easier. The final ESRC proposal identified a number of advantages in combining digital recorders with other cutting edge technology to facilitate the gathering and analysis of field data. These benefits included improved sound quality, the ability to save interviews as electronic files and the potential to apply a voice-recognition software package (DragonV9) to facilitate transcription. An extract from the original ESRC proposal below outlines the two distinct ways in which this technology would be used.

Extract from Research Bid to the ESRC's First Grant Scheme

1. Transcribing dictated digital recordings (i.e. the voice of one researcher)

Using the voice recognition software these files can be *automatically* transcribed with *no* additional manual effort. After each interview and observation, the researchers will make brief dictated notes under four headings: observations, key issues, triangulation and codes. These notes will be automatically transcribed and saved into word files.

This process has a number of advantages, including:

- Capturing initial thoughts or impressions that may be lost between interview and transcription,
- Identifying key issues to pursue in subsequent interviews,
- Identifying issues for triangulation,
- Sharing information between researchers in an efficient and timely manner, and
- Identifying codes for analysis that are firmly grounded in the data.

1. Transcribing interview digital recordings

Due to the time consuming nature of transcription, many researchers prefer to use professional transcribers. However, professional transcription is not without its problems. Even amongst the most experienced transcribers errors can creep in that can significantly affect the quality and accuracy of data. To ensure sensitivity to the data, many researchers prefer to undertake their own transcription. This debate forms the rationale for examining the voice recognition software. The voice recognition software can only identify the 'trained' voice so face-to-face interviews will require the researchers to *listen* to the recordings through a headset and, at the same time, *dictate* the transcripts. It is anticipated that this process will be significantly faster than *typing*.

Ayres was unfamiliar with this technology so began making some telephone enquiries about its application and use. It became apparent that in order for the equipment to operate at optimal level, some form of technical assistance and training would be required. Ayres contacted a leading UK voice recognition expert based in Cheltenham,

Neil Winton. After repeated telephone discussions and e mail correspondence with the author, Winton provided an outline of the equipment and training needed to fulfil the projects needs. Once, this had been agreed, Winton was able to send an invoice, detailing the required equipment and costs that the author attached to the ESRC bid. Details of the equipment and training costs and minimum requirements needed to run the software can be found in Appendices 1 and 2.

EVALUATING VOICE RECOGNITION SOFTWARE

The ESRC proposal outlined a small evaluation of the VR software which would take place after the project's fieldwork had been completed. An extract for the bid can be viewed below.

Validity and reliability of voice recognition transcribing

The project will include a small evaluation of the voice recognition software, to be completed as follows:

- A purposive sample will be used to select three digitally-recorded interviews, each with slightly different characteristics in terms of language and key words.
- The three interviews will be a) manually transcribed verbatim by a professional transcriber and b) transcribed verbatim using the voice recognition software.
- No *typed* edits or adjustments are to be made while using the voice recognition software.
- The two drafts will be given to three assessors who, unaware of how they have been produced, will be asked to rate the quality of the transcripts based on a set of criteria (e.g. ability to convey a clear message, punctuation, visual appearance of text).
- The objective will be to compare traditional methods and new technologies for transcription.
- The findings of this small study will be of value to both academics and practitioners who undertake transcription.
- The results, including a broader assessment of how the technology has been used during the project, will be written up as a research paper and will be disseminated via the web and amongst research methods groups and fora.

The methodology for the evaluation is as following:

Stage 1. Selecting the Sample

The first step in developing this evaluation was to identify a purposive sample of three digitally-recorded interviews. Each of the interviews contained slightly different characteristics to reflect the varying nature and style of digitally recorded interview data. Three criteria were used to select interviews, including:

- A high degree of technical language,
- A recording of an interview with high levels of background noise/interference,
- An interview of more than one interviewee.

Test 1. The Technical Interview

A potential problem with both professional transcription and VR software is the failure to understand or pick-up technical terms which may be used by the interviewee over the course of an interview. The VR software includes training and 'key words' functions which allow users to train the software to pick up unusual or complex language. For example, the software can be trained to pick up abbreviations such as RFAs (Regional Funding Allocations). By selecting this variable it is hoped that the evaluation will test the accuracy of the VR software and professional transcription services in dealing with complex, technical language.

