

ROCme! software for the recording and management of speech corpora

*Emmanuel Ferragne*¹, *Sébastien Flavier*², *Christian Fressard*²

¹ CLILLAC-ARP, EA 3967, Université Paris 7, France

² Laboratoire Dynamique du Langage, UMR 5596, CNRS – Université de Lyon, France

rocme@ish-lyon.cnrs.fr

Abstract

ROCme! – Recording of Oral Corpora Made Easy – has been designed to allow a sensible, autonomous, and dematerialized management of speech recordings. Users can create forms for the collection of metadata. Speakers autonomously fill in questionnaires, record, play, and save audio; and browse sentences (or other types of corpora) by clicking on buttons or using keyboard shortcuts. ROCme! can display text, optionally with HTML and CSS formatting, images, and play sounds or videos.

1. Introduction

ROCme! was developed to meet the needs of researchers involved in the collection of speech corpora. It is now available for free download at www.ddl.ish-lyon.cnrs.fr/rocme. ROCme! relies on Adobe AIR runtime system to offer a user-friendly graphical interface that is compatible with Windows and Mac platforms (Android soon). The basic requirements were to: i) allow speakers to record autonomously a series of sentences displayed on the computer screen, ii) collect speaker metadata with a fully customizable XML questionnaire, iii) guarantee consistency with the creation of projects.

2. Main features

ROCme! can help researchers to collect oral data with optimal consistency: users can define a number of constraints that will contribute to the homogeneity of a series of recordings: this is what ROCme! projects are for. In addition to this, users can create forms to collect speaker metadata. Finally, ROCme! can also be used as a simple recorder with user-defined templates for file names.

2.1. ROCme! projects

Through projects, users can choose to randomize the presentation of stimuli, display a mask with a fixation point between the beginning of the recording and the presentation of a new stimulus (in order to avoid missing the speaker's response). Projects can be set so that saving clipped signals to disk is disabled. Other options include i) the possibility to set a password which will prevent speaker's from accidentally exiting the prompt mode, ii) a clock display (which is useful when speakers are paid by the hour), iii) a special mode where each speaker only reads part of the corpus – and then a special view is available to show the amount of overlap among speakers (split mode), iv) another special mode – repeat mode – where one sentence (or the reaction to one stimulus) can be recorded several times. And of course, audio options (sampling rate, bit depth, number of channels) are fully customizable.

2.2. Speaker metadata

As far as speaker metadata are concerned, users can define an option to let speakers fill in the questionnaire on their own. ROCme! users can create questionnaires for the collection of speaker metadata thanks to XML tags. The XML code may be entered directly in a dedicated field in the GUI, and a preview of the final questionnaire is then available. Figure 1 shows an example of speaker metadata form, with the XML code first and the corresponding GUI preview below. The code includes the 6 special data types that ROCme! can handle (age, date, number, single, choice, multiple). Specifying a data type is optional but it is strongly recommended because ROCme! can then process these data, compute descriptive statistics, and display useful graphs allowing, for instance, the user to instantly spot potential imbalances in the project and then readjust her strategy to recruit participants. The metadata can then be exported to a spreadsheet program, and the graphs can be saved as picture files. Figure 2 shows a real example from a study by Stephan & Ferragne [1]. The pie chart shows a clear imbalance between the number of men and women in the study.

2.3. ROCme! as a simple recorder

ROCme! also features a compact mode corresponding to a simple recorder with basic functions like record, stop, replay, pause, which can be triggered by buttons or keyboard shortcuts. In compact mode, no project is required. Users can define an output directory and an automatic output file name template that guarantees maximal consistency in the naming of sound files. Figure 3 shows the dialog box for automatic file naming.

