Interspeech 2024
1-5 September 2024, Kos, Greece

# Aerodynamics of Sakata labial-velar oral stops

*Lorenzo Maselli*[1,2,3,4], *Véronique Delvaux*[4]

[1]Fonds Wetenschappelijk Onderzoek - Vlaanderen, Belgium
[2]Universiteit Gent, Belgium
[3]BantUGent, Belgium
[4]Université de Mons, Belgium

`lorenzo.maselli@ugent.be, veronique.delvaux@umons.ac.be`

## Abstract

The present contribution represents the first in-detail exploratory account of the aerodynamics of labial-velar oral stops in Sakata, a Bantu dialect cluster of southwestern Congo. Data collection took place at the phonetics laboratory facilities of Université de Mons with three speakers of central Sakata. Comparative data of labial-velar and plain bilabial oral stops are presented and analysed. Descriptive statistics of the relevant variables are discussed. Given each group of variables, MANOVA results are presented for specially tailored subsets of the whole dataset to investigate variance in the corpus. Sakata labial-velar stops are shown to differ from plain bilabials for duration, airflow, and pressure patterns. Voiceless labial-velar stops exhibit pressure and airflow values consistent with a more prominent lowering of the tongue root / larynx than their voiced counterparts. Matches and mismatches with the available typological literature are also delineated and discussed.

**Index Terms**: aerodynamics, labial-velars, Bantu languages, speech communication, phonetic documentation

## 1. Introduction

Labial-velar (LV) consonants are among the commonest double articulations in human speech, attested in around one tenth of the world's sound systems [1, 2, 3]; for the most part, they occur in Niger-Congo and Nilo-Saharan lects from northern sub-Saharan Africa, and are more sporadically reported in Papua-New Guinea, Oceania, and South America [4]. They are produced with overlapping labial and velar closures released almost simultaneously, with the velar release always preceding the labial [5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. Instrumental analyses of LV consonants are scarce, with only a handful of articulatory and aerodynamic descriptions available in the literature [15, 16, 17, 18, 19]. It is generally accepted that, during the articulation of LV stops, a certain degree of air rarefaction is attained between the two closures, causing the co-occurrence of ingressive and egressive airstream mechanisms upon release [20, 21, 22, 23]. This is most commonly due to a slight backward movement of the tongue root, often accompanied by a lowering of the larynx [24].

LV stops (both oral and nasal) have been reported in Sakata, a dialect cluster of Bantu (Niger-Congo) spoken by around 10-50,000 people in the Mai-Ndombe province of the Democratic Republic of Congo [25, 26]. Sakata is traditionally listed as C34 is Guthrie's alphanumeric classification of Bantu languages [27]. To this day, the available phonological literature on Sakata remains scant [28, 29, 30], with only one source providing detailed acoustic information [31]. This gap is particularly severe in that Sakata is spoken far to the south of the so-called "Macro-Sudan Belt", a vast linguistic macro-area of which LV have been traditionally considered a diagnostic feature [32, 33, 34]. Interestingly, LV have also been reported in other SW Congo lects, though their phonological status appears to be less well-established in those than in Sakata [35]. The phonological inventory of Sakata (adapted from [30, 31]) is reported below, Table 1.

Table 1: *Sakata phonology*

| **Consonants** | Bilabial | Labio-dental | Dental/Alveolar | Palatal | Velar | **Labial-velar** |
|---|---|---|---|---|---|---|
| Stops | p/b | | t/d | | k/g | kp/gb |
| Affricates | | | ts/dz | | | |
| Fricatives | | f/v | s/z | ʃ/ʒ | | |
| Nasals | m | | n | ɲ | ŋ | ŋm |
| Laterals | | | l | | | |
| Trills | | | r | | | |
| Approximants | | | | j | | w |

| **Vowels** | Front | | Back |
|---|---|---|---|
| High | i | | u |
| Mid-high | e | | o |
| Mid-low | ɛ | | ɔ |
| Low | | a | |

The present contribution represents the first empirical aerodynamic study of Sakata LV oral stops. Aerodynamic traces are essential in the study of LV articulations given their typological characteristics, including the occurrence of specific aerodynamic events such as air rarefaction between the velar and labial closures. This paper is particularly concerned with the production of LV oral stops (as well as plain bilabial stops) in two varieties of central Sakata, i.e., Sakata Mbantin and Sakata Mbamushie. The research made available in this venue is chiefly exploratory and should therefore be intended as primary phonetic documentation.

