Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 81 (2016) 61 – 66

Procedia Computer Science

5th Workshop on Spoken Language Technology for Under-resourced Languages, SLTU 2016, 9-12 May 2016, Yogyakarta, Indonesia

Parallel Speech Collection for Under-resourced Language Studies using the Lig-Aikuma Mobile Device App

David Blachon[a], Elodie Gauthier[a], Laurent Besacier[a,*], Guy-Noël Kouarata[b], Martine Adda-Decker[b], Annie Rialland[b]

[a]*Laboratoire d'Informatique de Grenoble (LIG) / GETALP group, France*
[b]*LPP, CNRS-Paris 3 / Sorbonne Nouvelle, France*

Abstract

This paper reports on our ongoing efforts to collect speech data in under-resourced or endangered languages of Africa. Data collection is carried out using an improved version of the Android application Aikuma developed by Steven Bird and colleagues [1]. Features were added to the app in order to facilitate the collection of parallel speech data in line with the requirements of the French-German ANR/DFG BULB (Breaking the Unwritten Language Barrier) project. The resulting app, called Lig-Aikuma, runs on various mobile phones and tablets and proposes a range of different speech collection modes (recording, respeaking, translation and elicitation). Lig-Aikuma's improved features include a smart generation and handling of speaker metadata as well as respeaking and parallel audio data mapping. It was used for field data collections in Congo-Brazzaville resulting in a total of over 80 hours of speech. Design issues of the mobile app as well as the use of Lig-Aikuma during two recording campaigns, are further described in this paper.

Peer-review under responsibility of the Organizing Committee of SLTU 2016

Keywords: Speech collection tool; under-resourced languages; language documentation

1. Introduction

The growing proliferation of smartphones and other interactive voice mobile devices offers new opportunities for field linguists and researchers in language documentation. The ease of collecting large volumes of data lowers the pressure of defining the best sampling selection process, which speakers and what data exactly to collect. We may thus envision endangered language documentation collections growing very large with many speakers and material to study a bunch of linguistic phenomena, from acoustic-phonetics to discourse analysis, including phonology, morphology and lexicon, grammar, prosody and tonal information. Field recordings should also include ethnolinguistic material which is particularly valuable to document traditions and way of living. However, large data collections require well organized repositories to access the content, with efficient file naming and metadata conventions.

* Corresponding author. Tel.: +33-4-76-63-56-95.
E-mail address: laurent.besacier@imag.fr

Peer-review under responsibility of the Organizing Committee of SLTU 2016
doi:10.1016/j.procs.2016.04.030

In the following, we describe our ongoing efforts to achieve an effective mobile device application which easily records primary linguistic data in the speakers' living area. The application automatically generates structured data files with raw speech data as well as field and speaker related metadata. We also describe our experience with collecting speech during two recent fieldtrips.

1.1. Context: the BULB project

The BULB[1] (Breaking the Unwritten Language Barrier) project aims at supporting the documentation of unwritten languages with the help of automatic speech recognition and machine translation technologies. The project relies on a strong German-French cooperation involving both linguists and computer scientists. The aim of the project is to design and adapt NLP methods to speed up linguistic analyses of oral, unwritten languages from the Bantu language family. Bantu includes about 500 languages (figures vary from 440[2] to more than 660[3,4] depending on the authors), among those many are exclusively oral and remain unstudied. Towards this aim, we have chosen three languages which are already partly studied and for which a few written resources exist: Basaa (A43), Myene (B11) and Mboshi (C25). In this contribution, we report on our experience in Mboshi data collection using Lig-Aikuma.

1.2. The respeaking concept

The model of Basic Oral Language Documentation, as adapted for use in remote village locations, which are far from digital archives but close to endangered languages and cultures, was initially proposed by Bird[5]. Speakers of a small Papuan language were trained and observed during a six week period. For this, a technique called respeaking, initially introduced by Woodbury[6], was used. Respeaking involves listening to an original recording and repeating what was heard carefully and slowly. This results in a secondary recording that is much easier to transcribe later on (transcription by a linguist or by a machine). The reason is that the initial speech may be too fast, the recording level may be too low, and background noise may degrade the content. For instance, in the context of recording traditional narratives, elderly speakers are often required (and they may have a weak voice, few teeth, etc.) compromising the clarity of the recording[7].