Test 2. The Noisy Interview

Ideally interviews should be carried out in quiet locations that provide high quality recordings. However, the reality of fieldwork often means that researchers may have to carry out interviews in noisy locations e.g. cafes or restaurants. Recordings containing a high level of audio interference can make it very difficult for professional transcribers to accurately pick up the dialogue. There is an assumption that the researcher who carried out the interview will be better placed to transcribe it as they were present at the time.

Test 3. The Multi-actor Interview

A common feature of qualitative research is for researchers to carry out interviews with several interviewees at the same time. This provides a test for professional transcribers to identify which individuals are speaking during the interview. Clearly researcher transcribed interviews should reflect the changes in speaker more accurately. This scenario will also test the VR software's formatting capabilities.

Stage 2. Transcription

After the three interviews were selected they were (i) manually transcribed by a professional transcriber and (ii) transcribed by a researcher using the VR software. A transcription service used by academics within the University of Bristol was selected for this purpose. For the voice recognition software, the following rules were enforced:

- No *typed* edits or adjustments are to be made while using the VR software *except* those that would normally be used in the normal day-to-day use of the software, such as spell check and some basic formatting functions.
- An accurate record of the time taken to transcribe an interview must be made by the researcher using the VR software.

All the formatting functions (e.g. new line, new paragraph) can be managed via the VR software. However, the application of the software as part of the working project was restricted to managing the interview discussions themselves. The researchers found that a combination of manual typing (for format) and voice-recognition software (for interview data) was the most effective means of using the technology. As such, it was deemed appropriate that some manual formatting should be allowed within the evaluation because it is primarily a test of how the software is actually used in practice by researchers.

Stage 3. Evaluation

In total, six interview transcripts were produced - three interviews by the professional transcription service and three using the voice recognition software.

These transcripts were given to three independent assessors who were asked to rate the quality of the transcripts based on a set of criteria. The assessors were other academics working within the School for Policy Studies at the University of Bristol and were not informed how the transcripts had been produced.

The evaluation used the following criteria as benchmarks of accuracy and quality:

1. Visual Appearance/Formatting - the overall appearance and format of the text.
2. Accessibility - the extent to which the transcript conveys a clear message.
3. Accuracy - the overall accuracy of the text in terms of 'making sense' and accuracy of technical terms.
4. Spelling/punctuation.

Assessors completed an evaluation sheet to rate the transcripts (see Appendix 3). For the first four criteria the evaluation sheet used a likert scale ranging from very poor to very good. Using this scale each transcript was given a score for each criteria:

Very poor = 1

Very good = 5

In addition, an indication of the cost of both methods is provided. This has been calculated using:

- The costs per transcript using professional transcription,
- The cost of transcription using a Research Associate (RA) on grade 35 on the university pay scale.

The cost of transcription using the RA has been calculated by noting the length of time taken to undertake each transcript and the RA's hourly rate of pay. Clearly if a more senior researcher uses the software to transcribe interviews then the financial costs to the project will increase. There are also the initial start-up/equipment costs to consider when factoring overall cost effectiveness (Appendix 1).

KEY FINDINGS

Comparing the quality of transcripts compiled using VR Software and Professional Transcription Services

The feedback from the three academic assessors is presented in Table 1. The results indicate that the three interview transcripts completed using VR software outscored the professional transcripts. This pattern was most significant in the ‘noisy interview’ where the VR software transcript scored 47 out of 60 across the four categories and the professional transcript scored 30 out of 60. The relative strengths and weaknesses of the VR software and professional transcription services are highlighted if the results are broken down in terms of their scores for individual criteria.

Table 1: Results from the Assessment

Interview	Visual Appearance/ Formatting		Accessibility		Accuracy		Spelling/ Punctuation		Interview Total	
	PRO	VR	PRO	VR	PRO	VR	PRO	VR	PRO	VR
Technical Interview	11	13	12	14	8	15	12	12	43/60	53/60
Noisy Interview	11	12	10	11	7	12	11	12	30/60	47/60
Multi-actor Interview	10	12	10	11	6	10	11	11	37/60	44/60
Criteria Total	32/45	37/45	32/45	36/45	21/45	37/45	34/45	35/45		

Several key findings can be identified:

1. *There were only marginal differences between the two modes of transcription in terms of the formatting and accessibility of transcripts*

This is perhaps unsurprising given that both the professional transcriber and research associate are likely to have had similar experience in terms of presenting interview data in a clear and accessible way. However, the marginal differences in terms of the scores given to the two can perhaps be explained by the inclusion of ‘ums’ and other asides within the professional transcripts which one assessor commented made the text difficult to read and at times the transcription appeared ‘too literal’.