## 2. Methodology and data

### 2.1. Data collection

Full datasets and metadata have been made available on OSF: `https://osf.io/cdfk8/?view_only=1bd311ad49cd4125a64c4c0913fb6658`.

Data collection took place at Université de Mons in March 2023. Data were collected with three L1 speakers of Sakata residing in Europe. All participants are multilingual, fluent in French and Lingala, and highly educated members of their community. They are the only speakers of Sakata we were able to recruit in Belgium and are likely to be among the very few living in Europe. Despite the limited number of participants, this sample is the most representative available as of 2024. Relevant biographical information (pseudonymised) is available below, Table 2.

A set of LV-containing and bilabial-containing Sakata

10.21437/Interspeech.2024-525

Table 2: *Speakers (metadata)*

| | **Sp. A** | **Sp. B** | **Sp. C** |
|---|---|---|---|
| **Age** | 51 | 63 | 64 |
| **Origin** | Bokoro | Mongobele | Mongobele |
| **Variety** | Mbantin | Mbamushie | Mbamushie |

words was produced in collaboration with the speakers, drawing upon information previously reported in the literature [31]. The set includes 22 lexical items, of which 11 contain LV and 11 bilabial stops; one LV–containing word presents two LV segments, and one bilabial-containing word presents 2 bilabial segments. Eight of the 11 LV–containing words display voiceless LV (with their voiced counterparts being regularly preceded by a nasal), while 7 bilabial-containing ones display a voiced bilabial; this lack of balance is largely due to the uneven quality of Sakata lexical documentation at the time of data collection. All segments are word-medial and in intervocalic position, with the exception of one word-initial voiceless LV and the voiced LV stops which are preceded by a nasal and followed by a vowel. A conventional orthography was agreed upon, as well as a carrier sentence. The carrier sentence reads as follows, with hyphens marking the position of the target word: *Mantshii - mba isa, - - -* (meaning: "I say - three times, - - -"; in practice, we elicited 4 repetitions per word, 1 in the carrier sentence and 3 in isolation). Aerodynamic data were collected with PcQuirer 516 (Scicon R and D Inc.). This allowed us to simultaneously record acoustic traces and three channels of aerodynamic data (mouth pressure, oral airflow, nasal airflow). The audio signal was digitised and sampled at 12 kHz, dc channels were sampled at 1 kHz. Mouth pressure was obtained by dint of a catheter introduced at the side of the mouth, bent behind the second molars, and connected to a pressure transducer. Volume flow from the mouth was collected with a Rothenberg mask, with a pressure tube inserted in one of the mask's outlet holes. Pressure and airflow signals were low-pass filtered at 50 Hz. The nasal airflow trace has not been analysed for the purposes of this study, as the focus of this paper does not extend to LV nasal stops.

## 2.2. Data analysis

The recordings were annotated with Praat [36]. Individual tokens deemed unsuitable for phonetic analysis (due to background noise and/or mispronunciations) were manually removed. Based on the acoustic trace, start and end points were identified for the relevant words and segments and marked on two interval tiers; four critical articulatory moments (T1-4) were identified on a third point tier based on the mouth pressure and oral airflow traces. These are described as follows: T1 - initial lowering of the oral airflow trace (associated with the first closure); T2 - initial lowering of mouth pressure (around the velar closure; only for LV); T3 - lowest mouth pressure value (only for LV); T4 - moment of maximum mouth pressure at the labial release. An example of annotated .wav file is presented in Figure 1.

