

# Using automatic speech recognition for phonological purposes: Study of Vowel Length in Punu (Bantu B40)

## 1. INTRODUCTION

Automatic speech technology offers great opportunities for investigating a wide range of issues in Laboratory Phonology (e.g. [1, 2]). However, its availability hardly extends beyond a small number of languages, and most spoken languages are thus under-resourced in this perspective. Nevertheless, an increasing effort is done to develop automatic speech recognition (ASR) for such languages<sup>1</sup> and some phonological issues may be tackled by now.

The primary goal of this paper is to assess the relevance of a multilingual ASR system for studying Vowel Length Contrast in Punu, a Bantu language (B40) spoken in Gabon. The additional goal is thus to examine the phonetic/phonological pattern of length contrast in Punu.

## 2. ELEMENTS OF PUNU PHONOLOGY

The vocalic system of Punu is made up of five different vowel qualities: /i, u, e, o, a/. Vowel length is phonologically distinctive and the corresponding long vowels are thus also considered: /i:, u:, e:, o:, a:/ [3]. The canonical construction of nouns in Punu is based on a CV.CV stem, preceded by a CV prefix, and length contrast only occurs in stress syllable, which is the first syllable of the stem for lexical items in isolation [4].

## 3. EXPERIMENTAL STUDY

3.1 Corpus Description. Audio recordings of nouns in isolation from one native speaker were used. The corpus consisted of 710 items along with their phonological transcriptions, for a total number of 1999 vowels (including 710 vowels in stress positions). A reference sub-corpus of 50 items (25 with a long vowel and 25 with a short vowel, each vowel quality equally represented) has also been manually aligned to provide reference values for vowel durations.

3.2 Automatic Alignment. An already available multilingual acoustic-phonetic decoder was used [5]. It was originally designed for English, Mandarin Chinese, Malaysian, and Vietnamese, and an a priori matching between Punu phonemes and similar phones of those languages has been made. Using each word transcription, an automatic alignment was performed on the whole corpus, resulting in automatic estimations of the vowel durations.

## 4. RESULTS

The accuracy of the automatic procedure is evaluated on the sub-corpus by comparing the estimated durations resulting from the automatic alignment with their reference values. The actual durations (Fig. 1 – solid lines) confirm the bimodal distribution of the vowel length in isolated words and automatic estimations (Fig. 1 – dashed lines) demonstrated a very good accuracy.

Considering the whole corpus, a similar pattern discriminates between short and long vowels in stress position (Fig. 2 – solid lines). Additionally, Fig. 2 provides the duration distributions for the 1289 vowels of the corpus in a non-stressed position (either prefix vowel or final vowel, in dashed lines), resulting in a more general outline of the phonetic/phonological pattern of length contrast in Punu.

## 5. 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. However, this step towards Laboratory Phonology needs further improvements. First, duration estimations will definitely be improved by filtering out the obvious alignment errors (often occurring in glide-vowel contexts). Besides, the ASR system will be soon adapted with a significant amount of data from several Bantu languages, with the expected result of improving the multilingual models for some problematic sounds (as the velar fricative [ɣ] for instance).

---

<sup>1</sup> See for instance the recent International Workshops on Spoken Languages Technologies for Under-resourced languages (SLTU): <http://www.mica.edu.vn/sltu-2010/>, <http://www.mica.edu.vn/sltu/>

## 6. 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.
- [3] Kwenzi-Mikala, J.-T., (1990), "Yipunu", *Revue Gabonaise des Sciences de l'Homme*, 2, pp. 299-306.
- [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.

## 7. FIGURES

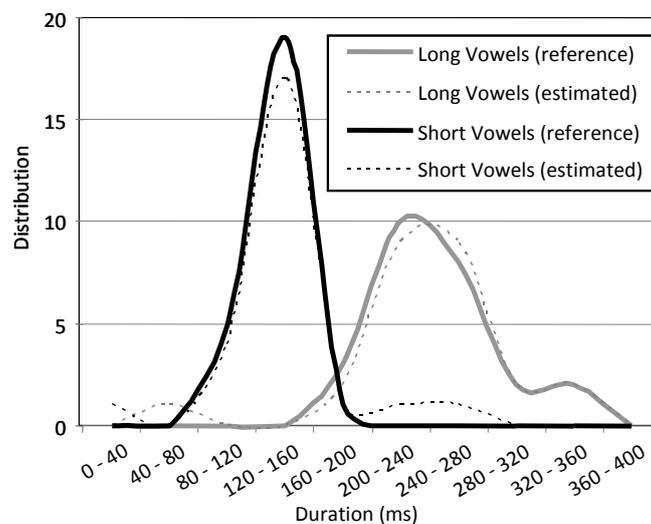


Figure 1. Duration distributions for the short and long vowels in stress positions in the 50-word sub-corpus. Manual reference durations (solid lines) and automatic estimated durations (dashed lines) are displayed.

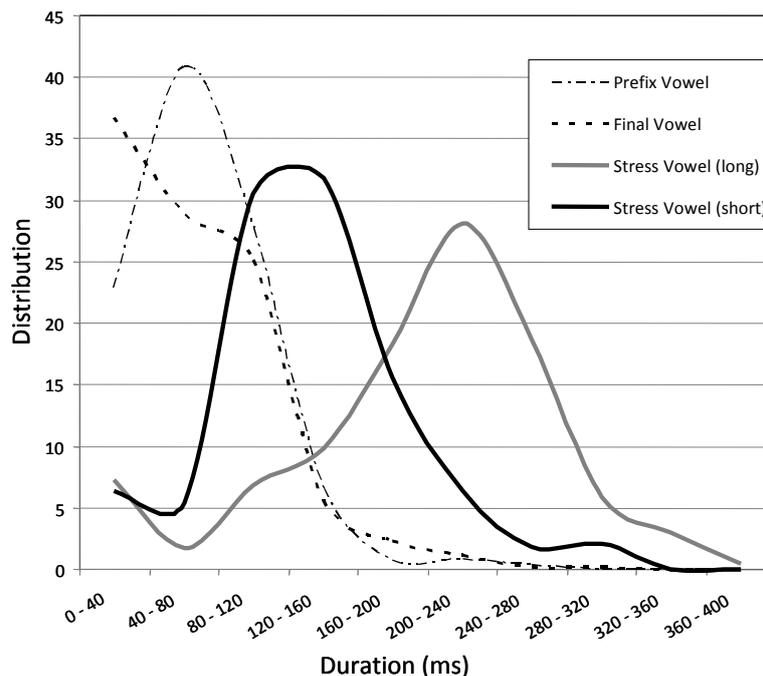


Figure 2. Distributions of estimated durations for all vowels as a function of their position (prefix position: 580 vowels; stress position: 710 vowels; final position: 709 vowels).