

# Machine Assisted Analysis of Vowel Length Contrasts in Wolof

Elodie Gauthier, Laurent Besacier, Sylvie Voisin

► **To cite this version:**

Elodie Gauthier, Laurent Besacier, Sylvie Voisin. Machine Assisted Analysis of Vowel Length Contrasts in Wolof. Interspeech 2017, Aug 2017, Stockholm, Sweden. Interspeech 2017, pp.2138 - 2142, 2017, <10.21437/Interspeech.2017-268>. <hal-01580902>

**HAL Id: hal-01580902**

**<https://hal.archives-ouvertes.fr/hal-01580902>**

Submitted on 3 Sep 2017

**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.

# Machine Assisted Analysis of Vowel Length Contrasts in Wolof

Elodie Gauthier<sup>1</sup>, Laurent Besacier<sup>1</sup>, Sylvie Voisin<sup>2</sup>

<sup>1</sup>Laboratoire d’Informatique de Grenoble (LIG), Univ. Grenoble Alpes, Grenoble, France

<sup>2</sup>Laboratoire Dynamique Du Langage (DDL), CNRS - Université Aix Marseille, France

elodie.gauthier@imag.fr, laurent.besacier@imag.fr, sylvie.voisin@cnrs.fr

## Abstract

Growing digital archives and improving algorithms for automatic analysis of text and speech create new research opportunities for fundamental research in phonetics. Such empirical approaches allow statistical evaluation of a much larger set of hypothesis about phonetic variation and its conditioning factors (among them geographical / dialectal variants). This paper illustrates this vision and proposes to challenge automatic methods for the analysis of a not easily observable phenomenon: vowel length contrast. We focus on Wolof, an under-resourced language from Sub-Saharan Africa. In particular, we propose multiple features to make a fine evaluation of the degree of length contrast under different factors such as: read vs semi-spontaneous speech ; standard vs dialectal Wolof. Our measures made fully automatically on more than 20k vowel tokens show that our proposed features can highlight different degrees of contrast for each vowel considered. We notably show that contrast is weaker in semi-spontaneous speech and in a non standard semi-spontaneous dialect.

**Index Terms:** computational phonetics, vowel length contrast, automatic speech recognition, wolof language, under-resourced languages

## 1. Introduction

Growing digital archives and improving algorithms for automatic analysis of text and speech create new research opportunities for fundamental research in linguistics and phonetics. This vision is shared by [1] where audiobooks (large amount of recordings in many languages and dialects, distributed in a natural way across a wide variety of speakers) are used for corpus-based phonetics. In their work, authors claim that - for the phonetic events observed - “the data used from audiobooks offers more tokens than have been examined in the entire 50-year history of sociolinguistic study of Spanish”. In a similar trend, we have recently shown the value of stochastic and neural acoustic models for analyzing, at a relatively large scale, vowel length contrast in two under-resourced african languages [2]. Such empirical approaches allow statistical evaluation of a much larger set of hypothesis about phonetic variation and its conditioning factors (among them geographical / dialectal variants). This paper illustrates this vision and proposes a detailed analysis of vowel length contrast in Wolof under different factors such as: read vs semi-spontaneous speech ; standard (Dakar) Wolof vs dialectal (Faana-Faana) Wolof.

**Paper contributions.** The first contribution of this paper is a large scale analysis of vowel length contrast on Wolof read speech. Multiple features are proposed to judge the degree of bimodality in the distribution (of durations) for a given vowel. Our measures made on 14k vowel tokens show different degrees of contrast according to the vowel considered. We also show, in a second contribution, that in the case of read speech, the need of manual transcriptions can be relaxed since the use

of automatic speech recognition (ASR) can lead to very similar measurements and to the same conclusions. Our third contribution is an application of our machine-assisted methodology to study vowel length contrast in more spontaneous speech for Wolof and for one of its dialectal variant (Faana-Faana). For reproducible research, a Wolof ASR VM and the data of this study are also made available online<sup>1</sup>.

