Anonymising long sounds for prosodic research.

Daniel Hirst[1,2]

[1]LPL, UMR 7309 CNRS, Aix-Marseille University, Aix-en-Provence, France
[2]School of Foreign Languages, Tongji University, Shanghai, China

`daniel.hirst@lpl-aix.fr`

Abstract

It is more and more standard practice, in speech research, to make publicly available the data used in the research, in particular the speech recordings. This can potentially raise the problem of how to respect the anonymity of the speakers, particular if the recordings consist of unmonitored conversations, which may contain references to people by name or other material which it may be thought preferable not to make public. This paper describes a simple procedure, originally proposed by Paul Boersma, which has been implemented as a Praat script. The script replaces portions of the original recording annotated with a key word by a *hum* sound, which reproduces the prosodic characteristics (fundamental frequency and intensity envelope) of the corresponding original speech signal. The script described is freely available from the *Speech and Language Data Repository* (http://sldr.org/sldr000526/en).
Index Terms: database, speech prosody, anonymisation.

1. Introduction

Research on speech in the last decades has shown a tendency to make use of larger and larger quantities of speech recordings for analyses. It has, furthermore, become standard practice to make the primary data, as far as possible, publicly available in order to make it possible for other researchers to check and replicate the analyses.

Making speech recordings publicly available can, in some cases, come into conflict with the need to respect the anonymity of the speakers. This is particularly true when the material recorded consists of unmonitored spontaneous conversation material in which the speakers may make explicit reference to themselves or to other people by name. Besides the question of preserving anonymity, there may be other reasons for not wishing to allow some parts of the recorded material to be in full public access, when, for example, one of the speakers gives some intimate information about someone else.

One radical solution to this problem is that which has been adopted for the oral part of the British National Corpus (BNC [2]), and which consists of replacing the portions of the recording which contain personal information by a silence.

The obvious disadvantage of this solution is that it makes it impossible to use those parts of the corpus for prosodic analysis since the prosodic information has been removed along with the speech signal. A preferable solution would be to modify the speech signal so that the lexical content of the signal is no longer recognisable but without removing the prosodic information.

2. Anonymising speech signals

In psycho-acoustic experiments, one solution which is often used is to 'invert' the spectrum of the speech signal ([3], chap 33 p. 619). This renders the speech signal inintelligible without removing the prosodic information. Unfortunately, a second inversion of the spectrum of the acoustic signal will restore the original signal and consequently the technique is not appropriate for the task we are describing.

A second technique, also used in pyscho-acoustic experiments, renders the speech signal inaudible by applying a (usually low-pass) filter, which removes most of the spectral information without removing the prosodic information. The problem with this technique is that even quite severe filtering can sometimes still leave the original signal intelligible. In order to be sure that the signal is not intelligible, the filtering tends to make the original signal sound so muffled that it is difficult to use it for perceptual evaluation of prosodic information, for example.

A different technique was suggested by Paul Boersma on the Praat-users mailing list in 2006, (http://uk.groups.yahoo.com/group/praat-users/message/2537) in response to a query on how to anonymise speech data.

The algorithm proposed consisted of the following steps:

1. Select the word in the editor window
2. Extract the selected word as a Sound to the Objects list (File menu)
3. Choose Get Intensity (dB) from the Query menu
4. Choose To Pitch
5. Choose To Sound (hum)
6. Select the original extracted word in the Objects list
7. Choose To Intensity
8. Choose Down To IntensityTier
9. Select the hummed sound plus the IntensityTier
10. Choose Multiply
11. Select the sound and choose Scale Intensity from the Modify menu
12. in the Scale Intensity window, type the value from step 3 and click OK
13. In the window with the original sound, select the same times as in step 1.
14. Choose Cut (Edit menu; the word disappears).
15. Select the hummed sound and choose Edit.

16. In the sound window, select the whole sound (the hum) and choose Copy (Edit menu).

17. Go back to the window with the original sound and choose Paste (Edit menu).

Paul Boersma noted that since there may be some irregularities at the edges, some of the cutting and copying could be carried out on zero crossings.

3. A Praat script to anonymise long sounds

The algorithm described in the preceding section was implemented as a Praat script which can be used to anonymise a whole folder of speech sounds which are treated as long sounds.

