Secondary Analysis of Audio Data. Technical Procedures for Virtual Anonymisation and Modification[1]

Henning Pätzold

Abstract: Qualitative material presented as audio data requires a greater degree of protecting of anonymity than for example textual data. Apart from the verbal content, it carries paraverbal aspects including voice characteristics, thus making it easier to identify the speaker. This complicates secondary analysis or reanalysis conducted by researchers who were not involved in the data collection. Difficulties increase if the chances are high that the researcher and the interviewee come in contact for example through a meeting.

This paper describes the technical procedures that are used to modify the sound of the audio source in a way that it reduces the possibility of recognition (i.e. similar to that of a carefully written transcript). A discussion of the technical possibilities of this procedure along with an exploration of the boundaries of anonymisation is presented.

Table of Contents

1. Introduction

The qualitative researcher's first encounter with the world is usually via sound not text (cf. GARZ & KRAIMER, 1994). Most qualitative data are collected using recording equipment and then further processed (of course this does not devalue other sources (from pictures to video studies), but they are used far less often). Literature on research methods recommends almost unanimously that transcripts be produced after data collection (cf. e.g. BORTZ & DÖRING, 2003; KOWAL & O'CONNELL, 2003). At this point we won't discuss the (probably mainly historical)

---

1   This is the English translation from a German text, published in *FQS* in January 2005. *FQS* thanks the author for providing the English version and Louise CORTI for copy-editing, May 2007.

reasons for the importance assigned to transcripts; it should be noted, however, that it is increasingly possible to bypass transcription, not least because of the use of new media. Initially such techniques were mainly used for sources which are very difficult to transcribe (especially video data), but in the meantime several references can be found to the possibility of analysing purely audio data based directly on the recording, omitting the transcription stage (KUCKARTZ, 2003, p.586). The reason is that nowadays software allows fast and easy non-sequential access to all the data—which formerly was one of the main disadvantages of recorded text over transcripts. Recent software for qualitative data analysis allows the whole research process (for example qualitative content analysis) to be performed with the aid of the computer, including the linking of categories to certain sections of the audio material. [1]

Although digitalised audio data are gaining importance as a source for qualitative analysis—and will do so even more in future—it is currently rather difficult to provide these data for secondary analysis. The main reason for this is the lack of anonymity. For example an interview series with students couldn't be handed over for secondary analysis by a colleague of the researcher without running the risk that this colleague could identify interview partners by hearing their voices in other contexts. This article will explain methods for overcoming this problem. [2]

2. Task: Modification

An audio source intended for analysis can be modified acoustically. Anonymised interviews on television are a well-known instance of this: a voice, and paraverbal utterances can be heard, while, at the same time, the voice sounds artificial and one can be sure that the speaker can not be identified easily from the recording. For the researcher, however, these procedures were much too expensive as they require a great deal of technical skill and tools. Nowadays an alternative exists—software for editing audio data. It provides a satisfactory result as far as anonymisation is concerned, though losses in quality, which complicate the analysis, must be taken into account (see below). The possibilities suggested here use a vocoder or pitch changing to anonymise the data. Both can be done with freely accessible software. A vocoder maps a given sound signal onto another, which results in a complex mixture of the characteristics of both signals. Pitch changing produces an effect known from records which are played too fast or too slowly, but without altering the speed of the recording. [3]

3. Procedure

The original audio source is used for editing. It is in the nature of the task that anonymisation goes hand in hand with a certain loss of quality and sound information which often leads to a loss of intelligibility. Therefore the quality of the audio source should be as good as possible, regardless of how modification is used. It can, for instance, be beneficial to raise the volume of silent data, use software to remove noise etc. The software introduced below (Audacity), for instance, offers possibilities for removing background noise. Loud annoying sounds (e.g. a push against the microphone) can be cut out with this programme.