1.3. Paper outline

This paper is organized as follows. Section 2 describes the original application, from which we evolved to Lig-Aikuma, described in section 3. Then section 4 describes first data collections made in Congo-Brazaville. Finally section 5 concludes and gives some perspectives.

2. Aikuma - The origins

2.1. The initial Aikuma application and its motivations

The initial smartphone application Aikuma was developed by Bird et al.[1] for the purpose of language documentation with an aim of long-term interpretability. According to the authors, the application is designed for a *future philologist*[1]: it collects enough speech and documentation to allow for a delayed (future) processing by a linguist. Indeed, the authors notice that, in general, language documentation projects lack of resources, especially human resources for the processing of the materials. As a consequence, data may be processed a long time after their collection. Moreover, in the case of an endangered language, there is a risk that the study starts after the language has disappeared. This is why the authors extended the concept of respeaking to produce oral translations of the initial recorded material. Oral translation is performed by listening to a segment of audio in a source language and spontaneously producing a spoken translation in a second language.

Aikuma is an Android application that suppports these requirements: recording of speech, respeaking of audio sources, and oral translation. In the next section, we describe its main features.

[1] http://www.bulb-project.org

2.2. Main features of the initial app

In Aikuma, both respeaking and oral translation work the same way: the speaker listens to segments of an audio source file, then speak them, either in the same language (respeaking) or in another one (oral translation). The produced segments are concatenated altogether and result in a single respoken (or translated) file. Once concatenated into a file, the segments may not be easy to distinguish, this is why a text file is also produced which contains the alignment of every segment of both audio files. Figure 1 illustrates such an alignment of segments.

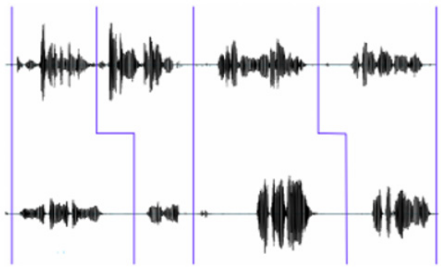

Fig. 1. Illustration of the alignment of segments from two audio files [1].

Additional features were thought for an easy and wide use. The user interface is composed of icons and is text-less, allowing anyone to use it. The application comes with a wireless "sync and share" feature that allows several devices to share the audio files. This is useful for expanding respeakings and translations among groups of persons. Recording documentation also includes the storage of basic metadata files including information on the speaker, the language and the location.

The authors report to have used Aikuma in remote indigeneous communities (e.g., in Brazil and in Nepal[7]) and have collected rather large amounts of audio this way.

3. From Aikuma to Lig-Aikuma

3.1. Motivations and specifications

With a tool like Aikuma, the motivation of the BULB project is to collect speech and perform processings on it. Specific use cases are identified within the project and associated to a series of tasks of interest: basic audio recording; respeaking and oral translation; and finally, elicitation of speech following the display of text, image or video.

Because the BULB project targets those use cases, some adjustments to the existing application were proposed. First, the user interface has been adapted so as to make it easier to focus on and select a task. Then, following the objective to make the application quick and easy to use, a feature is proposed for saving and loading the metadata of the latest recording. Then, when a user wants to make a new recording with the same speaker, the metadata form is already filled in. The next sections describe the most important features.

3.2. Recording modes

The core features of the initial Aikuma for recording, respeaking and translation have been kept, along with the storage of metadata about the speaker; also, some parts of the interface have been reused.

On top of that, new developments have focused on the setup of 4 modes, dedicated to the previously mentioned tasks. The home view is illustrated on Figure 2 (left). As one can see, the following four modes are identified:

- Free **recording** of spontaneous speech,
- **Respeaking** a recording (previously recorded with the app or loaded from a wav file): the respeaking allows now to listen (optionally) to the latest recording segment so as to check it and respeak it if needed, before going