Instead of using the editor as described in the algorithm, the script uses a TextGrid annotation object on which a specified tier contains target labels indicating the portions of the signal which need to be 'anonymised'. The original sound is then 'dissected' and the labeled segment of sound is treated as described in the algorithm before being splice back with the rest of the sound.
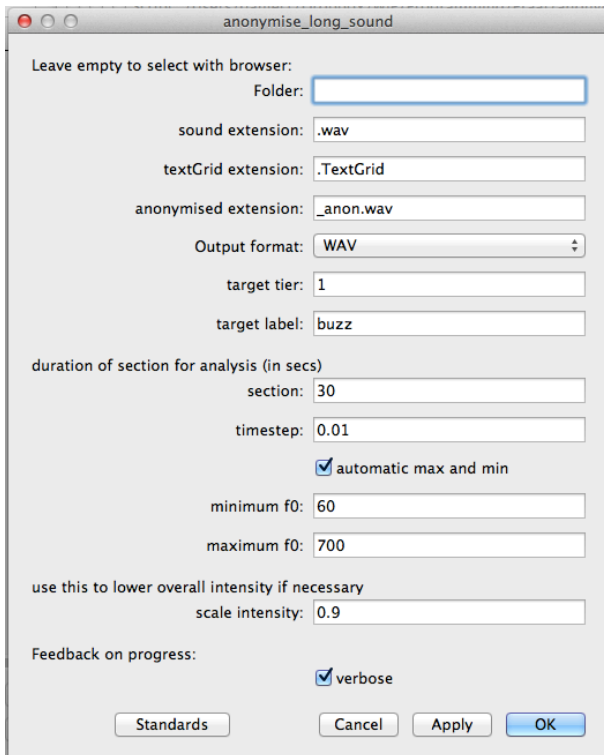

Figure 1: *Script window for the Praat script* anonymise_long_sounds.praat.

The path to the folder can either be specified directly in the script window, or selected with the browser (see Figure 1).

By default, the script expects to find Sounds and TextGrids in the folder with the extensions .wav and .TextGrid, respectively. The script can also read and write other sound formats.

Each sound is associated with a TextGrid containing at least one interval tier as target (the default target tier is 1). Intervals which are labelled with the *target label* (default is *buzz*) will be replaced by a *hum* sound which has the same pitch and intensity envelope as the original sound.

In order to be able to apply the script to sounds with a duration exceeding that which can be handled in the computer's RAM, the sounds are treated as LongSound objects, which means that instead of loading the whole sound into the computer memory, only one section of the sound (the duration of which is specified in seconds (default = 30) is treated at a time.

Figure 2 shows the sentence "Could you arrange to send an engineer on Tuesday morning please?" with the f0 display and intensity profile. The words "an engineer" have been marked with the label "buzz".
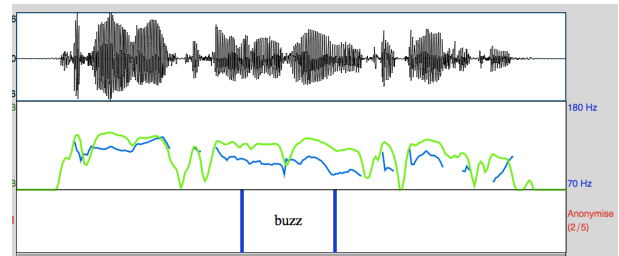

Figure 2: *The sentence "Could you arrange to send an engineer on Tuesday morning please?"*

In Figure 3, the words "an engineer" have been replaced by a buzz with the same pitch and intensity profile. The words are thus incomprehensible but the sentence can be used for the analysis of prosodic parameters without any appreciable loss of information.
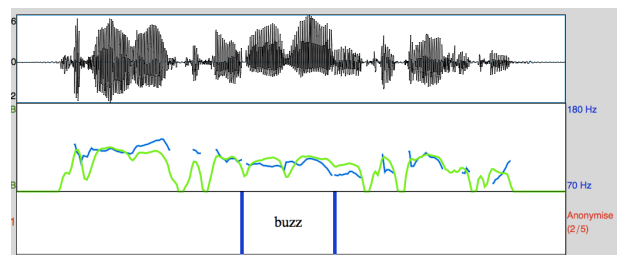

Figure 3: *The same sentence as that in figure 2 with the words "an engineer" replaced by a buzz with the same pitch and intensity profile.*

4. Conclusions

The script is freely downloadable from *SLDR*, the Speech and Language Data Repository, at the following address:

http://sldr.org/sldr000526/en

5. References

[1] Boersma, P.; D. Weenink, D. Praat, a system for doing phonetics by computer. http://www.praat.org [version 5.3.41, February 2013], 1992 (2013).

[2] Burnard, Lou (ed.) Reference Guide for the British National Corpus (XML Edition). British National Corpus Consortium, Research Technologies Service, Oxford University Computing Services, http://www.phon.ox.ac.uk/SpokenBNC 2007

[3] Smith, Steven W. Digital Signal Processing: A Practical Guide for Engineers and Scientists Elsevier Science, Burlington, MA. 2003.