HAL
open science

Automatic Speech Recognition for African Languages with Vowel Length Contrast

Elodie Gauthier, Laurent Besacier, Sylvie Voisin

To cite this version:

Elodie Gauthier, Laurent Besacier, Sylvie Voisin. Automatic Speech Recognition for African Languages with Vowel Length Contrast. 5th Workshop on Spoken Language Technologies for Under-resourced Languages (SLTU), May 2016, Yogyakarta, Indonesia. 10.1016/j.procs.2016.04.041 . hal-01350040

HAL Id: hal-01350040

https://hal.science/hal-01350040v1

Submitted on 29 Jul 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 00 (2016) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

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

Automatic Speech Recognition For African Languages With Vowel Length Contrast

Elodie Gauthier[a], Laurent Besacier[a,*], Sylvie Voisin[b]

[a] *Univ. Grenoble Alpes, LIG, F-38000 Grenoble, France*
[b] *Laboratoire Dynamique Du Langage (DDL), CNRS - Université de Lyon, France*

Abstract

This paper deals with ASR for two languages: Hausa and Wolof. Their common characteristic is to appear with vowel length contrast. In other words, two versions (short / long) of a same vowel exist in the phoneme inventory of the language. We expect that taking into account this contrast in ASR models might help and this is what we investigate in this pilot study. The experimental results show that while both approaches (vowel length contrast modeling or not) lead to similar results, their combination allows to slightly improve ASR performance. As a by-product of ASR system design, we also show that the acoustic models obtained allow a large scale analysis of vowel length contrast for phonetic studies.
© 2016 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the Organizing Committee of SLTU 2016.

Keywords: African languages; automatic speech recognition; vowel length contrast;

1. Introduction

Today is very favorable to the development of a market for speech in sub-saharan african languages. On this continent, people's access to information and communication technologies (ICT) is done mainly through mobile phones and the need for voice services can be found in all sectors, from higher priority (health, food) to more fun (games, social media).

1.1. Context: the ALFFA project

This paper is done in context of the ALFFA project (see `http://alffa.imag.fr`). ALFFA project aims to develop automatic speech recognition (ASR) and text-to-speech (TTS) technologies which would be used to build micro speech services for mobile phones in Africa. Moreover, the project is really interdisciplinary since it does not only gather technology experts but also includes fieldwork linguists/phoneticians.

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

1877-0509 © 2016 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the Organizing Committee of SLTU 2016.

1.2. Paper objective and contributions

In this paper we deal more precisely with ASR for two languages: Hausa and Wolof. Their common characteristic is to appear with vowel length contrast. In other words, two versions (short / long) of a same vowel exist in the phoneme inventory of the language. We expect that taking into account this contrast in ASR models might help and this is what we investigate in this pilot study.

More precisely, two state-of-the-art ASR systems are developped for both Hausa and Wolof languages. To our knowledge, the Wolof ASR system presented in this paper is the first large vocabulary continuous speech recognition system ever developed for this language (only a digit and small vocabulary recognition system in Wolof was previously presented in[1]). Then, we systematically compare approaches taking into account (or not) the vowel length contrast and present a preliminary solution to take advantage of both of them.

1.3. Focus on the languages studied

Hausa and Wolof are two african languages, largely spoken in the West of the continent.

Hausa is part of the Afro-Asiatic phylum. The language is one of the most spoken languages in all the continent with more than 35 million of speakers (mostly in Niger and Nigeria) as their first language. Hausa is also widely used as a common language with an additional 15 million people using it as second and third language. In the Latin based script, there are 5 vowels (i.e.: /a/, /e/, /i/, /o/, /u/) plus 2 diphthongs /ai/ and /au/ and 29 consons.

Wolof is part of the Niger-Congo phylum. The language is considered as one of the common languages of West Africa. It is mostly spoken in Senegal, in Gambia and in Mauritania. For our studies, we focus on the Senegalese Wolof. According to recent estimations more than 5 million Senegalese people speak Wolof as their mother tongue (40% of Senegal population) while more than 50% of Senegalese would speak Wolof as their second or third language[2]. There are 20 consons and the same 5 vowels as in Hausa, plus the phoneme /ɛ/, /ə/ and /ɔ/. Wolof is not learned at school since French is the official language. The orthography has no real standard rules and can be written in many ways. Nonetheless, the Center of Applied Linguistics of Dakar (CLAD)[1], coordinates the orthographic standardization of the Wolof language.