**Languages studied.** Wolof is the vehicular language of Senegambia (Senegal and Gambia), also spoken in Mauritania. This paper focuses on senegalese Wolof. We will use the term “standard” to refer to Wolof spoken in Dakar by native speakers of the language and “urban” for Wolof spoken by non-native speakers. In Senegal, there are also dialectal variants but mutual understanding exists between people living in the different areas. Linguists observe some phonetic or morpho-phonological variations, focusing on vocalism, on some forms of verbal inflection [3] and also on some morphological and syntactical variations [4], [5].

The Faana-Faana dialect studied in this paper is spoken in the region of Kaolack, also named *Wolof of the Saloum*. It is described by Dramé [6] and is closer to the Wolof of Gambia. This regional variant is not much influenced by other Wolof dialects. However, young people and men often spend part of their lives in Dakar and come back with influences from standard Wolof. Faana-Faana speakers live in a predominant Sereer speaking area which influences their own language, but they are not subject to other major linguistic influences.

In Wolof, the vocalic system is composed of 8 short vowels /i/, /e/, /ɛ/, /a/, /ə/, /ɔ/, /o/, /u/; each having a long counterpart (except /ə/). There is no tone in Wolof but phonemes can vary in length [7]. This means that word sense may differ depending on phoneme duration. For instance, the pronunciation of “fit” (*bravery*) and “fiit” (*trap*) varies only at the vowel length level, as well as “wall” (*to rescue*) and “waal” (*to take advantage of*), or “set” (*to be clean*) and “seet” (*to look for*). Same short and long vowels exists in the Faana-Faana variant. As can be seen in the examples above, reduplication of the vowel, in the spelling of Wolof, encodes the duration. One goal of this paper is to verify if this expected (phonological) contrast is also observed at the phonetic level.

**Paper outline.** This paper is organized as following. Section 2 reviews previous works on phonemic contrast analysis. In Section 3, we propose several features to measure degree of (length) contrast for a given unit. In Section 4, we present our multi factor analysis of vowel length contrast in Wolof read and semi-spontaneous speech (Dakar and Faana-Faana). Finally, Section 5 concludes this work and gives some perspectives.

<sup>1</sup>see [https://github.com/besacier/ALFFA\\_PUBLIC/blob/master/ASR/WOLOF/WOLOF-VM/](https://github.com/besacier/ALFFA_PUBLIC/blob/master/ASR/WOLOF/WOLOF-VM/) and [https://github.com/besacier/ALFFA\\_PUBLIC/blob/master/ASR/WOLOF/INTERSPEECH\\_2017](https://github.com/besacier/ALFFA_PUBLIC/blob/master/ASR/WOLOF/INTERSPEECH_2017)

## 2. Related Works

Vowel duration is a phonetic measure widely used in speech acoustic research. Many factors affect vowel duration such as its location within the vowel space ([8], [9]), position and length of the word [10], surrounding context of the vowel ([11], [12]), speech rate ([13], [14]) and position of the vowel within the word [15]. As raised by [16], main past studies of vowel duration were done through manual annotations. It is consequently a very time-consuming task and only few words were generally analyzed. We believe that use of automatic tools can lead to more objective and reproducible measures, at a larger scale.

As far as vowel length contrast is concerned, [17] studied its production and perception in Korean. They found that all Korean speakers of the study produced (length) contrasted vowels but they also concluded that short/long contrast is weaker in spontaneous speech. Vowel length contrast was also investigated to better understand language acquisition. [18] analyzed 11 hours of Japanese infant-directed speech, using statistical methods, to explore how infants learn to discriminate vowel length contrast existing in Japanese. They discovered that duration distribution for a given vowel is not clearly bi-modal since long vowels may be much less frequent than short vowels.

