USING AUTOMATIC SPEECH RECOGNITION FOR PHONOLOGICAL PURPOSES: STUDY OF VOWEL LENGTH IN PUNU (BANTU B40)

Hadrien GELAS[1] , Laurent BESACIER[2], Solange ROSSATO[2] & François PELLEGRINO[1]

(1) DDL, CNRS - Université Lumière Lyon 2 – (2) LIG, CNRS – Grenoble Universités

GOALS OF THE STUDY

➩ Automatic speech technology offers opportunities for investigating a wide range of issues in Laboratory Phonology (e.g. [1, 2]).
➩ But its availability hardly extends beyond a small number of languages, and most spoken languages are thus under-resourced in this perspective.
➩ **Is it possible to overcome this drawback?**
➩ This paper aims at assessing the relevance of a multilingual ASR (Automatic Speech Recognition) system for studying Vowel Length Contrast in Punu, a Bantu language (B40) spoken in Gabon.
➩An additional goal is to examine the phonetic/phonological pattern of length contrast in Punu.

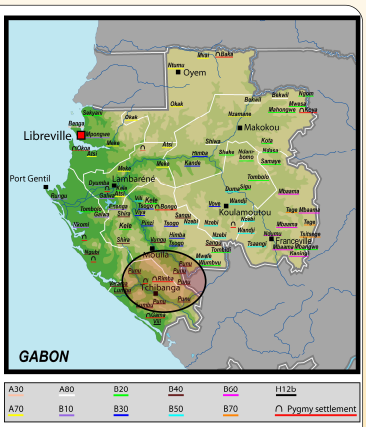
LINGUISTIC SITUATION

✓ The Punu Language

➩ Bantu language, B43 [6]
➩ South-west of Gabon
➩ Around 123 000 speakers [7]

✓Elements of Punu phonology

➩ The vowel system is based on 5 vowel qualities: /i, u, e, o, a/
➩ Vowel length is distinctive, leading to a 10-vowel system: /iː, uː, eː, oː, aː/
➩ The canonical construction of nouns in Punu is based on a CV.CV stem, preceded by a CV prefix.
➩ Length contrast only occurs in stressed position, which is the first syllable of the stem for lexical items in isolation [4].



Map 1. Languages in Gabon

➩ Tone system
 ⇘ Surface tones : level tones (Low, High)
 + register tones (Rising, Falling, Super-high-falling)
 ⇘ Noun tone patterns: LL, LH, HL, HH

CORPUS

➩ Audio recordings of nouns in isolation from one native speaker.
➩ 710 items along with their phonological transcriptions.
➩ Total number of 1,999 vowels (including 710 vowels in stressed position).
➩ A **sub-corpus** of 50 items (with 25 long and 25 short vowels in stressed position, each vowel quality equally represented) was manually aligned with their phonetic transcription and used as a reference for evaluating the automatic alignment.

METHODS

➩ Multilingual phonetic decoder

⇘ An existing multilingual acoustic-phonetic decoder [5] built upon the CMU Sphinx toolkit was used.
⇘ Phone models trained on English (EN), Mandarin Chinese (CH), Malaysian (MA), and Vietnamese (VI) data.
⇘ Duration modeling is implicit (short and long vowels refer to the same models) and tones are not considered.

➩ Phonetic matching between Punu & Multilingual models

⇘ For Punu phonemes *present* in the multilingual decoder, the matching was straight-forward.
⇘ For Punu phonemes *absent* from the decoder, a manual selection was done to pick up the most suitable phone (or phone sequence) from the multilingual model.

[β]➩V_EN	[t͡s] ➩ ts_CN	[m͡b] ➩ M_EN B_EN	[m͡v] ➩ M_EN V_EN
[ɓ] ➩ B_EN	[d͡ʒ] ➩ JH_EN	[n͡d] ➩ N_EN D_EN	[n͡z] ➩ N_EN Z_EN
[ɗ] ➩ D_EN		[ŋ͡g] ➩ NG_EN G_EN	

REFERENCES

[1] Ghosh, Prasanta K., and Narayanan, S.S., (2009), "Closure duration analysis of incomplete stop consonants due to stop-stop interaction", *J. Acoust. Soc. Am.* 126.
[2] Gendrot, C. & Adda, M., (2006), "Is there a universal impact of duration on formant frequency values of oral vowels? An automated analysis of speech from eight languages", *Laboratory Phonology X*, Paris, June 2006.
[4] Nsuka Nkutsi, F., (Ed), (1980). *Éléments de description du punu*, Lyon : CRLS Université Lumière Lyon 2, 217 p.
[5] Quenot, G., Tan, T.-P., Le, V.-B., Ayache, S., Besacier, L. & Mulhem, P. (2009), "*Content-based search in multilingual audiovisual documents using the International Phonetic Alphabet*", Multimedia Tools and Applications Journal, Springer US.
[6] Guthrie, M. (1967). *Comparative Bantu : an introduction to the comparative linguistics and prehistory of the Bantu languages.* Vol 1-4. Farnborough : Gregg press.
[7] Gordon, R. G. Jr. (2005). *Ethnologue: Languages of the World, Fifteenth edition*. Dallas : SIL International. Retrieved from http://www.ethnologue.com in 2009).

