

A Tool For Building Multilingual Voice Questionnaires

Alejandro Armando
Université de Genève

Pierrette Bouillon
Université de Genève

Manny Rayner
Université de Genève

Nikos Tsourakis
Université de Genève

ABSTRACT

We describe a prototype platform for creating multilingual voice questionnaires. Content is defined using a simple form-based language with units for questions, question-groups and answers; questionnaire definitions are compiled into efficient speech recognition packages and tables, and the resulting applications can be deployed over the web on both desktop and mobile platforms. We sketch our initial questionnaire application, which is designed for gathering information related to availability of anti-malaria measures in sub-Saharan Africa. It contains 114 question-groups and 218 questions.

1. Introduction

There are many circumstances where it is potentially useful to be able to administer multilingual voice questionnaires. A familiar example in Western society is admission to the accident and emergency room of a hospital: the nurse on duty will most likely start by asking for personal details, the nature of the patient's immediate problem, previous medical history, and so on. If the nurse and the patient do not share a common language, difficulties arise. Another example, which will occupy us more in this paper, is information gathering for demographic and health topics (DHS Program 2014).

We describe an easy-to-use architecture, inspired by the RAMP framework (Salihu 2013) which can be used to generate voice questionnaires of this type. Our questionnaires are deployed on mobile platforms — smartphones, tablets or laptops — linked over a 3G connection to a remote server. The person administering the questionnaire chooses the next question by speaking it in their own language. The application uses speech recognition, performed on the server, to identify it, speaks a pre-recorded translation in the respondent language, and displays a set of icons on the touch-screen. The respondent answers by pressing on the icons; each icon has an associated voice recording, in the respondent language, identifying its function.

In the rest of the paper, we briefly sketch the architecture, focussing on the formalism used to define questionnaires, and present an example questionnaire application.

2. Architecture of the system

The questionnaire designer specifies the questionnaire using a single file written in a simple formalism which supports three types of unit: Groups, Questions, and Answers. A Group specifies a top-level item on the questionnaire, a list of permitted Fillers, and a pointer to the next Group. A Question specifies one possible way to attempt to assign a Filler to a Group; it defines the Group which the Question belongs to, a list of surface realization of the question in the questionnaire administrator's language, a translation in each target language, and a list of permitted Answers, each one optionally associated with a Filler. An Answer defines an associated translation for each target language. The questionnaire description is compiled into an accurate limited-vocabulary speech recogniser and a set of tables. It is deployed over the web using methods developed at Geneva University and elsewhere in the context of various speech translation and CALL projects (Rayner et al 2006, 2014; Fuchs et al 2012).

To give a simple example of the formalism, the following lines specify a Group called Religion, which belongs to a questionnaire called `toy_questionnaire`.

```
Group
Questionnaire toy_questionnaire
Name Religion
PrintName What is your religion?
Code REL
Fillers christian muslim no_religion
Next WhichIslam If muslim
Next MotherTongue If christian no_religion skip
EndGroup
```

The other fields have the following meanings. `PrintName` is the title displayed at the top of the screen when the Group is reached, and `Code` is the identifier used to store the result when the filled questionnaire is printed out. There are three possible ways to fill the `Religion` slot: `christian`, `muslim` and `no_religion`. Once the slot has been filled, the questionnaire moves to the Group `WhichIslam` (what type of Muslim) if the selected filler was 'muslim', and moves to the Group `MotherTongue` on other fillers.

A Group will normally have several Questions associated with it; indeed, the value of the tool, compared to an application which mechanically asks a set of questions in a fixed order, is that it allows human judgement to be used to select an appropriate questions. In the current case, one Question might involve directly asking the interviewee what their religion is:

```
Question
Group Religion
Variant what is ( your | your family's ) religion
Variant what religion do ( you | your family ) belong to
Translation/French Quelle est votre religion ?
Translation/Spanish ¿Cuál es su religión?
```

```
Answers christian muslim no_religion
EndQuestion
```

Here, the `Variant` lines define different ways for the (English-speaking) interviewer to ask the question, the `Translation` lines define how it will be asked in the different target languages covered, and the `Answers` line lists possible responses; as shown, it is possible to use regular expressions in the `Variant` lines in order to list different forms more compactly.

Similarly, another question might instead ask the interviewee if they are Muslim:

```
Question
Group Religion
Translation/French Êtes-vous musulman?
Translation/Spanish ¿Es usted musulmán?
Variant are you ?a muslim
Answers yes=muslim no
EndQuestion
```

Here, the element `yes=muslim` in the `Answers` line indicates that the answer `yes` should be interpreted as indicating the filler `muslim`. Evidently, the situation may make some questions more appropriate than others; in certain cases, the difference may be very important.