The data must therefore exist in a format that can be processed by the computer. WAV and MP3 files[2] are easy to use, other formats may need to be converted (see Section 5). Once the data are prepared in this way, the actual editing can start. [4]

3.1 Preparation: Removing personal information

As with written transcripts, all personal data which give cues to the identification of the interviewee or other related people or organisations have to be marked and deleted. This can also be done using Audacity (see also Section 3.3.). Particular sections in the original audio file can be marked and deleted. The high resolution of the waveform picture on screen also allows single words or even parts of words to be edited easily. The simplest way is to replace certain words by silence, which is done by highlighting the particular area and choosing "silence" in the "edit"-menu. If one wants to make the deletion recognisable, it's better to insert a tone of the same duration (menu "generate"—"tone") and delete the original section afterwards. [5]

This procedure is just as simple, but it has the disadvantage that information is lost which could be of importance for secondary analysis. Therefore an alternative—"virtual anonymisation"—should be considered. This involves not deleting parts of the audio that might permit the speaker or related institutions to be identified, but replacing them. This is a little more complicated. In general it is similar to what one would do with a written manuscript, i.e. replacing any relevant passage (e.g. "Smith inc.") with an alternative text (e.g. "textile company"). But unlike written documents these "texts" must be produced in advance. This can be done directly using Audacity, if the computer has a microphone. One clicks the red "recording button" in the upper area of the window and stops it with the "stop button"[3]. The newly spoken parts can now be inserted into the places where the passages to be deleted are. If the recordings weren't made with Audacity, the original document is opened and the replacement files are imported. The specific passages in the original document can now be highlighted and deleted (menu "edit"—"cut") and from the alternative document the related passages are copied (menu "edit"—"copy") and inserted ("edit"—"paste"). To find the relevant passages it is helpful to pay attention to the timeline in the upper part of the screen and make only one audio track audible by clicking its "solo" button. [6]

---

2   Digitalised audio data very often come in one of these formats. WAV complies with the audio format used on CDs, MP3 is a compressed format, in which the data require much less storage space, with hardly any loss of quality. Data may come in other formats, for instance, because it was collected with special recording equipment. In this case conversion to WAV or MP3 can usually be carried out using the recording equipment software, which can also often be downloaded directly from the vendor.

3   The recordings can also be made with other equipment (e.g. with an MP3 recorder). In this case one needs to ensure that all recordings and the original file have the same sampling rate (a kind of resolution/speed) so that parts from one recording can be replaced with parts from the other without changing speed and pitch.

3.2 Method 1: Modification by vocoder programmes

For the task described here several public domain software tools exist; one of the well-known ones is the Zerius vocoder, which is used in the following description.
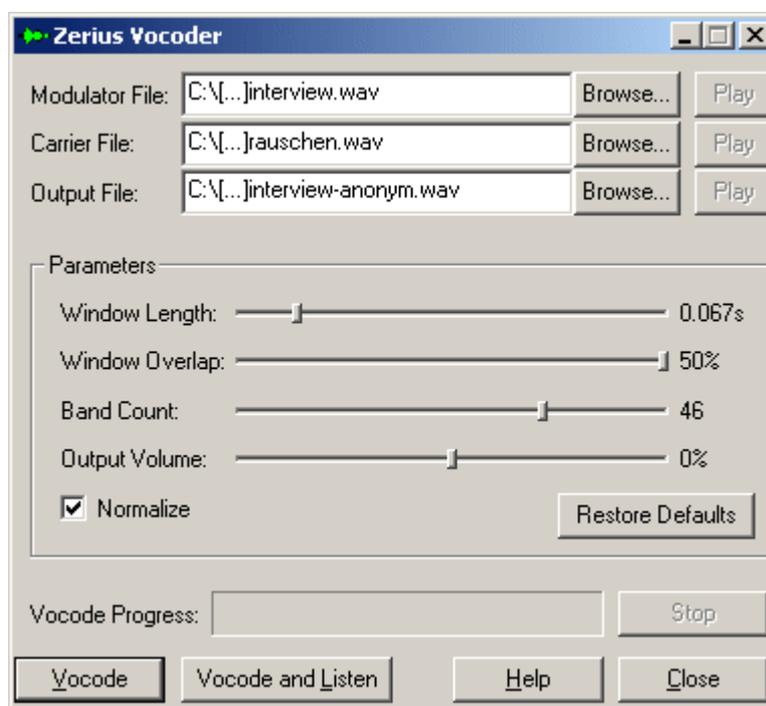