2. *The VR software significantly outperformed professional transcription services in terms of the accuracy of the interview transcripts*

In each of the interviews the VR software transcripts scored significantly higher than the professional transcripts in terms of accuracy. This would appear to confirm the hypothesis highlighted in Section 2 that professional transcripts tend to be more inaccurate than those completed by the researchers who carried out the interviews. This problem is exacerbated by the complexity of the language used in interviews, the quality of the recording and the number of interviewees. For example, all of the assessors commented that the professional transcript of the noisy interview included too many sections marked as inaudible or '?' and there were clear 'uncertain aspects' in the transcription. In addition, the surname of one of the interviewees within the multi-actor interview was incorrectly transcribed. The accuracy of the VR software also dipped in terms of the 'noisy' and 'multi-actor' interviews, although to a lesser degree.

3. *The standard of spelling and punctuation within the transcripts was virtually identical*

Perhaps an area where one would expect the VR software to potentially under perform was the quality of the spelling and punctuation within the text. However, the evaluation demonstrated that the assessors considered both professional and VR transcripts as being of a similar quality with regard these criteria.

Overall the results of the assessment would appear to suggest that the use of VR software provides higher quality transcripts than professional transcription services, particularly in terms of the accuracy of transcripts. However, as noted in the introduction, concerns regarding the sensitivity and accuracy of professional transcripts are a common concern amongst researchers. It is perhaps unsurprising that transcripts completed by a researcher, either manually or using VR software, would likely outscore a professional transcript.

Comparing the financial costs of VR Software and Professional Transcription Services

Table 2 sets out the time and costs of the three interviews transcribed as part of this evaluation (the times are rounded up to the nearest minute for convenience). An unexpected result was that under test conditions the 'noisy interview' was actually the quickest to be transcribed and the 'multi-actor interview' took the most by almost one hour. This result can be explained as follows:

1. In transcribing the 'noisy interview' the research associate took advantage of the noise cancellation function of the software included with the Olympus digital recorder. However, the noise cancellation function is only able to operate at 100% speed, which had the unintended effect of speeding up the transcription process. We would recommend using this function and speed for poor quality recordings but this approach to transcribing is emotionally exhausting and difficult to sustain over a long period.

2. The second unforeseen factor was the relative low quality of the ‘multi-actor interview’. Although the actual sound recording was of a fairly high quality, one of the interviewees was very softly spoken and therefore their voice was at times difficult to pick up even though they were sitting closest to the microphone.

The overall cost of the VR transcriptions is calculated using the hourly rate of the Research Associate completing the transcripts, which in the case of this research project was approximately £21.47.

Table 2: The Costs of Voice Recognition transcriptions

Interview	Time	Cost
Technical Interview	2:00	£42.94
Noisy Interview	1:55	£41.15
Multi-actor Interview	2:53	£61.91

Clearly the cost of voice recognition is dependent on the experience and hourly rate of the researcher carrying out the transcriptions. In the case of this evaluation the research associate was grade 35 on the university pay scale. The time taken to carry out the transcriptions was produced in optimum conditions after the research associate had used and trained the Dragon software for an extended period (approximately 18 months). Clearly the first interviews transcribed using the software take longer to complete whilst the user familiarises themselves with the software. However, one might expect an hour long interview to take approximately 3 hours within a relatively short space of time.

The overall cost of the professional transcripts for the three interviews included within this evaluation was £300 (inc VAT). Therefore, the VR software provided a saving of £154 - over 50% cheaper than professional transcriptions. However, two additional factors need to be considered:

1. The start up costs of the VR software are relatively high, for example, on this project the hardware, software and training for two researchers came to approximately £1900 (inc VAT). (see Appendix 1) However, given the savings outlined within this small study it can be argued that savings in real terms can be seen after approximately 30-40 interviews depending on the speed of transcription and the costs of professional transcription.
2. The inaccuracies inherent in professional transcripts means that researchers need to edit professional transcripts in order to verify the data. This can be a time-consuming process in itself. Although some errors remain, the VR software negates the need to spend additional time checking transcripts. It also enables the researcher to have a much closer relationship with the data providing the foundation for the

CONCLUSIONS

The evaluation of the VR software carried out within this study has highlighted the potential benefit of using VR software as an alternative to manual transcribing and using professional transcription services. We conclude by answering a number of simple questions in the hope of shedding light on the validity of this software.