Hausa and Wolof have length contrast at the phone level. Nonetheless, contrast can be located on vowels (except for [ə]) as well as on consonants in Wolof while in Hausa this feature is only for vowels. In addition, there are three tones in Hausa : high, low tone and falling tone[3], while Wolof is a non-tonal language. Length contrasts and tones modify the meaning of a word. Both languages can be written by a variant of the Arabic script (called *'Ajami* in Hausa and *Wolofal* in Wolof) or the Latin script (called *Boko* in Hausa). The latter is however considered as the standard nowadays for both languages but short/long distinction and tones are not marked (while it is in Arabic script).

1.4. Paper outline

This paper is organized as following. In Section 2, we summarize the works done in automatic speech recognition for under-resourced languages and mention some studies dedicated to vowel length contrast modeling in ASR. In Section 3, we describe the data we collected and used for our experiments. Then, Section 4 illustrates the vowel length contrast in Hausa using a large scale machine assisted analysis. In Section 5, we compare our ASR systems for Hausa and Wolof when handling a vowel duration model or not. We also present a simple way to combine length-contrasted and non length-contrasted CD-DNN-HMM models for ASR. Finally, Section 6 concludes this paper and gives a few perspectives.

2. Background & Related Works

In the last decades, several works have been conducted in order to aid for the identification, exploration and legitimization of under-resourced languages. Besacier et al., in 2005, was the first to propose working on ASR

[1] see `http://clad.ucad.sn/`

for Vietnamese and Khmer low-resourced languages[4]. Later on, work was published on African languages: Somali[5], Amharic[6], South African languages[7]. Vu et al.[8] developed a method for a rapid adaptation of an automatic speech recognition system for Vietnamese. Recently, a survey on ASR for under-resourced languages has been published[9]. Moreover, concerning the description and the analysis of language features, Gelas[10] has shown the relevance of multilingual acoustic models to study, at a large scale, particular phenomena of languages (like vowel length contrast in Punu, a Bantu language spoken in Gabon). More precisely, his work assessed the relevance of an automatic alignment for studying duration patterns, even in a language previously unknown from the automatic alignment tool. Results reported that the expected bimodal distribution of vowel durations could be correctly (and automatically) detected. Also, since phonemic analysis is a fundamental part of the description and documentation of a language, the use of automatic methods has been investigated to identify the contrastive units of sounds in language. This process consists in taking a phonetic transcription and checking for contrast between pairs of phones, which can be tedious so[11] proposed to automate parts of the process.

As far as duration modeling in speech recognition is concerned,[12] modeled word duration at the acoustic modeling level. He improved the word error rate (WER) of the speech recognition system by rescoring N-best lists with the duration models built. Few years later,[13] worked on duration modeling on both word and phone levels. Lattice rescoring was used but the WER reduction was limited. Nonetheless, the technique implemented allows to improve the transcription's quality when combining systems. Pylkkönen also focused on the phone length, in Finnish where duration is important for the meaning of the word. He compared in[14] different duration modeling techniques. A 8% relative reduction of the letter error rate compared to the baseline system (not handling duration) was obtained. However, the decoding speed was reduced. Finally,[15] used these approaches on phone duration modeling and applied them to the Estonian case. He analyzed the phone duration probabilities to create a duration model. Three different phone lengths exist in Estonian (short, long and overlong). A method was presented based on linguistic and phonological characteristics of phone as well as their surrounding phones context to calculate a decision tree that classifies phones into groups of similar durations. Word error rate of his ASR system was, again, only slightly improved.

3. Experimental Settings

3.1. Speech and text data used

3.1.1. The Globalphone Hausa corpus

We used the Hausa GlobalPhone Speech Corpus[16] to train our ASR system in Hausa. We used all the resources provided. This corpus is a read speech corpus, recorded in Cameroon with native Hausa speakers. The training set is composed of 24 males, 58 females and contains 5,863 utterances corresponding to 6.6 hours of speech data. For the decoding, the development set provided is composed of 4 male, 6 female and represents 1,021 utterances (total speech duration is 1h).