In Wolof, very few phonetic studies were published, especially on vowel length contrast. One exception is the work of [19] who studied a dialectal variant of Gambian Wolof, close to Faana-Faana analyzed in this paper. The author compared 3 minimal pairs, each containing /i/, /a/ and /u/ vowels (read speech) and noticed that length contrast was more important for vowel /a/ than for /i/ and /u/. Moreover, less (length) contrast was observed in rapid speech rate compared to normal speech rate. Finally, in 2006, [7] pointed out that a large analysis of Wolof phonetics was lacking and to the best of our knowledge this is still the case at present.

## 3. Measuring Vowel Length Contrast

It is not trivial to objectively analyze the degree of bimodality in the distribution of durations for a given vowel. One reason is that - for some vowels - there may be much more short occurrences than long ones [20]. Eye-looking at distributions is a possibility but more objective features are needed if we want a fine evaluation of the degree of contrast across different speech styles and dialects (see [18] for Japanese). This section proposes different criteria (features) to estimate the degree of bimodality for the (duration) distribution of a given vowel. These features are not extracted from true distributions of short and long vowels, but from their normalized gamma approximations<sup>2</sup> - see Figure 1 for the notations used: (1) ratio  $r_1$ , (2) ratio  $r_2$ , (3) area  $\mathcal{A}$  between both (short/long) gamma distributions and (4) delta  $\Delta$  between modes of both gamma distributions.

We define  $d_S(x)$  and  $d_L(x)$  as representing respectively the distribution of the short and long units of a vowel (for instance  $d_{/i/}(x)$  and  $d_{/ii/}(x)$ ). In accordance with this definition,  $r_1$  is defined by equation (1) and is the ratio between  $d_S(a)$  and  $d_L(a)$ , when  $a$  is the global maximum value of  $d_S(x)$ . A high value of  $r_1$  means a large amount of short tokens compared to long tokens at the maximum peak of  $d_S(x)$ . In the same way,  $r_2$  defined in equation (2) is the ratio between  $d_L(b)$  and  $d_S(b)$ , when  $b$  is the global maximum value of  $d_L(x)$ . A high value of  $r_2$  means a large amount of long tokens compared to short tokens at the maximum peak of  $d_L(x)$ . For both ratios, the bigger the value, the stronger the duration contrast is.

<sup>2</sup>We preferred *Gamma* distributions to *Gaussian* for their skewness.

$$r_1 = \frac{d_S(a)}{d_L(a)} \quad (1)$$

where  $a = \arg \max_x(d_S(x))$ .

$$r_2 = \frac{d_L(b)}{d_S(b)} \quad (2)$$

where  $b = \arg \max_x(d_L(x))$ .

$\mathcal{A}$  corresponds to the computed area between both curves when  $d_S(x) < d_L(x)$ , as shown in equation (3). The larger the area, the stronger the duration contrast should be. We consider that a significant contrast should give an area  $\mathcal{A} > 0.40$ .

$$\mathcal{A} = \int_I^{\infty} d_L(x) - d_S(x) dx \quad (3)$$

We also compute  $\Delta$  which is the difference between both modes of  $d_S(x)$  and  $d_L(x)$ , as represented in equation (4). The greater the value of  $\Delta$ , the more significant the contrast is. Figure 1 displays duration histograms, associated gamma curves and notations, for phoneme /a/.

$$\Delta = \arg \max_x(d_L(x)) - \arg \max_x(d_S(x)) \quad (4)$$

Finally, it is important to note that we did not use Hartigan's Dip test of unimodality [21] since our preliminary measurements have shown that this test always concludes to the bimodality of our distribution - even for extremely weak contrasts.

