Automatic Speech Recognition For African Languages With Vowel Length Contrast

Elodie Gauthier*, Laurent Besacier*, Sylvie Voisin**
*LIG, CNRS and Univ. Grenoble Alpes (France)
**DDL, CNRS and Univ. Lyon 2 (France)
African Languages and Information Technologies

Address under-resourced languages from Africa

- Focus on West Africa
  - Hausa, Wolof, Fulfulde, Zarma, Bambara
- 2 East African languages
  - Amharic, Swahili

Data collection methodology

- ASR (speech-to-text) and TTS (text-to-speech)
- French Partners: LIG (Grenoble), LIA (Avignon), DDL (Lyon), Voxygen (Lannion), [http://alffa.imag.fr](http://alffa.imag.fr)
This paper focuses on ASR

Four languages covered so far in ASR

- Hausa, Wolof, Amharic, Swahili

<table>
<thead>
<tr>
<th>Task</th>
<th>WER %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swahili Broadcast News</td>
<td>20.7</td>
</tr>
<tr>
<td>Hausa Read Speech</td>
<td>10.0</td>
</tr>
<tr>
<td>Amharic Read Speech</td>
<td>8.7</td>
</tr>
<tr>
<td>Wolof Read Speech (under dev.)</td>
<td>28.6</td>
</tr>
</tbody>
</table>

Data and Kaldi [1] scripts released on Github

https://github.com/besacier/ALFFA_PUBLIC

Vowel length contrast in Hausa and Wolof (1/3)

- 2 African languages largely spoken in the West of the continent
- Vehicular languages also spoken as a 2d or 3d language
- Both have length contrast at the phone level
  - Contrast of vowels (Hausa) and contrast of vowels + consonants (Wolof)
  - Contrast changes the meaning of a word
  - Literature mention contrast for 5 vowels in Hausa and 7 vowels in Wolof

Length mark in text?

- Hausa: **no** (vowel length depends on its position into the syllable)
  - **bos** (bus) => /o/ expected short
  - **aboki** (friend) => /o/ expected long
- Wolof: **yes** (duplication of the graphemes)
  - **kaar** (bus) => /a/ expected long
  - **xarit** (friend) => /a/ expected short
Vowel length contrast in Hausa and Wolof (2/3)

📖 Related works in ASR duration models

- [1] rescoring N-best lists with duration models
- [2] modeled duration at word and phone levels + lattice rescoring
  - WER reduction limited
  - Moderate WER reduction

**Comment:** these approaches do not challenge the duration contrast phenomenon while for under-documented languages, it might be important to verify its empirical realization

Taking into account length contrast in Hausa and Wolof

- Train ASR systems with and w/o vowel length modeling
  - Different phone inventories and lexicons
- Combine these systems (instead of re-scoring framework)
- Vowel length modeling in Wolof
  - Contrastive phone units for short and long vowels (a gives /a_short/ and aa gives /a_long/)
- Vowel length modeling in Hausa
  - Contrastive phone units for phones if they are in a closed or open syllable (o gives /o_closed/ or /o_open/ depending on its position in the syllable)
  - Sometimes, not decidable: /o_unk/ label
Data sets

- Hausa Globalphone Corpus [1]
- Wolof in-house Corpus [2]

Wolof

- 21h of read speech recorded in Dakar (Senegal), 18 speakers (10+8)
- In-house docs + text data crawled from the Web (but very few!!)

<table>
<thead>
<tr>
<th>Task</th>
<th>LM corpus (#words)</th>
<th>AM training (h)</th>
<th>Eval. Set (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hausa read speech</td>
<td>8M</td>
<td>7h</td>
<td>1h</td>
</tr>
<tr>
<td>Wolof read speech</td>
<td>0,6M</td>
<td>17h</td>
<td>2h</td>
</tr>
</tbody>
</table>

ASR Systems

- Kaldi CD-GMM-HMM and CD-DNN-HMM systems
- 2.9k CD states for Hausa, 3.4k CD states for Wolof
- Kaldi scripts made available
  - https://github.com/besacier/ALFFA_PUBLIC/tree/master/ASR/HAUSA
  - https://github.com/besacier/ALFFA_PUBLIC/tree/master/ASR/WOLOF

Baseline results (no length modeling)

<table>
<thead>
<tr>
<th>Task</th>
<th>WER (GMM-HMM)</th>
<th>WER (DNN-HMM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hausa read speech</td>
<td>13.0 %</td>
<td>8.0 %</td>
</tr>
<tr>
<td>Wolof read speech</td>
<td>31.7 %</td>
<td>27.2 %</td>
</tr>
</tbody>
</table>
Empirical verification of vowel length contrast in Hausa

- 5,863 training sentences forced-aligned
- Measured Hausa vowel length in closed vs open syllabic context
- Differences in vowel length only observable for /e/ and /o/

For /a/, /i/ and /u/ contrast was not empirically observed
Hausa Results

- Only /e/ and /o/ contrasted in the lexicon with _closed/_open/_unk labels
- Slight gain, but not significant
- Modeling contrast for all vowels degrades the results (not reported here)

<table>
<thead>
<tr>
<th>Method</th>
<th>Hausa WER (DNN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No length modeling</td>
<td>8.0 %</td>
</tr>
<tr>
<td>Length modeling</td>
<td>7.9 %</td>
</tr>
<tr>
<td>Combination</td>
<td>7.8 %</td>
</tr>
</tbody>
</table>
Wolof Results

- All 7 vowels contrasted in the lexicon with \_short_/\_long
- Slight gain with combination only
- Did not verify the length contrast for Wolof at the time of paper submission (see next slide)

<table>
<thead>
<tr>
<th>Method</th>
<th>Hausa WER (DNN)</th>
<th>Wolof WER (DNN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No length modeling</td>
<td>8.0 %</td>
<td>27.2 %</td>
</tr>
<tr>
<td>Length modeling</td>
<td>7.9 %</td>
<td>27.7 %</td>
</tr>
<tr>
<td>Combination</td>
<td>7.8 %</td>
<td>26.3 %</td>
</tr>
</tbody>
</table>
Updated results on Wolof

- Only 5 vowels for which contrast was empirically verified
- Wolof eval. set cleaned since then (from 2,000 to 1,120 utterances)

<table>
<thead>
<tr>
<th>Method</th>
<th>Hausa WER (DNN)</th>
<th>Wolof WER (DNN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No length modeling</td>
<td>8.0 %</td>
<td>20.5 %</td>
</tr>
<tr>
<td>Length modeling</td>
<td>7.9 %</td>
<td>20.0 %</td>
</tr>
<tr>
<td>Combination</td>
<td>7.8 %</td>
<td>19.1 %</td>
</tr>
</tbody>
</table>
Conclusion

**First LVCSR system developed for Wolof language**
- Train ASR systems with and w/o vowel length modeling
- Complementarity of systems but small performance gains
- Acoustic models with length contrast useful for phonetic studies or TTS
  - Length contrast phenomenon not empirically observed on all vowels, even on read speech
  - What about spontaneous speech?

**ALFFA project outcomes**
- Kaldi ASR systems and resources online
- Started to work on Pulaar Fulfulde