Figure 1: Interface of the Zerius vocoder [7]

The software is quite simple to use: the original audio file is selected as the "modulator file". As "carrier file" for the purpose in mind one uses white noise, which is a signal theoretically consisting of all frequencies of the audible spectrum. The programme itself is delivered with a white noise file called white.wav. The name of the modified file ("output file") can now be inserted. Clicking "vocode" starts the actual modification process. The result is an audio file in which the voice modulation of the speaker is transferred to the white noise, so that the voice is no longer recognisable (the sex of the speaker is also unidentifiable in the modified file)[4] The further switches of the programme usually work well with the defaults. If there are problems with acoustic comprehensibility, it might be helpful to try alternative values for the "window length"(the duration of passages which are processed separately by the programme). [8]

3.2.1 Problems

The most problematic aspect of this procedure is the loss of audio quality in the modified file. As long as the source was recorded carefully it is still reasonable, though understanding the speakers becomes a little more difficult, and it still

---

4 The speaker's voice becomes completely unrecognisable—but this doesn't mean that the speaker can't be identified at all. Identification might still be possible due to linguistic characteristics, from pronunciation through breaks to a particular vocabulary (see also Section 4).

keeps open all possibilities for further processing. Hesitations and breaks remain unaltered, stresses and nonverbal sounds are also still recognisable, though of less intensity. If, however, the source file is of poor quality, the resulting file may not be wholly comprehensible and analysable. [9]

Furthermore, due to the vocoder the noise replaces the speaker's voice. Therefore a mapping of statements to speakers by voice is no longer possible. In the case of interviews with one person it is helpful to have a stereo recording which allows one to separate the tracks and alter only that of the interviewee (or both with different parameters). Separating and recombining the tracks can be done with Audacity (the Zerius vocoder unfortunately does not allow different tracks to be treated differently in stereo recordings). [10]

3.3 Method 2: Modification by pitch changing

Pitch changing moves the whole spectrum of frequencies up or down—voices in the modified audio file are therefore higher or lower than in the original. The difference in pitch between the different voices and other characteristics are preserved with greater authenticity than with the vocoder. For pitch changing, public domain software also exists. The following description refers to the above-mentioned programme, Audacity.
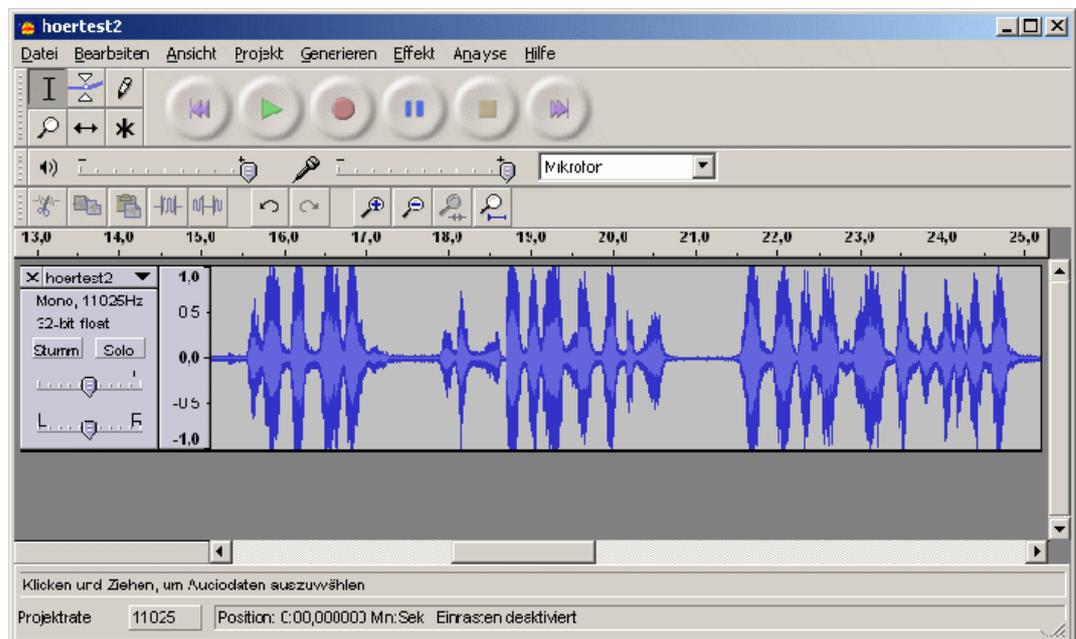