## 4. Machine Assisted Analysis of Vowel Length Contrasts in Wolof

### 4.1. Data and ASR System

In addition to our existing in-house (Dakar standard) Wolof read speech corpus [22], we recently collected data during a field trip in Senegal. We collected semi-spontaneous speech of Wolof (Dakar standard) and dialectal variants. In total, we gathered around 1.5 hours of elicited speech from 22 speakers (6 Faana-Faana speakers, 2 Lebu speakers, 3 speakers of urban Wolof and 11 speakers of standard Wolof). Each speaker had to watch a series of 76 short videos designed to express trajectory [23]. This data can be considered as semi-spontaneous speech.

Our best Wolof ASR system was used to decode new recorded speech. This is a standard context dependent DNN-HMM hybrid system trained with Kaldi speech recognition toolkit [24]. More details on this system can be found in [2] and it is made available through a VM<sup>3</sup>. We used 5 transcriptions of Faana-Faana (over 6) and 3 transcriptions of standard Wolof (over 11), because only a subset of ASR hypotheses were corrected by Wolof linguists. Table 1 summarizes each data set on which we will measure vowel length contrast in this paper.

Table 1: *Wolof speech data overview.*

Data Set	Male	Female	#Utt	#Words	Duration
Wolof (read)	8	6	1,120	10,461	1h12 mins
Wolof (semi-spontaneous)	2	1	254	2,825	14 mins
Faana-Faana (semi-spontaneous)	5	0	454	3,365	19 mins

<sup>3</sup>see [https://github.com/besacier/ALFFA\\_PUBLIC/blob/master/ASR/WOLOF/WOLOF-VM/](https://github.com/besacier/ALFFA_PUBLIC/blob/master/ASR/WOLOF/WOLOF-VM/)

## 4.2. Analysis on Wolof Read Speech

### 4.2.1. Forced Alignment with Human Transcriptions

In a first phase, we extract vowel durations by force-aligning human transcriptions of development (*dev*) set described in [2] (1,120 utterances, 1h12mn of speech) and made up of Wolof read speech (see Table 1). Forced-alignment is done with our CD-DNN-HMM-based acoustic model (length contrasted acoustic models with different units for short and long vowels). The 7 contrasted vowels are tagged as /short/ or /long/ depending on the duplication of the grapheme within the word. Data is partitioned in different sets denoted by  $\mathcal{D}_l^v$  where  $v \in \mathcal{V} = \{i, e, \varepsilon, a, \text{ɔ}, o, u\}$  is the studied vowel and  $l \in \mathcal{L} = \{S, L\}$  is the expected length of the vowel (short or long). We computed vowel durations and built their histogram for each vowel after deleting outliers (we keep observations  $x$  so that  $\mu - 3\sigma < x < \mu + 3\sigma$ ). We also approximate our real distribution by the probability density function of a *Gamma* distribution. Eye-looking at normalized distributions for each vowel confirms that bimodality exists for all of them. However, the degree of contrast differs for each vowel. For instance, strong duration contrast is observed for vowel /a/ (Figure 1) whereas weak contrast is observed for vowel /ɔ/ (Figure 2).

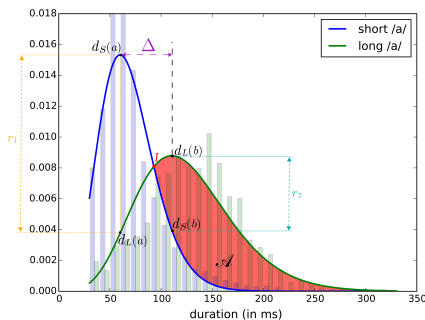


Figure 1: Histogram and Gamma Distribution for vowel /a/ in Wolof Read Speech - Strong Contrast

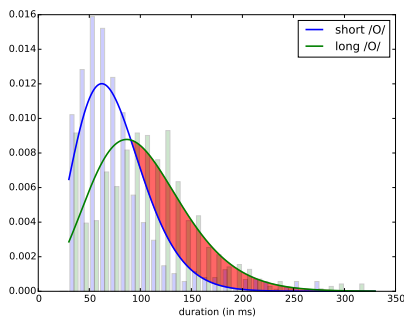


Figure 2: Histogram and Gamma Distribution for vowel /ɔ/ in Wolof Read Speech - Weak Contrast