3.1.2. Speech and text data collection for Wolof

We collected our own data to build an ASR system for Wolof. We exploit a set of very few electronic documents available for this language, gathered as part of[17] to build our read speech corpus. It was recorded in Dakar (Senegal) with native Wolof speakers (10 males, 8 females - from of 24 to 48 years old). The 18,000 recorded utterances represent 21.3 hours of signal. The training set is composed of 8 male, 6 female and contains 13,998 utterances (about 16.8 hours of speech data) while the decoding (dev) set is composed of 1 male, 1 female and contains 2,000 utterances (about 2.2 hours of speech data). We also own an evaluation (test) set - not used so far - composed of 1 male and 1 female, representing 2.3 hours of speech (2,000 utterances read). Later, in order to build a language model with more textual data written in Wolof, we collected on the Web the Universal Declaration of Human Rights, the Bible and a book written by a humanist. We also crawled the Wikipedia database in Wolof using Wikipedia Extractor[18]. All these texts were cleaned (i.e: by converting to lower case, by removing HTML/XML tag and punctuation marks). More description about the data collected in Wolof is available in[19].

3.2. Baseline ASR systems for Hausa and Wolof

We used Kaldi speech recognition toolkit[20] for building our ASR systems for both languages. Two systems based on different acoustic modeling techniques were built: one based on the classical hidden Markov model and Gaussian mixture model (CD-HMM/GMM) approach and another one using deep neural networks (DNNs). To build DNNs, we trained the network using state-level minimum Bayes risk[21] (sMBR) and the network had seven layers, each of the six hidden layers had 1024 hidden units. For the CD-HMM/GMM system, the acoustic models were built using 13 Mel-frequency cepstrum coefficients (MFCCs) and GMMs on 6.6h training data for Hausa and on 16.8h for Wolof. We trained triphone models by employing 2,887 context-dependent states for Hausa and 3,401 context-dependent states for Wolof, and 40K Gaussians for both. The standard Kaldi recipe for hybrid CD-HMM/GMM and CD-DNN-HMM ASR training was used (see[19] for more details). Concerning the language model (LM), we used a trigram model, in the decoding process, for each Hausa and Wolof ASR systems. For Hausa, we used the language model provided by[16]. It was built from a text corpus of about 8M words. Its perplexity, calculated on the decoding set (first-mentioned in 3.1.1) is 88 (0.19% of out of vocabulary words (OOVs)). For Wolof, we built our own statistical language model with the SRILM toolkit. We interpolated a language model built from two LMs: the first model was built from the very few electronic documents mentioned in 3.1.2 and the second from Web data we crawled. Finally, this interpolated language model of Wolof is built from a text corpus of 615,631 words. Its perplexity that we measured on the decoding set (abovementioned in 3.1.2) is 314 (5.4% of OOVs). About the pronunciation dictionary, we also used the one of[16] for the training and decoding stages of the Hausa ASR. Initially, it included some tags related to tone of the vowels but we have not used these annotations. This pronunciation dictionary contains 38,915 entries and 33 phonemes. For Wolof, we used our in-house pronunciation dictionary which contains 32,039 entries and a phoneme inventory of 34 phonemes. At this stage, phoneme inventories (for both languages) do not take into account any vowel length contrast.

We can see in the table 1 below the performance for the first Hausa and Wolof ASR systems, trained using both CD-HMM/GMM and CD-DNN-HMM approaches. These baseline performances show that our first Hausa ASR system can reach WER below 10%, even without any special modeling of the vowel length, which is very encouraging. The acoustic models of Hausa ASR will be used for machine analyzing vowel length in the next section.

Table 1: Results according to the baseline acoustic models - Hausa and Wolof ASR - no modelisation of length so far

Language	**CD-HMM/GMM**		**CD-DNN-HMM**	
	WER (%)	CER (%)	WER (%)	CER (%)
Hausa	13.0	3.7	8.0	2.1
Wolof	31.7	12.9	27.2	10.2

The poor performance of the ASR system for Wolof can be explained by the high perplexity of the language model, by the fact that we only have two speakers in the test set (while 10 speakers for Hausa). Moreover, compared to the Hausa where in the language model we had 0.19% of unknown words, for Wolof we had 5.4% of unknown words in the language model. Another problem with Wolof ASR is the lack of normalisation in the writing of words which penalizes both language model and WER. For this reason, we also display the character error rate (CER) in our results tables since CER is less sensitive to spelling normalization issues.

4. Illustration Of Vowel Length Contrast For Hausa

In this section, we focus on the Hausa language only. Same analysis on the Wolof language is currently investigated.