The surface forms of the answers are defined by `Answer` units; a typical example is the following, which specifies how to realize the `Answer` `yes`.

```
Answer
Content yes
PrintName Yes
Translation/French Oui
Translation/Spanish Sí
EndAnswer
```

3. Example questionnaire application and user interface

Our initial case study has used a questionnaire consisting of 114 `Groups` and 218 `Questions`, designed for gathering information related to availability of anti-malaria measures in sub-Saharan Africa. The overall goal is to measure nine internationally recognized vector control indicators that can be generated through household surveys. Some of the questions are about preventive measures — namely, the use of mosquito nets and indoor residual spraying — and bear a more immediate relevance to such indicators; others are about the characteristics of the household's dwelling unit and are important to establish a connection between the socioeconomic status of the population and public health issues. All questions were taken from two model questionnaires recommended by the cluster of humanitarian actors that developed the indicators, which include the World Health Organization and the Roll Back Malaria Partnership.

When using the application, the first step is to choose the specific questionnaire the user wants to administer and the language of the interviewee. At the moment, the available choices

are to administer malaria-relevant modules of the Multiple Indicator Cluster Survey (MICS) or the Malaria Indicator Survey (MIS) in French or Spanish. Adding other target languages more suitable for the intended context of application would only involve writing extra Translation lines in the relevant Question definitions.

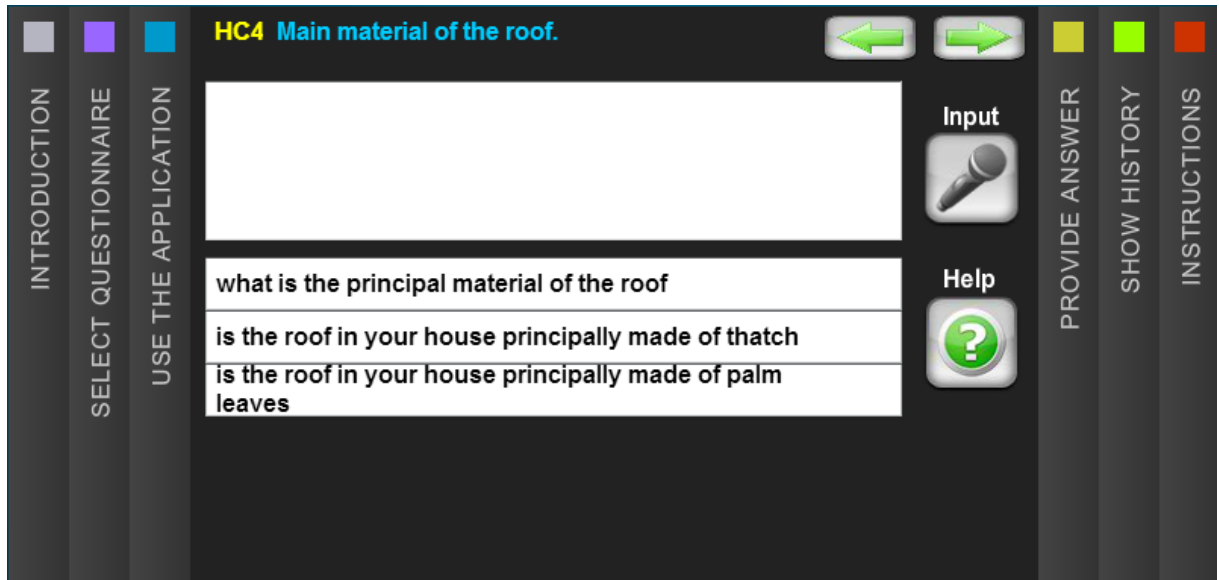


Figure 1: User interface example, main tab. The current Group, “HC4: Main material of the roof” is shown at the top; the user has just pressed the Help button (bottom right) to get a list of example questions (bottom center). They can ask a question by holding down the Input button (top right) and speaking. The interface shows the recognition result. If this is correct, the questionnaire administrator switches to the PROVIDE ANSWER tab.