Table 2 shows measurements of length contrast. Vowels are sorted according to their height. In addition to the contrast features described in Section 3, we also display in third column the mean duration  $\mu$  (in ms) for each short and long vowel. Vowel /a/ is the one that appears most frequently (both short and long) while vowel /o/ is the one that appears most rarely. This is easily explained because words containing the vowel /a/ are very common while those containing vowel /o/ are rare in Wolof. We observe that 2 articulatory features affect vowel duration: height and backness. Indeed, mean duration of short vowels in-

Table 2: Contrast Features Extracted on Wolof Read Speech.

Phoneme	#occurrences	$\mu$ (in ms)	$r_1$	$r_2$	$\mathcal{A}$	$\Delta$ (in ms)
short /i/	2,149	76	2.54	1.42	0.44	49
long /i:/	133	131				
short /e/	227	79	2.63	1.52	0.45	37
long /e:/	178	120				
short /ɛ/	1,264	81	2.64	1.50	0.45	46
long /ɛ:/	557	131				
short /a/	4,673	69	4.07	2.21	0.56	50
long /a:/	880	125				
short /ɔ/	881	73	1.62	0.93	0.27	24
long /ɔ:/	710	102				
short /o/	60	68	2.85	1.27	0.46	34
long /o:/	69	108				
short /u/	1,893	67	2.34	1.09	0.40	36
long /u:/	111	110				

creases with the aperture of the jaw, as described in [19], except for /a/. The phonological status of /a/ is still in debate and [7] raises the fact that linguists are not all unanimous on the issue. The same rule is not observed on long vowels. Mean duration also shows that back vowels (/ɔ/, /o/ and /u/) are shorter than front vowels (/i/, /e/, /ɛ/), for both short and long phonemes.  $\Delta$  varies from 24 ms to 50 ms and  $\mathcal{A}$  from 0.27 to 0.56. Vowel /a/ is the one with the strongest length contrast, with large  $r_1$  and  $r_2$  ratios, as well as large area  $\mathcal{A}$  and large  $\Delta$ . Though /ɔ/ is the vowel with the least distinguishable length contrast, with low  $r_1$  and  $r_2$  ratios, small  $\mathcal{A}$  and moderate  $\Delta$ , features unveil that all vowels are length-contrasted. The table also shows that contrast features are correlated but they are complementary to describe the shape of the vowel length distributions. To conclude on this sub-section, this analysis (made fully automatically on 14k vowel tokens) show that our proposed features can highlight different degrees of contrast for each vowel considered and confirm - at a larger scale - previous analyses made.

### 4.2.2. Forced Alignment with Automatic (ASR) Transcriptions

In this sub-section, we try to see if manual transcriptions can be replaced by ASR hypotheses while keeping same trends/conclusions. In that case, we relax the constraint of having manual transcription of the data set. We computed vowel durations from forced alignment obtained with ASR transcripts (from our baseline Wolof ASR system, trained on held-out data - around 20% WER on read speech) and built gamma distributions as in previous section. For each vowel, we compared both distributions (manual transcription vs ASR transcription) using Kolmogorov-Smirnov statistical test [25] (the null hypothesis  $H_0$  was that both distributions obtained after manual and ASR transcriptions are similar). For each vowel  $v$ , no significant difference was found. To illustrate this result, Figure 3 shows duration histograms and associated gamma curves for phoneme /u/ when human (*ref*) or ASR (*hyp*) transcriptions are used for forced-alignment. Both curves are very similar and this confirms that, for read speech, the need of manual transcriptions can be relaxed since the use of ASR leads to very similar measurements and to the same conclusions. For the next sub-sections (semi-spontaneous speech), ASR will be also used to produce transcripts but they will be further corrected by humans due to the more spontaneous nature of the data<sup>4</sup>.

