Speech Technologies for African Languages: Example of a Multilingual Calculator for Education

Laurent Besacier, Elodie Gauthier, Mathieu Mangeot, Philippe Bretier, Paul Bagshaw, Olivier Rosec, Thierry Moudenc, François Pellegrino, Sylvie Voisin, Egidio Marsico, et al.

Speech Technologies for African Languages: Example of a Multilingual Calculator for Education

Laurent Besacier[1], Elodie Gauthier[1], Mathieu Mangeot[1], Philippe Bretier[2], Paul Bagshaw[2], Olivier Rosec[2], Thierry Moudenc[2], François Pellegrino[3], Sylvie Voisin[3], Egidio Marsico[3], Pascal Nocera[4]

[1]Laboratoire d'Informatique de Grenoble (LIG), Univ. Grenoble Alpes, Grenoble, France
[2]Voxygen S.A. - Pleumeur Bodou, France
[3]Laboratoire Dynamique Du Langage (DDL), CNRS - Université de Lyon, France
[4]Laboratoire d'Informatique d'Avignon (LIA), Univ. Avignon, France
laurent.besacier@imag.fr

Abstract

This paper presents our achievements after 18 months of the ALFFA project dealing with African languages technologies. We focus on a multilingual calculator (Android app) that will be demonstrated during the *Show and Tell* session.

Index Terms: ASR, TTS, Android app, African languages

1. Introduction

1.1. Context: the ALFFA project

Today is very favorable to the development of a market for speech technologies in African languages. People's access to ICT is done mainly through mobile (and keyboard) and the need for voice services can be found in all sectors, from higher priority (health, food, education) to more fun (games, social media). For this, overcoming the language barrier is needed and this is what we propose in the ALFFA project[1] where two main aspects are involved: fundamentals of speech analysis (language phonetic and linguistic description, dialectology) and speech technologies (ASR and TTS) for African languages. In the project, developed ASR and TTS technologies will be used to build micro speech services for mobile phones in Africa. For this, speech fundamental knowledge for targeted languages has to be upgraded while African language technologies are still at their very beginning.

For these reasons, the ALFFA project is really interdisciplinary since it does not only gather technology experts (LIA, LIG, Voxygen) but also includes fieldwork linguists/phoneticians (DDL).

1.2. Demo Content

This paper describes our achievements after 18 months of project. We notably present the last version of a multilingual calculator prototype in several African languages (including Wolof, Hausa and accented French). This calculator - developped by the industrial partner of the ALFFA project (Voxygen) - was already presented at the Francophonie's summit 2014 in Dakar (Senegal)[2]. Voxygen is currently looking for partners, especially in the world of education, who could deploy this educational tool at a large scale.

[1]http://alffa.imag.fr

[2]http://voxygen.fr/fr/content/voxygen-devoile-sa-calculatrice-parlante-dakar (Web page in French)

2. ASR and TTS systems developed

2.1. Target languages

Language choice for the project is mainly governed by population coverage and industrial perspectives. We focus on Hausa spoken by around 60 million people, as first or second language. Enlarging the sub-region coverage, we will also consider Bambara, Wolof, and Fulfulde languages, to cover major West Africa languages. Bambara is largely spoken in West Africa by around 40 million people. Wolof is mainly spoken by 10 million people. Fulfulde is a set of dialects spoken in all West Africa countries by 70 million people. Those languages cover more than half of the 300 million people of West Africa. They include both tonal and non-tonal languages.

As far as East Africa is concerned, we designed ASR system for Swahili which is the most widespread language in the East of the continent: more than 60 million people. We also work on Amharic, mostly spoken in Ethiopia (20 millions).

A linguistic description of the targeted languages (white paper) is planned for each target language. So far, white papers were written for Wolof and Swahili.

2.2. ASR

ASR systems for Swahili, Hausa and Amharic have been built so far. All the data and scripts to build a complete ASR system for Swahili and Hausa are already available to the public on a github repository [3]. We used Kaldi speech recognition toolkit [1] for building our ASR systems. For the Swahili and Amharic ASR systems, the transcribed speech corpora, pronunciation lexicons and LMs are also made available while for Hausa ASR, users need to buy the corpus and the lexicon at ELDA first.

More details on the Swahili corpus and how it was collected can be found on [2]. For Hausa, the GlobalPhone Speech Corpus [3] was used. The Amharic system was retrained from the corpus described in [4]. A summary of the ASR performance obtained for the three languages is given in table 1 but more experimental details can be found in the README files of the *github* repository. Current data collection and ASR developments include Wolof language.