The resulting corpus includes 74 [kp] items, 32 [gb], 73 [p], 82 [b].

Selected variables were extracted from Praat. These include: duration values (entire word, relevant segments, adjacent segments if nasals or vowels, segment boundaries to T1, T2, T3, and T4 for LV and to T1 and T4 for bilabials), oral airflow (sampled at T1, T2, T3, and T4 for LV, and at T1 and T4 for bilabials), and mouth pressure (sampled at T1, T2, T3, and T4 for LV, and at T1 and T4 for bilabials). Duration is expressed in s, oral airflow in l/s, mouth pressure in $cmH_2O$. The relevant datasets were manually inspected. Outliers were excluded from final analysis; missing values for bilabial stops at T2 and T3 were assigned NAs.

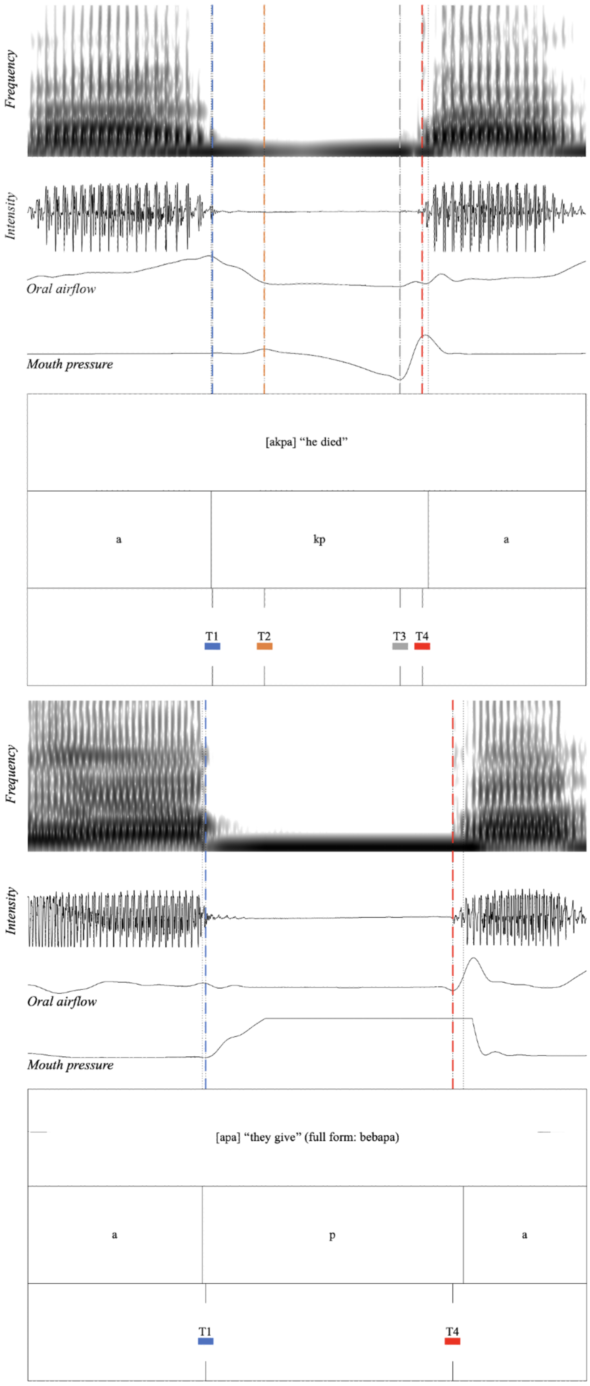

Figure 1: *Overview of standard annotation (LV on top, bilabial at the bottom)*

Descriptive statistics summarising the resulting dataset are presented as bar graphs and scatter-plots in §3.1-3.2. Inferential statistics were performed with RStudio [37] for duration, airflow, and pressure values. We employed Multivariate Analysis of Variance (MANOVA, [38]) to explore the effects of “segment type” (independent variable) on the multiple dependent quantitative variables extracted from Praat. These variables were categorised into three groups (“duration values”, “oral airflow”, and “mouth pressure”), and MANOVAs were run on two datasets with different “segment type” and dependent variable compositions. The first dataset includes four “segment type” values (kp, gb, p, b) and was analysed for all three dependent variable groups. Due to the multidimensional nature of the “duration values” group, a Principal Component Analysis (PCA) was performed to reduce dimensionality and extract meaningful patterns. This approach allowed us to assess the overall effect of “segment type” on the combined dependent variables within each group. The second dataset is limited to LV values of “segment type” (kp, gb) but expands the number of dependent variables, introducing additional measures taken at T2 and T3. PCA was again utilised for “duration values” to condense the variables into principal components, which were then analysed using MANOVA. The application of PCA prior to MANOVA for “duration