<sup>4</sup>Preliminary measurements have shown that the ASR transcriptions on spontaneous speech are too noisy to be used directly. We got around 31% WER for Wolof and 66% WER for Faana-Faana.

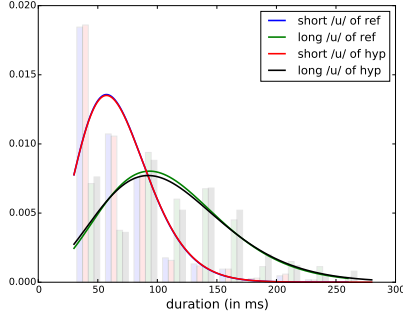


Figure 3: Histogram and Gamma Distribution for /u/ in Wolof Read Speech - Using Human (ref) or ASR (hyp) Transcripts

### 4.3. Analysis on Wolof Semi-Spontaneous Speech

Table 3: Contrast Features Extracted on Wolof Semi-Spontaneous Speech.

Phoneme short long	#occurrences	$\mu$ (in ms)	$r_1$	$r_2$	$\mathcal{A}$	$\Delta$ (in ms)
/i/	1,757	72	<b>1.06</b>	<b>1.01</b>	<b>0.10</b>	<b>11</b>
/i:/	252	83				
/e/	161	71	<b>1.10</b>	<b>1.14</b>	<b>0.19</b>	<b>12</b>
/e:/	213	83				
/ɛ/	518	69	<b>1.40</b>	<b>0.91</b>	<b>0.21</b>	<b>18</b>
/ɛ:/	225	90				
/a/	1,815	60	<b>2.56</b>	<b>1.32</b>	<b>0.44</b>	<b>35</b>
/a:/	324	100				
/ɔ/	360	67	<b>1.22</b>	<b>0.84</b>	<b>0.09</b>	<b>5</b>
/ɔ:/	190	74				
/o/	62	51	<b>6.12</b>	<b>3.26</b>	<b>0.61</b>	<b>35</b>
/o:/	123	89				
/u/	755	51	†	<b>5.95</b>	<b>0.73</b>	<b>44</b>
/u:/	16	96				

† The ratio can not be computed because there were no data for the long unit of the phone ( $d_L(a) = 0$ ) at point corresponding to the mode of the short phone distribution.

We computed same features shown in Table 2 on our Wolof semi-spontaneous corpus. Results are presented in Table 3.

Looking at the mean duration of the vowels  $\mu$ , our first remark is that it is lower in semi-spontaneous speech compared to read speech (for both short and long units). These conclusions were expected but they confirm that our machine-assisted methodology allows usable measurements at a larger scale. Comparing  $\mu$  in read and semi-spontaneous context, we observe that long vowels are the most affected by the speaking style, especially front vowels (/i:/, /e:/ and /ɛ:/), while short units are the least impacted among the vowel set. Results for /u/ have to be taken with caution, since we only have 16 long occurrences, as well as for /o/~ /o:/ for which we have less occurrences compared to other vowels. All computed features show that length contrast on /ɔ/~ /ɔ:/ pair is significantly reduced in semi-spontaneous speech in comparison to what was observed in read speech. In addition, the vowel height has no longer influence on the duration. These findings are consistent with [26] who described that spontaneous speech have an effect on the vowel pronunciation which tends to be more centralized when pronounced shorter.

Table 4: Contrast Features Extracted on Faana-Faana Semi-Spontaneous Speech.