3. Future enhancements

Based on user feedback, ROCme! is in constant evolution to meet the needs of the research community involved in the recording of speech corpora. Firstly, as mentioned earlier, ROCme! will soon be ported to Android systems. Working with mobile devices constitutes a good opportunity for linguists and speech scientists to dramatically reduce the weight and size of their recording hardware. Secondly, for the time being, the latest stable version of ROCme! can only record sound files; we are planning to add the possibility to record videos very soon. Thirdly, while, for the moment, users must manually create a text file containing the corpus to be displayed (e.g. with explicit path names to media files), a special GUI will soon be available for the creation of corpus files.

```

<user>
  <field id="Name" required="required"></field>
  <field id="Sex" type="single" required="required">
    <field id="Male"></field>
    <field id="Female"></field>
  </field>
  <field id="Age" type="age"
required="required"></field>
  <field id="Birthdate" type="date"></field>
  <field id="Number_of_children"
type="number"></field>
  <field id="Place_of_residence" lbl="Place of
residence" type="choice">
    <field id="North"></field>
    <field id="South"></field>
  </field>
  <field id="Languages_spoken" type="multiple">
  <group>
    <field id="Language"/>
  </group>
</field>
</user>

```

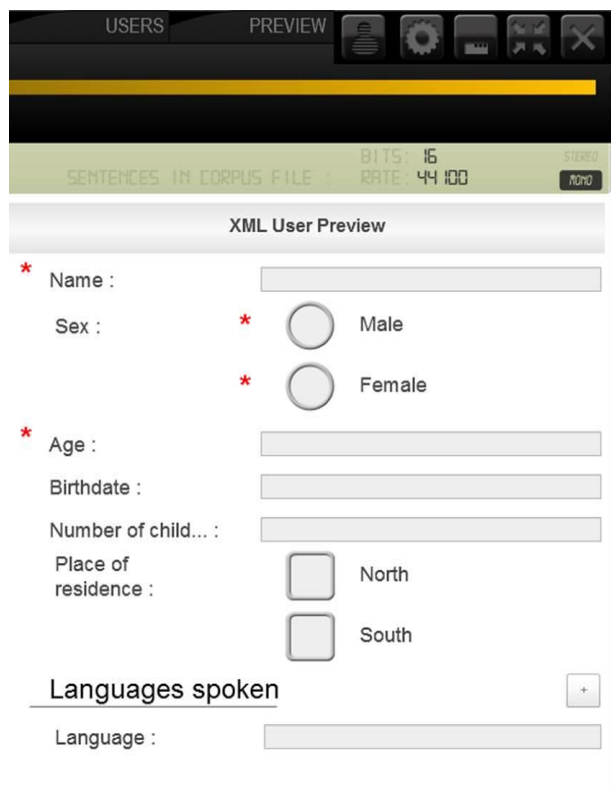


FIGURE 1: XML code and GUI version of a ROCme! questionnaire

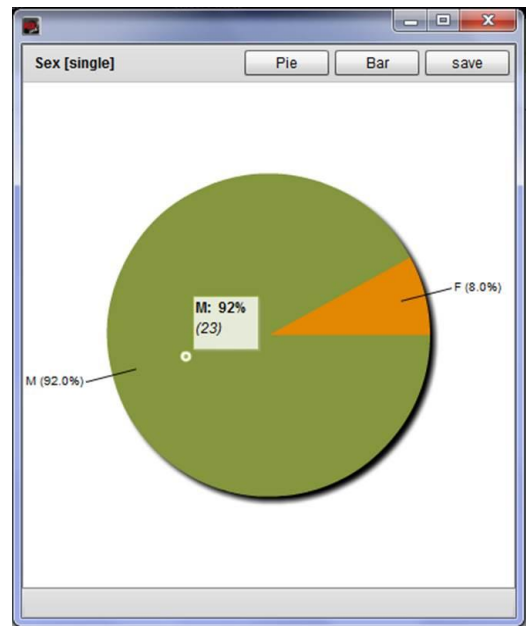


Figure 2: Pie chart showing a great bias towards male (M) speakers in a ROCme! project

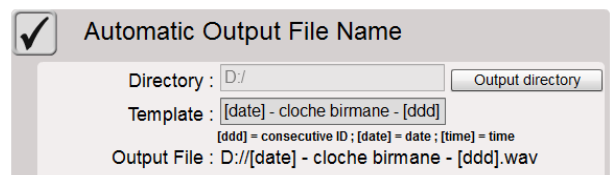


Figure 3: Dialog box for automatic file naming

4. Acknowledgements

The ROCme! project is supported by a BQR grant from Université Paris 7 Diderot and receives technical support from Laboratoire Dynamique Du Langage.

5. References

[1] Stephan, P. and Ferragne, E. "Acoustic analysis of atypical contrasts in Northern- Irish English", *Variation and Language Processing 2 Conference*, Christchurch, New Zealand, 16-18 January, 2013.