values” in both datasets was key in reducing data complexity and highlighting the underlying structure of the data. This step was crucial for handling the high dimensionality and collinearity among variables, thus ensuring a more robust and interpretable analysis. “Oral airflow” and “mouth pressure” were handled without PCA as they did not raise the same collinearity concerns as “duration values”.

## 3. Results

### 3.1. Duration

Duration differences are more prominent in the case of LV than bilabials. On average, voiceless [kp] are longer than voiced [gb]; see Figure 2.

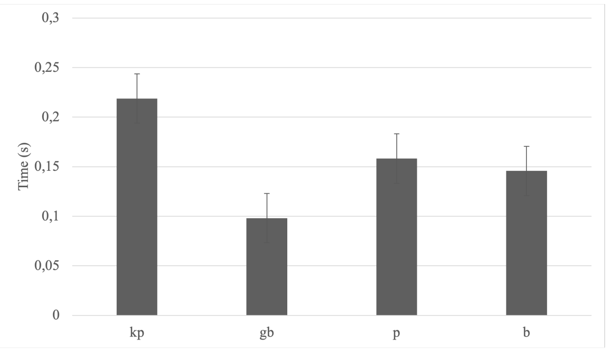

Figure 2: *Segment duration values (with standard error bars)*

Segments preceding LV appear to be longer (for [kp]) or shorter (for [gb]) than those preceding plain bilabials. Following segments are only shorter than preceding ones in the case of [kp]. Absolute highest following-segment duration values are those reported for [b], with all other sound classes behaving similarly in that position. Though absolute highest preceding-segment duration values are reported for [kp], they appear to fall within the same error range as [b]; see Figure 3.

MANOVA results for "duration values" reveal a significant effect of "segment type" on the relevant quantitative variable (i.e., the score of the PCA limited to its first two components, which by themselves explain over 89% of the variance in the corpus). This is true of both the four- and the two-level independent variable datasets, as summarised in Table 3.

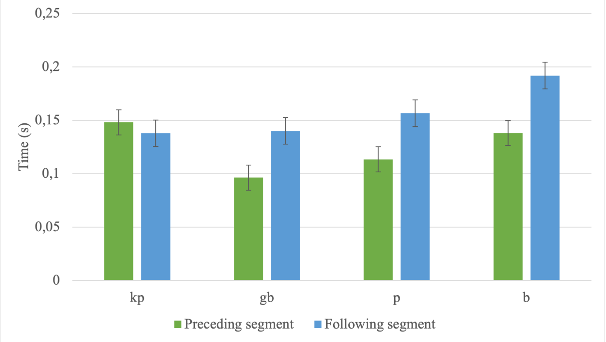

Figure 3: *Adjacent segment duration values (with standard error bars)*

Table 3: *MANOVA results ("duration values")*

| **Independent variables** | Df | approx F | *p*-value |
|---|---|---|---|
| Segment type (4 lvl) | 3 | 115.05 | <.001 |
| Segment type (2 lvl) | 1 | 78.017 | <.001 |

### 3.2. Aerodynamics

Considerably more airflow can be detected at T1 and T4 on LV than plain bilabials, with the latter displaying some negative flow (air suction) upon release. This is expected right before the typical surge in airflow accompanying explosive bursts (see Figure 1). Voiceless LV exhibit more positive airflow upon release than their voiced counterparts; see Figure 4.