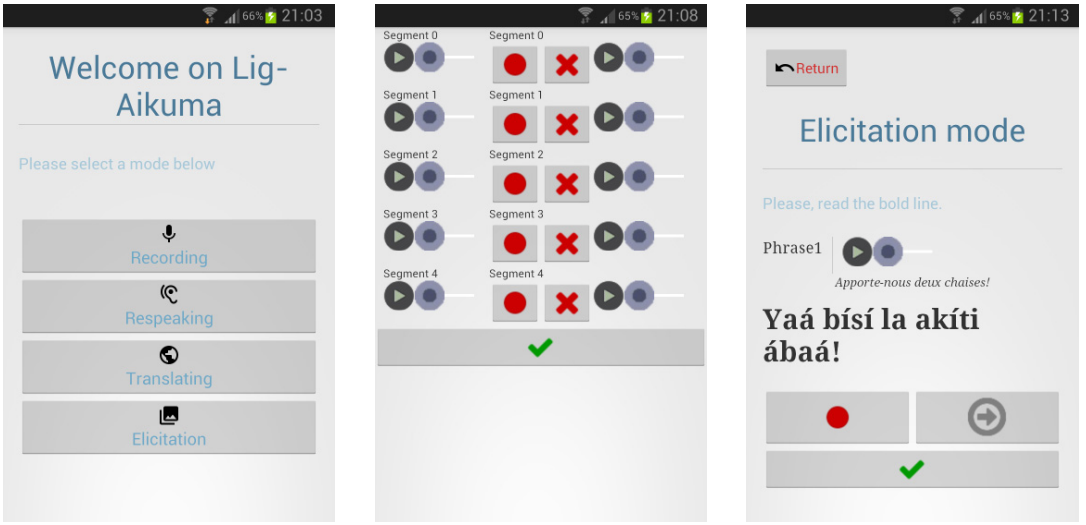

Fig. 2. Screenshots from the Lig-Aikuma application: from left to right, i) the home view ; ii) the summary view after respeaking is done, the speaker may play and edit every segment; iii) the elicitation mode

to the next segment. Also, once the respeaking is finished, a summary view displays the new segments and their corresponding original segments and allows to (optionally) listen to or respeak any of them before finishing the session. On Figure 2 (middle), one can see that original segments are aligned with respoken ones. Both can be played while the latter can also be recorded if necessary, which is useful for double checking and error correction,

- **Translating** a recording (previously recorded or loaded): same features as for the respeaking mode except that the source and target languages must be different,
- **Eliciting** speech from a text file (image and video media will follow very soon): the user loads a text file within the app, then reads the sentence, speaks, listens to the recording for checking and goes to the next sentence, etc. This mode was specifically required for the data collection which took place in Congo-Brazaville during summer 2015 (described in Section 4). Figure 2 (right) illustrates the text elicitation mode.

3.3. Metadata and file naming

For every mode, a metadata file is saved with the recording file. Metadata are filled in a form before any recording. Unlike the initial version of Aikuma which proposed four different views for that purpose, the form summarizes all the questions within a single interface.

In addition, metadata have been enriched with new details about the languages (language of the recording, mother tongue of the speaker and other spoken languages) and about the speaker (name, age, gender, region of origin). Moreover, in order to save time, a feature saves the latest metadata as a session and uses it to pre-load the form the next time it is necessary.

The files are now named using the specific following format: *DATE-TIME_DEVICE-NAME_LANG*. As an example *2015_07_22_17_00_00_Samsung_tablet_fra* is the name of a recording made on July, 22nd 2015, at 5pm (17:00:00), in French language, on a Samsung device.

3.4. Other updates

The interface has been adapted for the large screens of tablets (10 inches), so the app works both on Android powered smartphones and tablets. Apart from the specifications, based on multiple discussions with linguist colleagues, this new version was developped in approximately 3 man-months and generated 5000+ lines of code. All the new

Fig. 3. Examples of the use of Aikuma on Android tablets for data collection: elicited verb conjugations spoken by a native Mboshi woman (left) and free conversations involving several speakers (right).

code has been put on the LIG forge and is accessible open source[2] for use or development on demand. Following the licence of the initial Aikuma application, the Lig-Aikuma application will be released under the same GNU Affero General Public License[3].

The application Lig-Aikuma has been successfully tested on different devices (including Samsung Galaxy SIII, Google Nexus 6, HTC Desire 820 smartphones and a Galaxy Tab 4 tablet).

3.5. Downloading Lig-Aikuma

Users who just want to use the app without access to the code can download it directly from the forge direct link[4]. In a web browser, type in the url address and a popup window may appear to require the install confirmation (user must have authorized the install from other sources than Play Store).

4. First data collection

4.1. Objectives and data preparation

Successful fieldwork, in particular data collection, requires careful preparation. In our project, we specifically had to consider both linguistic and technological requirements, as the overall aim is to develop NLP tools for linguistic work. Prior to the trip, we discussed the specifications to meet both linguists' and computer scientists' requirements.