2.3. TTS

A viable manner of addressing language development for TTS stems from the incremental approach followed by Voxygen for several years now. The key point of this methodology is that instead of developing the entire (under-documented and under-resourced) language at once, only part of the language is ad-

[3]https://github.com/besacier/ALFFA_PUBLIC/tree/master/ASR

Table 1: ASR performance for Swahili, Hausa and Amharic - HMM/SGMM acoustic modeling - all scripts available on *github*

Task	WER (%)
Swahili broadcast news	20.7
Hausa read speech	10.0
Amharic read speech	8.7

dressed at each stage. *Step 1* consists in developing a TTS module able to read any numerical information (such as time, volume, price, date ...). Application specific sentences may then be designed with numerical slots that are dynamically filled by TTS and automatically inserted seamlessly in the sentences (e.g. a speaking calculator, a mobile service providing agricultural market prices, etc.). *Step 2* extends to the ability to fill any slot with a single word (merchandise name, company name, place name...) in pre-defined sentences. *Step 3* (iterative) aims at full language coverage taking into account word functions and contexts, to provide fluid speech with adequate prosody. In the ALFFA project, first step TTS prototypes have been developed for Hausa and Wolof and some fillers have been recorded for the multilingual speaking calculator needs. New developements concern Fulfulde, Zarma en Serere.

3. Multilingual calculator for education

Education is one of the straightforward domain for which speech enabled micro-applications are needed. In the ALFFA project, we propose a multilingual speaking calculator to demonstrate the potential of speech technologies for African languages. The prototype, which is an android app, offers different features: a simple *voice-enabled calculator* where the user enter the operation from the numeric keyboard (or vocally[4]) then the application reads the operation and the result ; a *multiplication table mode* where the user chooses a number then the calculator recites the corresponding table ; and a *quizz mode* where the user is invited to submit a complete equation and the calculator confirms if the answer is correct or not. A button allows to switch the language at any time or to make the calculator repeat the last utterance in the new selected language.

The ALFFA project will progressively extend the speaking calculator with new African languages: Fulfulde, Zarma and Serere are coming shortly. Bambara and Swahili will rapidely follow. Field testing of the multilingual speaking calculator prototype is envisaged in the project.

4. Machine-assisted language analysis

Automatic speech technology offers great opportunities for investigating a wide range of issues in laboratory phonology. However, its availability hardly extends beyond a small number of languages. In the ALFFA project, we develop ASR for African languages, so linguists can, in the meantime, benefit from automatic tools (automatic annotation tools, forced alignment and segmentation, etc.). For instance, we did a preliminary study on vowel duration in Hausa depending on their syllabic context. Figure 1 shows that for two hausa vowels /e/ and /o/, the vowel length contrast (short / long) can be highlighted with automatic tools. We see on the figure that duration in closed syllabic context is mostly between 20 and 80 milliseconds while it is going from 90 to 240 milliseconds for open context. This figure confirms, at a large scale (5k utterances), linguists' description of vowel contrast phenomenon in Hausa.

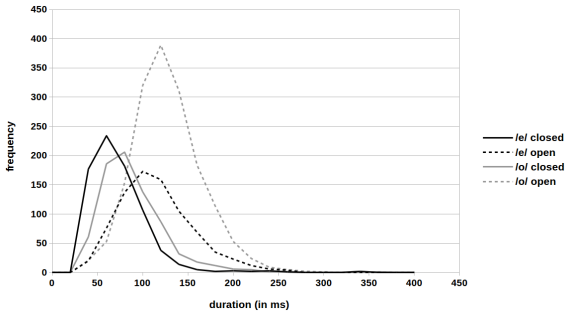

Figure 1: *Illustration of the use of automatic tools for laboratory phonology: distribution of /e/ and /o/ in Hausa, depending on their syllabic context (from HMM forced-alignments)*

5. Project highlights and demo content

We present below some other highlights and achievements of the ALFFA project which also tries to promote speech technology for under-resourced (and especially African) languages by organizing events (conferences, workshops, special sessions).

- Swahili ASR system of LIG is now included in the KALDI[5] trunk
- Voxygen presented its multilingual calculator at the Francophonie's summit 2014 in Dakar (Senegal)
- LIG is involved in French-German BULB project (*Breaking the Unwritten Language Barrier*)
- Fulfulde – English – French dictionaries were computerised and made available on the Web[6]
- DDL organized the *Sénélangues 2015* spring school on west african languages description

A video (in French) presenting the multilingual calculator was recently published on Youtube [7]. We will present a poster with the last achievements obtained within ALFFA project and the multilingual calculator (Android app) will be demonstrated.

6. Acknowledgements

This work was done in the framework of the ALFFA project, funded by ANR (*Agence Nationale de la Recherche*).

7. References

[1] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlíček, Y. Qian, P. Schwarz *et al.*, "The kaldi speech recognition toolkit," in *ASRU*, 2011.

[2] H. Gelas, L. Besacier, and F. Pellegrino, "Developments of swahili resources for an automatic speech recognition system," in *SLTU*, Cape-Town, South Africa, 2012.

[3] T. Schlippe, E. G. K. Djomgang, N. T. Vu, S. Ochs, and T. Schultz, "Hausa large vocabulary continuous speech recognition." in *SLTU*, 2012.

[4] M. Tachbelie, S. T. Abate, and L. Besacier, "Using different acoustic, lexical and language modeling units for asr of an under-resourced language - amharic," *Speech Communication*, vol. 56, 2014.

[4] this latter feature in currently under development

[5] `http://kaldi.sourceforge.net/`

[6] see `http://papillon.imag.fr/` choose "DicoFulNiger" - "DicoFulUS"

[7] `https://www.youtube.com/watch?v=vpIxYZN8rt4` will be screened during S&T