Figure 2: Main interface of Audacity [11]

This programme is also fairly easy to use. First the audio source is opened (Audacity handles different audio formats, including WAV and MP3). The "change pitch" option is in the "effects" menu. Here you select the amount of change (interval). It is not possible to give recommendations in this case, as the quality of the results of anonymisation and comprehensibility closely depend on the characteristics of the source file. Convincing modification which preserves good

comprehensibility is achieved by changing the pitch to a middle range which means raising low voices and lowering high ones. Good results are often achieved with a variation of about 3-5 half tone steps. Of course it is always necessary to compromise if the voices differ greatly in pitch; in case of doubt the anonymisation of the interviewee is of course the most important goal. [12]

Audacity offers a variety of other functions which can be useful for the preparation of audio data for (secondary) analysis. Some of them have already been mentioned (see section 3.1). Instead of describing these functions in detail some general remarks are in order: a central procedure is the highlighting of sections using the mouse. To do so the entire section in question should be visible in the window. This is done using the magnifying glass which allows the section to be enlarged or made smaller. If a procedure is to be applied to the whole file, it can be highlighted by pressing "ctrl-a". [13]

The result of modification can be tested by clicking the "preview" button. But changes can also be undone afterwards (menu "edit—undo"). If several tracks are to be edited simultaneously (e.g. copying parts of one track onto another), each track has to be imported (menu "project—import audio"). The result can be saved by highlighting the whole track ("ctrl-a") and exporting it afterwards (menu "file—export WAV"). [14]

3.3.1 Problems

Anonymisation using pitch shifting is "weaker" because more characteristics of the voice are kept; at the same time it makes further analysis easier. An audio file of bad quality, which would be totally unintelligible using the vocoder, might still be usable after pitch shifting. The deciding factor is the amount of anonymisation required. [15]

Voices which differ markedly in pitch can cause problems for pitch shifting. In such cases, much like the situation described above, a stereo recording is helpful as it allows one to modify different tracks in opposite directions. Furthermore in principle this allows only the voice of the interviewee to be changed, while leaving the interviewer's unaltered. [16]

4. How Far Anonymisation Goes

Today qualitative social research is subject to ethical standards which take anonymisation as a matter of course. "Anonymity of interviewed or tested persons must be ensured" (Deutsche Gesellschaft für Soziologie und Berufsverband Deutscher Soziologen, 1992, B5), for example, is the ethical statement of the German Sociological Society. But what is easy enough to fulfil within large-scale nationwide or international studies can turn out to be a difficult task for detailed qualitative studies with a regional focus (see HOPF, 2003). This problem occurs in principle with every secondary analysis; if you take, for instance, narrative interviews with a large biographical component, deleting names and other directly personal aspects from a transcript (or an audio file) wouldn't suffice to ensure

anonymity, if you couldn't ensure that researcher and interviewer aren't acquainted with each other. At the same time any anonymisation step goes hand in hand with a loss of information which reduces data quality. Therefore the researcher faces the dilemma of editing the information in a way that, on the one hand, preserves anonymity of the interviewee, and, on the other hand, keeps information loss to a minimum. This problem occurs with written data (or visual sources) as well as with audio data. [17]