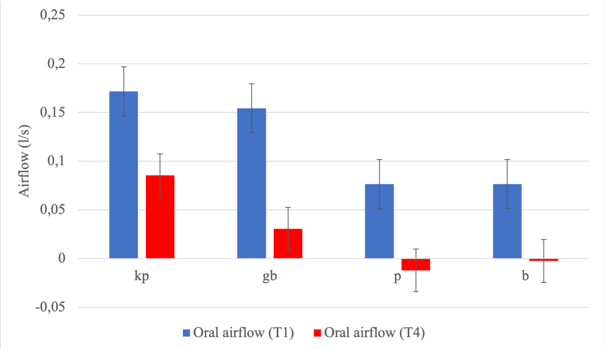

Figure 4: *Oral airflow at closure and release (with standard error bars)*

Mouth pressure values are consistently higher at T4 than T1. Mouth pressure values at T1 are highest for [gb] and lowest for [kp]. However, mouth pressure values at T4 are importantly higher for bilabials, and [p] in particular, than LV. Standard error indicates that negative mouth pressure can be found at T4 on both LV (and at T1 on [kp]); see Figure 5.

Scatter-plots are presented in Figure 6, representing LV oral airflow (top) and mouth pressure (bottom) values at the four articulatory moments, over time (0 on the time axis represents

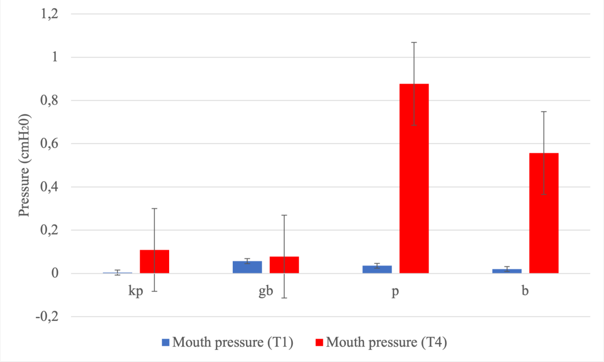

Figure 5: *Mouth pressure at closure and release (with standard error bars)*

the beginning of the relevant segment). For the two aerodynamic measures, LV display similar aerodynamic trajectories irrespective of voicing, with suction detected at T2 and T3. T3 mouth pressure levels are markedly lower for [kp] than [gb]. As concerns the timing of the four events, the first closure typically occurs before the sound's acoustic trace is audible, though voiceless and voiced LV differ in that, on average, voiced-LV T2 also precedes the sound's left edge. Inter T2-T3 time is longer for [kp] than [gb]. T4 immediately follows T3 in both voicing contexts.