In Hausa, the length contrast is not marked in the standard orthography. Vowel length vary depending on several factors: the syllabic structure type in which the vowel is, its position into the word (initial, middle, final) and also if the vowel is in pre-pausal position[22]. In this section, we only focus on the place of the vowel in the syllable. Therefore, we decided to use the forced-alignments provided during ASR system training. We split the words of the utterances into syllables. The syllable structure of Hausa can be closed (i.e: the syllable ends by a consonant, its shape is CVC) or open (i.e. the syllable ends by a monophtong or a diphthong, its shape is CV or CVV).[3] explains that, in a closed syllable, vowels are always short whereas in an open (but not prepausal) syllable a length contrast can appear.[23] specifies that vowels /e/ and /o/ have a longer duration in open syllabic context than in closed context. Consequently, for

closed syllabes we expect to have shorter vowels than for open syllables. According to the literature, we automatically assigned tags to the vowels depending on the structure of the syllable. Information about the initial or final phoneme was given thanks to the forced-alignment. We applied a window of 4 phonemes (1 before the current phoneme, the current phoneme, 2 after the current phoneme) in order to know if the syllable was open or closed. Thus, this simple labelling (closed versus open syllables) allows us to analyze the duration of the vowels according to the syllabic context.

We forced-aligned the 5,863 training set utterances with the CD-HMM/GMM acoustic model, extracted vowels durations thereof and calculated their distribution according to the closed/open label. We measured for each of the five Hausa vowels the delta of their mean duration in closed versus open syllabic context of the word, using forced (CD-HMM/GMM) alignments. We conclude that differences in vowel length are only significant and observable for /e/ and /o/. Indeed, we measured a delta of 41.2 milliseconds for /e/ and 38.2 milliseconds for /o/ while for /a/, /i/ and /u/ the delta was between 2 and 9 milliseconds only. We think this is because the length of /e/ and /o/ vowels is more predictable according to their syllabic context. Indeed, as a reminder,[23] explains that vowels /e/ and /o/ have strict rules: these are short in closed syllable while they can only occur long in an open syllable. For the three others vowels /a/, /i/ and /u/, the linguists do not always agree about their phonetic realization and our empirical results confirm that the syllable context (open/closed) alone does not allow to predict their duration. Other rules are necessary, as taking into account the vowel position into the word. We can observe on the figure 1 the distribution of /e/ and /o/ Hausa vowels. The figure shows that vowels have a different duration depending on whether they are part of a closed or an open syllable. As we expected, vowels in a closed syllable are shorter than those in an open syllable. These empirical measurements, at a large scale, encourage us to try to model vowel length contrast in our ASR systems. This is what we propose in the next section.

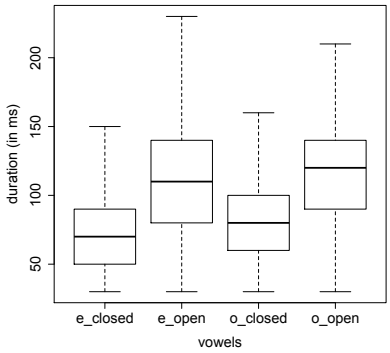

Fig. 1: Distribution of /e/ and /o/ depending on their syllabic context, obtained by CD-HMM/GMM forced alignment - Hausa language.

5. Modeling Vowel Length Contrast In ASR

To see if taking into account the short/long vowel duration of /e/ and /o/ helps the ASR system, we trained new Hausa acoustic models by forcing the system to represent this contrast (i.e. having different units in the phone set). While in Hausa the duration of the vowels varies depending on its syllabic context, in Wolof the length mark can be deduced from the orthography by a duplication of the grapheme, so it was easier to label the length contrast in Wolof (marked as either a "_short" or a "_long" label). Based on these facts, we added the length mark in the pronunciation dictionary according to the language treated. Moreover, to follow our experiments of previous section on the Hausa language, we only marked length contrast for vowels /e/ and /o/ into the pronunciation dictionary. For the wolof, we distinguish the length contrast for all the vowels of the language. In the table 2, we summarize for each ASR system the label set used for vowel contrast. In other words, for the Hausa, we added on /e/ and /o/ either a "_closed" either

an "_open" label, depending to the syllabic context [2]. Thus, the Hausa pronunciation dictionary was modified to take into account this according to the position of the vowel into the syllable. For example, the word "aboki" which means "friend, companion" has a V-CV-CV syllabic structure. So, /o/ was labeled "o_open" (expected long). As an another exemple, "bos" which means "bus" has a CVC structure. The /o/ is annotated "o_closed" (expected short) given that it is surrounded by consonants. For the Wolof, we added on the vowels either a "_long" label if the grapheme was duplicated in the word orthography, either a "_short" if the grapheme was single. So, for the word "ameel" (which means "to owe something to somebody"), the vowel /a/ was annotated "a_short" and the double /e/ constitutes now a single phone labeled "e_long".