We thus come to a basic problem for anomymisation of qualitative data, which we discuss for the sake of completeness, and which as far as audio sources are concerned, can be expressed as follows: a data handling method that meets ethical standards, requires that:

1. the interviewees know that the data gathered from them might be handed to other researchers under certain conditions, be it as audio source or as a transcript—the consent hereto may be obtained later;
2. anonymisation works; what is meant by "works" depends largely on the person who wants to analyse it. If researcher and interviewee are acquainted with each other, anonymisation requirements are very stringent (pitch shifting for instance usually wouldn't be enough). In certain cases secondary analysis might not be possible at all—and an analysis of text data might be objectionable if the anonymity of the interviewee can't be assured for reasons of social proximity between him or her and the researcher. [18]

This leads to a grey area as far as the technical part of anonymisation is concerned; it can only be responsibly clarified by the researcher. If the data are to be handed over to researchers, who are definitely not acquainted with the interviewees, pitch changing should be adequate to ensure anonymity—even if researcher and interviewee later meet each other by chance, recognition would be impossible as long as voice and articulation aren't very unusual. If, on the other hand, there is the possibility that the researcher knows the interviewee (or has at least a cursory acquaintance), then a higher degree of anonymisation—for example using a vocoder—is desirable, even if it leads to a lack of audio quality. [19]

Anonymisation of qualitative data is therefore a task which at the end of the day depends not only on technical options but on the responsible decisions of a researcher. This however applies to the reuse of any qualitative data. Technical methods like the ones described here allow the anonymisation of audio data to an extent, which in case of doubt provides no more cues for identifying the interviewee than a carefully made transcript does. To this extent they offer— without undergoing transcription—nearly identical opportunities for secondary analysis. [20]

5. Hints on Software

The software mentioned in the text can be found at the following sources:

1. Audacity: Programme for editing audio data which fulfils the following tasks: importing audio data in different formats (converting, if necessary); pitch shifting; deleting or altering certain sections within the audio file. Source: http://www.audacity.de/.

2. Zerius Vocoder: Vocoder-like modification of recordings. Source: http://www.zerius.com/. [21]

References

Bortz, Jürgen & Döring, Nicola (2003). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin: Springer.

Deutschen Gesellschaft für Soziologie und Berufsverband Deutscher Soziologen (1992). *Ethik-Kodex*, http://www.bds-soz.de/Ethik.pdf [17.12.2004].

Garz, Detlef & Kraimer, Klaus (1994). *Die Welt als Text*. Frankfurt/M.: Suhrkamp.

Hopf, Christel (2003). Forschungsethik und qualitative Forschung. In Uwe Flick, Ernst von Kardorff & Steinke, Ines (Eds.), *Qualitative Forschung. Ein Handbuch* (pp.588-600). Reinbek: Rowohlt Taschenbuch.

Kowal, Sabine & O'Connell, Daniel (2003).Zur Transkription von Gesprächen. In Uwe Flick, Ernst von Kardorff & Ines Steinke (Eds.), *Qualitative Forschung. Ein Handbuch* (pp.437-447). Reinbek: Rowohlt Taschenbuch.

Mayring, Philipp (1999). *Einführung in die qualitative Sozialforschung*. Weinheim: Beltz.

Kuckartz, Udo (2003): Qualitative Daten computergestützt auswerten: Methoden, Techniken, Software. In Barbara Friebertshäuser & Annedore Prengel (Eds.), *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft* (pp.584-595). Weinheim: Juventa.

Author

Dr. *Henning PÄTZOLD*, Jun.-Professor for adult education. Areas of interest: Teaching and learning in adult education, counselling, pedagogy of media, teacher education.

Contact:

Dr. Henning Pätzold

TU Kaiserslautern
Dpt. of Social Sciences
P.O. Box 3049
D 67653 Kaiserslautern, Germany

Tel.: ++ 49 / (0)89 / 2180 – 3802

E-mail: paetzold@rhrk.uni-kl.de
URL: http://www.uni-kl.de/paedagogik/

Citation