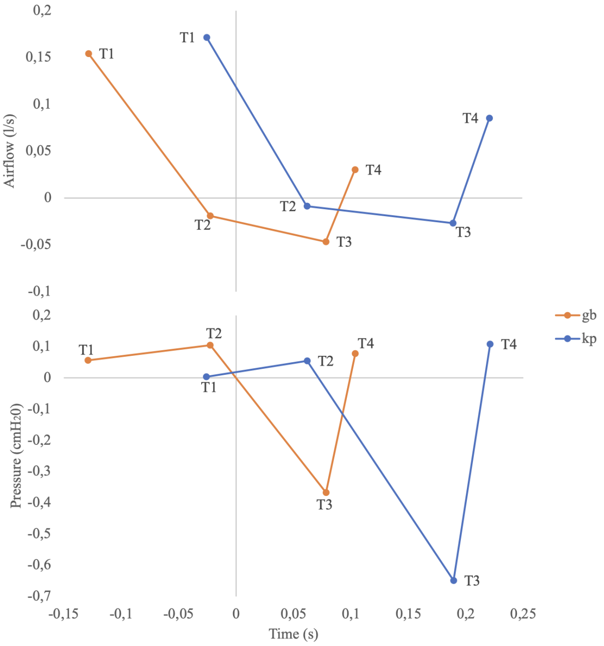

Figure 6: *Schematic aerodynamic trajectories of LV at the four relevant articulatory moments*

MANOVA results for "oral airflow" and "mouth pressure" reveal a significant effect of "segment type" on the variables, as can be seen in Table 4.

Across all analyses, significant *p*-values consistently point to the critical role of "segment type" in explaining corpus variance. The difference in effect sizes between the two variable groups and across the different levels of "segment type" complexity suggests that the impact of "segment type" on "mouth pressure" is more pronounced than on "oral airflow". This pattern holds true irrespective of segmentation granularity, albeit with nuanced differences in magnitude between categorisations.

Table 4: *MANOVA results (aerodynamics)*

| **Qualitative variables** | Df | approx F | *p*-value |
|---|---|---|---|
| Oral airflow (4 lvl) | 3 | 7.4233 | <.001 |
| Oral airflow (2 lvl) | 1 | 7.2378 | <.001 |
| Mouth pressure (4 lvl) | 3 | 62.191 | <.001 |
| Mouth pressure (2 lvl) | 1 | 10.636 | <.001 |

## 4. Discussion

Data from Sakata confirm previous findings in the literature concerning the relative duration of [kp] and [gb] [39, 5], with the former exhibiting higher values than the latter, but no specific effects of segment type on LV- and bilabial-adjacent segment duration were detected. Differences in duration between [p] and [b] are less marked but in line with common typological differences between voiced and voiceless segments. Oral airflow values differ significantly between LV and bilabials, with the former displaying higher volume flow at both first closure and release. Mouth pressure build-up prior to release is higher for bilabials than LV. Air rarefaction after completion of the second closure is more important for [kp] than [gb]. Differences in inter T2-T3 time appear to explain the durational difference between [kp] and [gb], indicating a more prolonged period of air rarefaction in the articulation of [kp] than [gb]. This matches previous findings concerning the relative salience of voiceless and voiced LV in other Bantu languages of the region of interest [35, 31], and is compatible with mouth pressure levels at T3, for which the two segments have been shown to differ significantly (the longer duration of the articulatory event leading to air rarefaction between the two closures causes a more significant drop in mouth pressure). A cursory analysis of spectral differences between [kp] and [gb] does not offer clear-cut cues to differences in low-frequency energy concentrations ("voicing bar") within the LV category. This might indicate the presence of a glottalic airstream mechanism [15, 40], i.e., a lowering of the larynx, in the articulation of LV in the corpus. Differences in T1-T2 timing relative to the acoustic left edge of the LV might be attributable to the fact that voiced LV are often preceded by nasals, suggesting that the labial closure already occurs during the nasal articulation. While this consideration exceeds the scope of the present contribution, it might indicate that pre-LV nasals in Sakata are also articulated as LV and therefore assimilate to [gb] [41, 42, 43, 44, 45]. T1-T4 timing is comparable across LV types, in spite of general duration differences between voiced and voiceless LV. Negative airflow at T2 indicates that larynx, and possibly tongue body, lowering (and, consequently, air rarefaction in the oral cavity) is initiated before the labial closure is complete. Positive oral airflow values at T4 on LV mark the presence of an egressive pulmonic airstream mechanism with a sudden rise in pressure upon release of the two closures [22, 11]; lack of a similar phenomenon in bilabials is marked by oral airflow values approximating 0 at T4. Higher mouth pressure values at T4 for bilabials also reflect this difference, with pressure building up throughout the articulation (i.e., without the typical dip at T3 which marks their LV counterparts).