EXPERIMENTAL STUDY

➩ Using each word transcription, an automatic alignment was performed on the whole corpus, resulting in automatic estimations of the vowel durations. Evaluation consisted in:
 ① Comparing automatic estimations with the hand-made reference values for the sub-corpus (50 vowels).
 ② Studying the vowel length distribution to the whole corpus (710 vowels in stress position, 1,999 vowels in all positions).
➩ Considering the tiny size of the reference sub-corpus (i.e. 5 items per vowel quality and vowel length), no statistical procedure was performed.

RESULTS

① Comparison between reference and automatic vowel duration estimation

➩ 50 word subset
➩ Results
 ⇘ High consistence between distributions yielded from manual and automatic processes
 ⇘ Clear-cut dichotomy between short and long vowels (caveat: isolated words)
 ⇘ Some gross errors (see below)
 ⇘ Dependency to vowel quality

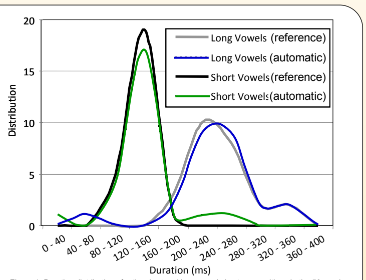
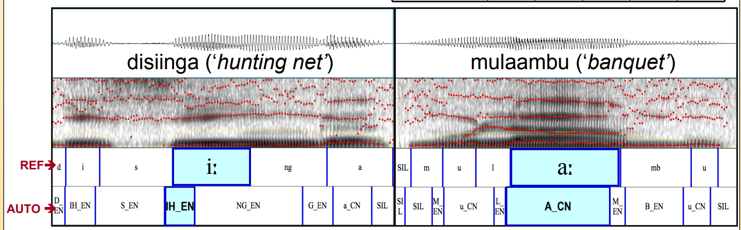


Figure 1. Duration distributions for the short and long vowels in stress positions in the 50-word sub-corpus. Manual reference durations (black and grey lines) and automatic estimated durations (green and blue lines) are displayed.

Vowel quality	a	e	i	o	u
Correlation (r^2)	0.97	0.80	0.22	0.74	0.97



② Length Contrast in Punu

➩ Automatic estimation applied to the whole corpus (1,999 vowels) provide a more general outline of the phonetic/phonological pattern of length contrast in Punu
 ⇘ Clear-cut dichotomy between short and long vowels
 ⇘ Clear distinction between vowels in non-stressed position and vowels in stressed position
 ⇘ Probable vowel reduction in final position

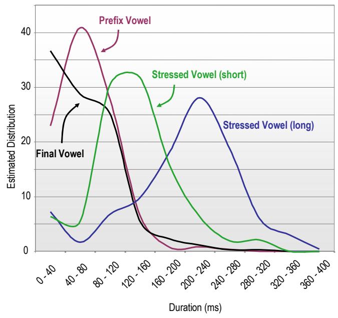


Figure 2. Distributions of estimated durations for all vowels as a function of their position - (prefix position: 580 vowels; stressed position: 710 vowels; final position: 709 vowels).

CONCLUSION

➩ Discussion

⇘ This paper assesses the relevance of an automatic alignment for studying durational patterns, even in a language previously unknown from the ASR system.
⇘ Results show that the expected bimodal distribution of durations is correctly detected, even with a large mismatch between the acoustic-phonetic models (based on European and Asiatic languages) and the target language.
⇘ Though the distribution estimation is correct, gross errors are present

➩ Further improvements

⇘ Filtering out the obvious alignment errors (often occurring in glide-vowel contexts) to provide more accurate duration estimations
⇘ Determining the best phonetic model for each Punu phoneme
⇘ Taking into account language-specific phonotactics (more data needed)
⇘ Adapting the ASR system with data from more closely-related languages (for a better phonetic system coverage and adaptation)

Acknowledgement: This research was done in the framework of the project ANR PI (Spoken Language Technologies for PI-languages, PI: Laurent Besacier)