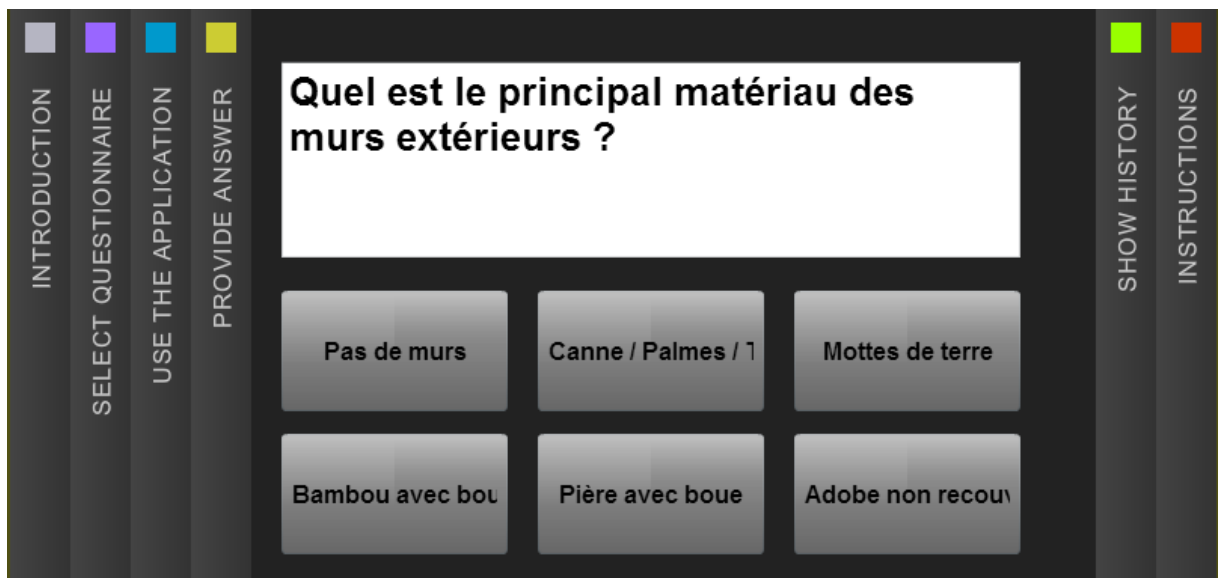


Figure 2: User interface example, PROVIDE ANSWER tab. The top panel shows the interviewee language translation of the question (here in French). The clickable buttons at the bottom are the available answers, and are also displayed in the interviewee language.

After choosing the questionnaire and the language, the user is directed to the operative tab of the application (Figure 1), where they can start administering the questionnaire. The title at the top of the screen corresponds to a question in the original paper-based questionnaire. Just as in any conventional situation, the user can read the prompt out loud or use their own judgement to rephrase the question or ask something else. The prototype allows for some degree of syntactic and lexical reformulation, as well as for a number of relevant alternatives, including all possible yes-no questions stemming from predefined answers. For instance, instead of asking “What is the floor made of?” the interviewer can opt for “Is the floor made of wood?” if he or she deems it more appropriate in view of the context. When the interviewer has spoken the question, it is displayed in the Provide Answer tab together with the available answers (Figure 2).

The application reproduces the sequence structure of the original questionnaire, so the next question is always determined by the answer of the question at hand. For instance, the application will automatically skip question HC12 in the MICS questionnaire (“How many hectares...?”) if the respondent gives a negative answer to HC11 (“Do you own any agricultural land?”). Similarly, the application will force the interviewer to probe for additional information whenever it is required. All questions and answers are registered and shown in the show history tab, and there is a final tab, instructions, that contains question-specific guidelines for the interviewer.

This application is intended for use in a context where household surveys are conducted by mixed teams of international and local interviewers, as it is often the case for this type of programme-monitoring surveys. A mobile device integrating speech recognition and automatic translation can help international organizations reduce the need of training ad hoc teams of local people or having to rely on unqualified interpreters.

References

- DHS Program. 2014. <http://dhsprogram.com/>
- Fuchs, M., Tsourakis, N. and Rayner, M. 2012. A Scalable Architecture for Web Deployment of Spoken Dialogue Systems. Proc. LREC. Istanbul, Turkey.
- Rayner, M., Hockey, B.A. and Bouillon, P. 2006. *Putting Linguistics into Speech Recognition*. CSLI Press.
- Rayner, M., Baur, C. and Tsourakis, N. 2014. CALL-SLT Lite: A Minimal Framework for Building Interactive Speech-Enabled CALL Applications. Proc. Second Workshop on Child-Computer Communication. Singapore.
- Salihu, M. 2013. Piloting the first-ever RAMP survey using DHIS mobile. Proc. 141st APHA Annual Meeting. Boston, MA.