## 5. Acknowledgements

The first author acknowledges the support of FWO (Fonds Wetenschappelijk Onderzoek - Vlaanderen) PhD fellowship n°11D7221N. The second author is a Research Associate of the Fonds de la Recherche Scientifique – FNRS, which she wishes to acknowledge here. We acknowledge the help of three anonymous reviewers as well as the support and contribution of the three pseudonymised consultants that made this research possible.

## 6. References

[1] I. Maddieson, "Phonetics and african languages," Berlin, pp. 546—601, 2018.

[2] S. Moran and D. McCloy, "Phoible 2.0," 2019.

[3] M. Cahill, "Why labial-velar stops merge to /gb/," *Phonology*, vol. 25, pp. 379—398, 2008.

[4] ——, "Labial-velars: A questionable diagnostic for a linguistic area," in *Proceedings of the 8th World Congress of African Linguistics Kyoto 2015*, 2017.

[5] P. Ladefoged and I. Maddieson, *The Sounds of the World's Languages*. Oxford: Blackwell, 1996.

[6] I. Maddieson and P. Ladefoged, "Multiply-articulated segments and the feature hierarchy," *UCLA Working Papers in Phonetics 72: 116-138*, 1989.

[7] C. Painter, "Implosives, inherent pitch, tonogenesis, and laryngeal mechanisms," *Journal of Phonetics*, vol. 6, pp. 249–274, 1978.

[8] B. Connell, "Temporal aspects of labiovelar stops," *Work in Progress, 20, Dept. of Linguistics, University of Edinburgh*, pp. 53–60, 1987.

[9] ——, "Accounting for the reflexes of labial-velar stops," *Proceedings of the XII ICPhS*, vol. 3, 1991a.

[10] ——, "Phonetic aspects of the lower cross languages and their implications for sound change," Ph.D. dissertation, University of Edinburgh, 1991b.

[11] ——, "The structure of labial-velar stops," *Journal of Phonetics*, vol. 22, pp. 441–476, 1994.

[12] G. Dogil, "On the acoustic structure of multiply articulated stop consonants (labio-velars)," *Wiener Linguistische Gazette*, vol. 42, no. 43, pp. 3–55, 1988.

[13] M. Cahill and J. Hajek, "Why [kp]?" 2001.

[14] M. Cahill, *Labial-velars of Africa: Phonetics, phonology, and historical development*. Routledge, 2018.

[15] P. Ladefoged, *Airstream mechanisms and double articulations*. Cambridge University Press, 1968, ch. 2.

[16] I. Maddieson, "Investigating ewe articulations with electromagnetic articulography," *UCLA Working Papers in Phonetics, Vol. 85*, 1993.

[17] D. Demolin and B. Teston, "Aspects aérodynamiques et articulatoires des occlusives labio-vélaires," *XXIèmes journées d'études sur la parole. Avignon, 111-114*, 1996.

[18] D. Demolin and A. Soquet, "Double articulations in some mangbutu-efe languages," *Journal of the International Phonetic Association*, vol. 29, 1999.

[19] S. Grawunder, B. Winter, and J. Atoyebi, "Voicing of labiovelar stops in yoruba," *ICPhS XVII*, 2011.

[20] E. Mills, "Senoufo phonology: discourse to syllable," *Dallas: SIL*, 1984.

[21] P. Vogler, "Le parler vata," *Strasbourg: Université des Sciences Humaines de Strasbourg*, 1987.

[22] D. Demolin, "Les consonnes labio-vélaires du mangbetu," *Pholia (6): 85-105*, 1991.

[23] P. Ladefoged, "Linguistic aspects of respiratory phenomena," *Annals of the New York Academy of Sciences, 155: 141-151*, 1968.