- Does the software successfully manage the tasks as set out in the original project brief? *Answer: Yes*
- Has the technology performed better or worse than you predicted? *Answer: Better*
- From your initial impressions, is transcribing using voice recognition software faster than manual typing? *Answer: Yes*
- When using voice recognition software, is the quality of the transcripts better than professional transcription services? *Answer: Yes*
- Is voice recognition software cheaper than using professional transcription services? *Answer: Yes – in the medium term.*
- Would you recommend this software to other qualitative researchers? *Answer: Yes, but it is important that equipment and training is tailored to the needs of the individual and research project.*
- Would you continue to use this software in future research projects? *Answer: Yes*

Appendix 1: Equipment details and costs

8 December 2005

Voice Recognition Systems
 70 Century Court, Montpellier Grove,
 Cheltenham, GL50 2XR,
 Tel: 01242 702804, Fax: 01242 702934
 E-mail: info@pc-voice.co.uk,
 Website www.pc-voice.co.uk

Company Name and Address:

Contact:		
Signed:		
Position		
Date:		
Tel:		
E-mail		

Naturally Speaking Preferred V9/Olympus DS-2200 digital recorder

2 *	Dragon Naturally Speaking Preferred V9 @	110.00	220.00
2 *	DS-2200 Digital Dictation Recorder @ includes 128 MB memory xD-picture card, Remote control DSS player pro software, ME51S stereo microphone USB download cable, leather case, batteries	225.00	450.00
2 *	VRS Headset Microphone @ Provides superior recognition with Dragon Naturally Speaking	65.00	130.00
2 *	AS-2000 Transcription kit @ Includes: DSS transcription pro software foot-switch and stereo headset	110.00	220.00
1 *	One day installation and training for two people @ VRS training material Set-up of specialist dictionary, install specialist shortcuts and 30 days on-line support.	600.00	600.00
Total Price £1620.00 ex vat			

Optional Extras

1 *	Olympus 512Mb xD picture card provides over 17 hours of stereo recording	55.00 ex vat	<input type="checkbox"/>
-----	--	---------------------	--------------------------

Please Note

Price includes assembly of equipment, full installation and training, plus 30 days on-line support. All prices subject to VAT @ 17.5%.

Appendix 2: Minimum Software requirements

Minimum System Requirements: Windows

Dragon Naturally Speaking Preferred 9

- Intel® Pentium® / 1 GHz processor (for example, Pentium® M, Pentium® 4), or equivalent AMD® processor - Faster processors will yield faster performance
- 1 GB RAM
- 1 GB free hard disk space
- Microsoft® Windows® XP (SP1 or higher) Home and Professional, 2000 (SP4 or higher), Vista
- Creative® Labs Sound Blaster® 16 or equivalent sound card supporting 16-bit recording
- Microsoft® Internet Explorer 5 or higher
- CD-ROM drive
- Nuance-approved noise-cancelling headset microphone (included)
- Speakers
- A web connection is required for activation

Olympus DSS Player

- IBM PC/AT compatible PC
- OS: Windows ME/2000 Professional/XP Pro/XP Home Edition
- CPU: Pentium II class 333 MHz processor or higher. (If recording to a hard disk directly with the WMA format, please use in the range of 500 MHz or more)
- RAM: 128 MB or more (256 MB or more recommended)
- Hard disk space: 50 MB or more
- Drive: 2 x or faster CD-ROM, CD-R, CD-RW, DVD-ROM drive.
- Sound card: Creative Labs Sound Blaster 16 or 100% compatible sound card.
- Display: 800 x 600 pixels or more, 256 colours or more
- Browser: Microsoft Internet Explorer 4.01 SP2 or later

One free USB port, earphone output or speakers.

Appendix 3: Evaluation Sheet

Evaluation Sheet					
Reviewer					
Interview Number					
Criteria	1 (v. poor)	2 (poor)	3 (average)	4 (good)	5 (very good)
Visual Appearance/ Formatting					
Accessibility					
Accuracy					
Spelling/ Punctuation					
Any further comments.....					
Signed Date.....					

ACKNOWLEDGEMENTS

This work is funded through the Economic and Social Research Council. Project title: English Regionalism: Rhetoric or Substance? Evaluating Decision Making Procedures for Regional Funding Allocations, Award number RES-061-23-0033.