Phoneme short long	#occurrences	$\mu$ (in ms)	$r_1$	$r_2$	$\mathcal{A}$	$\Delta$ (in ms)
/i/	882	69	<b>0.91</b>	<b>1.14</b>	<b>0.09</b>	<b>8</b>
/i:/	167	75				
/e/	77	74	<b>0.87</b>	<b>1.41</b>	<b>0.21</b>	<b>11</b>
/e:/	116	83				
/ɛ/	197	69	<b>1.17</b>	<b>1.06</b>	<b>0.18</b>	<b>17</b>
/ɛ:/	176	87				
/a/	909	63	<b>1.76</b>	<b>1.02</b>	<b>0.32</b>	<b>27</b>
/a:/	188	94				
/ɔ/	197	63	<b>1.12</b>	<b>0.90</b>	<b>0.06</b>	<b>3</b>
/ɔ:/	112	68				
/o/	24	53	<b>2.76</b>	<b>1.40</b>	<b>0.46</b>	<b>21</b>
/o:/	50	77				

† /u/ is not represented because we do not have enough data for a comparison.

### 4.4. Analysis on a Dialectal Variant of Wolof

We computed same features shown in Table 2 and Table 3 on our Faana-Faana semi-spontaneous corpus (see Table 4).

As we can see in Table 4, long vowels /e:/ and /o:/ still appear more frequently than their short counterpart, as in semi-spontaneous (standard) Wolof. We observe that the duration increases with vowel height, for front long vowels (/i:/, /e:/, /ɛ:/) but not for their short counterparts. By looking at the value of the features, we note that distinction between short and long pronunciation of vowels is tenuous. The length contrast on vowel /ɔ/ is also weakened, as in semi-spontaneous (standard) Wolof. These results do not allow to demonstrate that there exists in Faana-Faana a strong opposition of vowels length as observed in (standard) Wolof. In the mean time, we can not affirm that vowel length contrast does not exist in Faana-Faana. In the descriptions of this dialect, as in the Gambian Wolof, the short/long opposition is described, so we can hypothesize that dialectal differences in Wolof are not based on this lack of contrast. In addition, two-sample Kolmogorov-Smirnov tests revealed that /e/, /ɛ/, /a/, /ɔ/ vowel distributions in semi-spontaneous Wolof data set were not found significantly different from those in semi-spontaneous Faana-Faana data set but /i/, /o/ and /u/ vowel distributions were. Finally, since this variant has been little studied, we hope that our analysis represent one first stone in the study of phonemic contrast in Wolof dialects.

## 5. Conclusion

We presented in this study a large scale analysis (compared to previous phonetic studies) of vowel length contrasts in Wolof. We worked on different speaking styles but also on one dialectal variant (Faana-Faana). We proposed correlated but complementary features to describe the shape of the vowel length distributions and to highlight different degrees of length contrast given a vowel. Another important result is that relaxing the constraints on the transcriptions (by using ASR transcriptions instead of manual transcriptions) is possible for read speech since it leads to very similar distributions of durations. Future work will be dedicated to leveraging computational models and machine learning for large scale speech analysis and laboratory phonetics. Further work will analyze the relation between these distinctive features of the length contrast distribution and the functional load concept developed by [27].