[24] ——, *A Course in Phonetics, 3rd ed.* Fort Worth : Harcourt Brace Jovanovich College Publishers, 1993.

[25] G. V. Bulck, *Notions de grammaire et début de vocabulaire de kisaghata. Kimbari (Congo-Kinshasa?): Vill. de Makaw, Kote Moke.*

[26] L. Mundeke, *Réflexes consonantiques du sakata. Kinshasa: Centre de Linguistique Théoretique et Appliquée (CELTA).*

[27] M. Guthie, *Comparative Bantu: An Introduction to the Comparative Linguistics and Prehistory of the Bantu languages. Volume 2: Bantu Prehistory, Inventory and Indexes*. London: Gregg International, 1971.

[28] K. M. Matabisi, "Phonologie, morphonologie et morphologie de la langue sakata," Master's thesis, Kinshasa: IPN, Kinshasa, 1979.

[29] T. Tylleskär, "Phonologie de la langue sakata (bc 34): langue bantoue du zaïre parler de lemvien nord," Master's thesis, Université de la Sorbonne Nouvelle (Paris 3), 1987.

[30] J. B. Bosonkie, "Structure phonologique du kisakata c34. travail de fin d'études," 1997.

[31] L. Maselli, V. Delvaux, J.-P. Donzo, S. Pacchiarotti, and K. Bostoen, "Labial-velar stops in sakata (bantu c34): A preliminary phonetic and diachronic-phonological approach," *Studies in African Linguistics*, 2023.

[32] N. Clements and A. Rialland, "Africa as a phonological area, in heine, b., nurse, d. (eds), a linguistic geography of africa, cambridge, cup."

[33] T. Güldemann, "The macro-sudan belt: Towards identifying a linguistic area in northern sub-saharan africa. in b. heine d. nurse (eds), a linguistic geography of africa, 151-185. cambridge/new york: Cambridge university press."

[34] D. Idiatov and M. Van de Velde, "The lexical distribution of labial-velar stops is a window into the linguistic prehistory of northern sub-saharan africa," *Language*, vol. 97, pp. 107 – 72, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:234242736

[35] L. Maselli, S. Pacchiarotti, and K. Bostoen, "Labial-velar stops outside the macro-sudan belt: New evidence from lwel (west-coastal bantu, b862)," *Africana Linguistica*, 2021.

[36] P. Boersma and D. Weenink, "Praat, a system for doing phonetics by computer," *Glot International*, no. 9/10, pp. 341–345, 2001.

[37] RStudio Team, *RStudio: Integrated Development Environment for R*, RStudio, PBC., Boston, MA, 2020. [Online]. Available: http://www.rstudio.com/

[38] B. G. Tabachnick, L. S. Fidell, and J. B. Ullman, *Using multivariate statistics*. pearson Boston, MA, 2013, vol. 6.

[39] D. Demolin, "Le mangbetu : Etude phonétique et phonologique," Ph.D. dissertation, Université Libre de Bruxelles, 1992.

[40] B. Siertsema, "Some notes on yoruba phonetics and spelling," *Bulletin de l'Institut Français d'Afrique Noire XX B : 576-95*, 1958.

[41] H. Goldie, *Dictionary of the Efik Language (in two parts). Edinburgh.*

[42] W. E. W. 1968, "Jukun of wukari and jukun of takum," *Occasional Publication 16. Ibadan.*, 1968.

[43] W. E. Welmers, *African Language Structures*. UCP, 1973.

[44] M. E. Ryder, "An autosegmental treatment of nasal assimilation to labial-velars," *Proceedings of the Chicago Linguistics Society 23.2:253-65*, 1987.

[45] M. Cahill and F. Parkinson, "Partial class behavior and feature geometry: Remarks on feature partial class behavior and feature geometry: Remarks on feature class theory," *Proceedings of the North East Linguistic Society 27*, 1997.