Table 2: Summary of the annotation of the labels for each ASR systems.

New system	Number of vowels contrasted	Mark
Hausa ASR	2	_closed / _open / _unk
Wolof ASR	7	_short / _long

Finally, for each Hausa and Wolof dictionaries, we have several modeling of a same phone instead of one. In other words, we now have a set of 15 vowels (instead of 8 vowels initially) in our Wolof pronunciation dictionary (i.e.: /a_short/, /a_long/, /e_short/, /e_long/, /ɛ_short/, /ɛ_long/, /i_short/, /i_long/, /o_short/, /o_long/, /ɔ_short/, /ɔ_long/, /u_short/, /u_long/, plus the central vowel /ə/ which has no long counterpart). In the same way, we have 11 vowels (instead of 5 vowels at the start) in our Hausa pronunciation dictionary (i.e.: /a/, /i/, /u/ and /e_closed, /e_open, /e_unk/, /o_closed, /o_open, /o_unk/).

Using these new pronunciation dictionaries, we trained for each language a new system with the Kaldi toolkit. We used the same protocol than in 3.2. We trained a triphone model using 2,969 context-dependent states for Hausa, 3,393 context-dependent states for Wolof and 40K Gaussians. The performance of these new systems is shown in the table 3.

Table 3: Results according to the new acoustic models - with modelisation of vowel length.

Language	CD-HMM/GMM		CD-DNN-HMM	
	WER (%)	CER (%)	WER (%)	CER (%)
Hausa	12.9	3.7	7.9	2.1
Wolof	31.9	12.9	27.7	10.5

We can see on table 3 that we slightly improved the acoustic models for the Hausa ASR system (respectively, -0.1% and -0.3% of WER). We do not notice any improvement of the character error rate (CER). About the Wolof ASR system, we observe a slight degradation of the performance at this stage. While similar in performance, we believe that both ASR systems (duration modeling or not) are complementary and the remaining of the section is dedicated to their combination. Nonetheless, even if the labelisation of the vowels does not increase the performance of the ASR system, this modelisation puts an extra information for the linguists who want to annotate their corpus (in a speech analysis scenario).

Table 4: Performance of the Hausa and Wolof ASR systems combining system w/o vowel length labelling and system with vowel length labelling. Column "Gain" shows the absolute word error rate reduction compared to the best WER of both systems.

Language	CD-DNN-HMM		
	WER (%)	CER (%)	*Gain (%)*
Hausa	7.8	2.1	*0.1*
Wolof	26.3	10.5	*0.9*

[2] Actually, we also have a few vowels labelled /e_unk/ and /o_unk/ since we still have some /e/ and /o/ non labeled, due to the large variability of the vowels phonetic realization depending on their position into the word (ie. Section 4)

For Hausa and Wolof, we combined both DNN-based acoustic models generated from systems handling or not vowel duration. The table 4 shows that, compared to the best word error rate, we gain 0.1% for the Hausa ASR system and 0.9% for the Wolof ASR.

6. Conclusion

This paper shows the results obtained when we modeled or not the phone length contrast existing for Hausa and Wolof languages. Combining these systems slightly improves the performance of both ASR systems.

In the future, concerning the tuning of our ASR systems for these both languages, we will experiment to consider sharing the same root phone for the longer vowels and their shorter counterparts. In addition to work on better duration models, we intend to model tones for the Hausa ASR system. For the Wolof, we want to deal with text normalization and data augmentation techniques in order to improve ASR performance.
The resources in Wolof (the transcribed speech corpora, pronunciation lexicons and language model) will be soon available on our GitHub repository[3] as well as the Kaldi recipe to reproduce ASR experiments. For Hausa, users will need to buy the corpus and the lexicon at ELDA first.

Acknowledgements

This work was realized in the framework of the French ANR Blanc project ALFFA (ANR-13-BS02-0009).