## 6. Acknowledgements

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

## 7. References

- [1] N. Ryan and M. Liberman, "Large-scale analysis of Spanish /s/-lenition using audiobooks," in *Proceedings of the 22d International Congress on Acoustics*, Buenos Aires, Argentina, 2016.
- [2] E. Gauthier, L. Besacier, and S. Voisin, "Speed perturbation and vowel duration modelling for ASR in Hausa and Wolof languages," in *Proceedings of Interspeech*, San Francisco, California, USA, September 2016 2016.
- [3] S. Robert, "Le wolof," in *Dictionnaire des langues*, ser. Dicos Poche, J. B. . A. P. Emilio Bonvini, Ed. Quadrige/P.U.F., 2011, pp. 23–30. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-00600630>
- [4] S. Voisin and M. Dramé, "Inaccompli et complexe verbal dans différentes variantes du wolof," *Africana Linguistica*, still in prep.
- [5] S. Voisin, "Le wolof et ses variantes," *JWAL*, still in prep.
- [6] M. Dramé, "Phonologie et morphosyntaxe comparées de trois dialectes wolof," Ph.D. dissertation, UCAD, Dakar., 2012.
- [7] M. T. Cissé, "Problèmes de phonétique et de phonologie en wolof," *Revue électronique internationale de sciences du langage SudLangues*, vol. 6, pp. 1–41, 2006.
- [8] B. Lindblom, "Vowel duration and a model of lip mandible coordination," *Speech Transmission Laboratory Quarterly Progress Status Report*, vol. 4, pp. 1–29, 1967.
- [9] I. Lehiste, *Suprasegmentals*. MIT Press, Cambridge, MA, 1970.
- [10] B. Lindblom, B. Lyberg, and K. Holmgren, *Durational patterns of Swedish phonology: do they reflect short-term motor memory processes?* Indiana University Linguistics Club, 1981, vol. 3.
- [11] A. S. House, "On vowel duration in English," *The Journal of the Acoustical Society of America*, vol. 33, no. 9, pp. 1174–1178, 1961.
- [12] I. Maddieson, "Phonetic cues to syllabification," *UCLA Working papers in phonetics*, vol. 59, pp. 85–101, 1984.
- [13] T. Gay, "Mechanisms in the control of speech rate," *Phonetica*, vol. 38, no. 1-3, pp. 148–158, 1981.
- [14] H. S. Magen and S. E. Blumstein, "Effects of speaking rate on the vowel length distinction in Korean," *Journal of Phonetics*, no. 21, pp. 387–410, 1993.
- [15] S. Myers, "Vowel duration and neutralization of vowel length contrasts in Kinyarwanda," *Journal of Phonetics*, vol. 33, no. 4, pp. 427–446, 2005.
- [16] Y. Adi, J. Keshet, E. Cibelli, E. Gustafson, C. Clopper, and M. Goldrick, "Automatic measurement of vowel duration via structured prediction," *The Journal of the Acoustical Society of America*, vol. 140, no. 6, pp. 4517–4527, 2016.
- [17] G. Lee and D.-J. Shin, "An acoustic and perceptual investigation of the vowel length contrast in Korean," *Journal of the Korean society of speech sciences*, vol. 8, no. 1, pp. 37–44, 2016.
- [18] R. A. Bion, K. Miyazawa, H. Kikuchi, and R. Mazuka, "Learning phonemic vowel length from naturalistic recordings of Japanese infant-directed speech," *PLoS one*, vol. 8, no. 2, p. e51594, 2013.
- [19] R. Sock, "L'organisation temporelle de l'opposition de quantité vocalique en wolof de gambie. sa résistivité aux conditions de durée segmentales et suprasegmentales." Ph.D. dissertation, 1983.
- [20] S. Sauvageot, *Description synchronique d'un dialecte wolof: le parler du Dyolof*. Institut Français d'Afrique Noire, Dakar., 1965, no. 73.
- [21] J. A. Hartigan and P. Hartigan, "The dip test of unimodality," *The Annals of Statistics*, pp. 70–84, 1985.
- [22] E. Gauthier, L. Besacier, S. Voisin, M. Melese, and U. P. Elin-gui, "Collecting Resources in Sub-Saharan African Languages for Automatic Speech Recognition: a Case Study of Wolof," *LREC*, 2016.
- [23] C. Grinevald, "On constructing a working typology of the expression of path," *Faits de langues*, no. 3, pp. 43–70, 2011.
- [24] 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," 2011.
- [25] F. J. Massey Jr, "The Kolmogorov-Smirnov test for goodness of fit," *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.
- [26] C. Gendrot and M. Adda-Decker, "Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German," *Variations*, vol. 2, no. 22.5, pp. 2–4, 2005.
- [27] E. Ferragne, N. Bedoin, V. Boulenger, and F. Pellegrino, "The perception of a derived contrast in Scottish English," in *International Congress of Phonetic Sciences*, Hong Kong SAR China, Aug. 2011, p. ICPHS. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-00613604>