Beyond standard ethical issues and consent forms, specifications included the collection of "linguistically dense" data e.g., elicited verb conjugation or small sentences designed to collect most useful data for basic oral language description complying with traditional fieldwork. Figure 3 illustrates the use of the Lig-Aikuma tablet to collect Mboshi verb conjugations (left) or more free conversations including several speakers (right). Our objective was to collect large volumes of data (at least several tens of hours) from dozens of speakers in different speaking styles. All possible written documents including a 2000-entry Mboshi dictionary[8], traditional tales and biblical texts in Mboshi were gathered. Furthermore, a large part of the 6000 reference sentences for oral language documentation[9] were translated and written in Mboshi by one of the authors (Guy-Noël Kouarata). These written sources were processed to enter the elicitation recording mode in Lig-Aikuma. A further objective of the fieldtrip was to take tradition and culture related images and videos. These can be used as language-independent speech eliciting material in later collections of purely oral languages. They can also be used to illustrate language resources, such as online dictionaries and education material, which can be seen as a fair payback to the speaker community who accepted to contribute to the study.

[2] https://forge.imag.fr/projects/lig-aikuma/
[3] https://www.gnu.org/licenses/agpl-3.0.en.html
[4] https://forge.imag.fr/frs/download.php/706/MainActivity.apk

4.2. Data collected

Practically, two 1-month fieldtrips to Congo-Brazzaville were done by one of the authors (Guy-Noël Kouarata) who is a native speaker of Mboshi. The data recording campaign was organised as follows: four *Galaxy Tab 4* tablets were given to four main contact persons who were responsible of the material and who were responsible of recruiting volunteer speakers from their relatives, friends and broader acquaintances. The main contact persons were also asked to contribute to the respeaking process. All speakers were paid and the main contact persons will be given the tablets at the end of the collection process as additional reward.

So far, within the BULB project, 48 hours of speech data in Mboshi language were collected with Lig-Aikuma in Congo-Brazzaville. The corpus is composed of 33 hours of spontaneous speech (mostly debates and stories), 9.5 hours of controlled speech (including conjugations), 2 hours of read sentences (collected using the elicitation mode) and 3.5 hours of read speech. Another collection of 1000 elicited sentences was performed. A collection of French translations of the respoken utterances is currently achieved.

5. Conclusion

This paper presented Lig-Aikuma, a recording application for language documentation. A first data collection of language resources in Mboshi was conducted and further developments and features are planned in a near future. One of the goals of the BULB project is also to automatically extract knowledge from the parallel speech data collected (like Mboshi / French parallel signals). This could be used, for instance, to automatically learn word units (as well as their pronunciation) in an unknown (and unwritten) language with very little supervision.

Acknowledgements

This work was realized in the framework of the French-German ANR-DFG project BULB (ANR-14-CE35-002) and also supported by the French Investissements d'Avenir - Labex EFL program (ANR-10-LABX-0083).

References

1. Bird, S., Hanke, F.R., Adams, O., Lee, H.. Aikuma: A mobile app for collaborative language documentation. *ACL 2014* 2014;1.
2. Guthrie, . *Comparative Bantu, Volume 2*. Farnborough: Gregg Press; 1971.
3. Mann, M., Dalby, D.. *A thesaurus of African languages*. London: Hans Zell Publishers; 1987.
4. Maho, J.. *A classification of the Bantu languages: an update of Guthrie's referential system*. Routledge; 2003, p. 639–651.
5. Bird, S.. A scalable method for preserving oral literature from small languages. In: *Proceedings of the 12th International Conference on Asia-Pacific Digital Libraries*. 2010.
6. Woodbury, A.C.. *Defining documentary linguistics*; vol. 1. Language Documentation and Description, SOAS; 2003, p. 35–51.
7. Bird, S.. Collecting bilingual audio in remote indigenous communities. COLING; 2014.
8. Roch Paulin, B., Chatfield, R., Kouarata, G., Embengue-Waldschmidt, A.. *Dictionnaire Mbochi-Français*. Congo (Brazzaville): SIL-Congo Publishers; 2000.
9. Bouquiaux, T., Thomas, J.. *Enquête et description des langues à tradition orale*. Paris: SELAF; 1976.