References

1. Tamgno, J.K., Barnard, E., Lishou, C., Richomme, M.. Wolof speech recognition model of digits and limited-vocabulary based on hmm and toolkit. In: *Computer Modelling and Simulation (UKSim), 2012 UKSim 14th International Conference on*. 2012, p. 389–395. doi:`10.1109/UKSim.2012.118`.
2. Leclerc, J.. Sénégal. `http://www.axl.cefan.ulaval.ca/afrique/senegal.htm`; ???? Visited on 2016-01-26.
3. Newman, P.. The hausa language. *An Encyclopedic Reference Grammar (New Haven & London)* 2000;.
4. Besacier, L., Le, V.B., Castelli, E., Sethserey, S., Protin, L.. Reconnaissance automatique de la parole pour des langues peu dotées: Application au vietnamien et au khmer. *TALN 2005* 2005;.
5. Nimaan, A., Nocera, P., Torres-Moreno, J.M.. Boites a outils tal pour les langues peu informatisees: Le cas du somali. *JADT'06: actes des 8es Journées internationales d'analyse statistique des données textuelles: Besançon, 19-21 avril 2006* 2006;(3):697.
6. Pellegrini, T.. *Transcription automatique de langues peu dotées*. Ph.D. thesis; Université Paris Sud-Paris XI; 2008.
7. Barnard, E., Davel, M.H., Van Huyssteen, G.B.. Speech technology for information access: a south african case study. In: *AAAI Spring Symposium: Artificial Intelligence for Development*. 2010, .
8. Vu, N.T., Kraus, F., Schultz, T.. Rapid building of an asr system for under-resourced languages based on multilingual unsupervised training. In: *Interspeech*. Citeseer; 2011, p. 3145–3148.
9. Besacier, L., Barnard, E., Karpov, A., Schultz, T.. Automatic speech recognition for under-resourced languages: A survey. *Speech Communication* 2014;**56**:85–100. URL: `http://dx.doi.org/10.1016/j.specom.2013.07.008`. doi:`10.1016/j.specom.2013.07.008`.
10. Gelas, H., Besacier, L., Rossato, S., Pellegrino, F.. Using automatic speech recognition for phonological purposes: study of vowel length in punu (bantu b40) 2010;.
11. Kempton, T., Moore, R.K.. Finding allophones: An evaluation on consonants in the timit corpus. In: *Tenth Annual Conference of the International Speech Communication Association*. 2009, .
12. Gadde, V.R.. Modeling word duration for better speech recognition. In: *Proceedings of NIST Speech Transcription Workshop*. 2000, .
13. Povey, D.. Phone duration modeling for lvcsr. In: *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*; vol. 1. IEEE; 2004, p. I–829.
14. Pylkkönen, J., Kurimo, M.. Duration modeling techniques for continuous speech recognition. In: *INTERSPEECH*. 2004, .
15. Alumäe, T., Nemoto, R.. Phone duration modeling using clustering of rich contexts. In: *INTERSPEECH*. Citeseer; 2013, p. 1801–1805.
16. Schlippe, T., Djomgang, E.G.K., Vu, N.T., Ochs, S., Schultz, T.. Hausa large vocabulary continuous speech recognition. In: *SLTU*. 2012, p. 11–14.
17. Nouguier Voisin, S.. *Relations entre fonctions syntaxiques et fonctions sémantiques en wolof*. Ph.D. thesis; Lyon 2; 2002.
18. Attardi, G., Fuschetto, A.. Wikipedia extractor. *Medialab, University of Pisa* 2013;.
19. Gauthier, E., Besacier, L., Voisin, S., Melese, M., Elingui, U.P.. Collecting resources in sub-saharan african languages for automatic speech recognition: a case study of wolof. *LREC* 2016;Accepted.

[3] see `https://github.com/besacier/ALFFA_PUBLIC`

20. Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., et al. The kaldi speech recognition toolkit 2011;.
21. Kingsbury, B.. Lattice-based optimization of sequence classification criteria for neural network acoustic modeling. In: *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*. 2009, p. 3761–3764.
22. Newman, R.M., Heuven, V.J.v.. An acoustic and phonological study of pre-pausal vowel length in hausa. *Journal of African Languages and Linguistics* 1981;**3**(1):1–18.
23. Jaggar, P.J.. *Hausa*; vol. 7. John Benjamins Publishing; 2001.