Effect of vowel context on stop place identification in Yoruba

Roslyn Burns and Jason Shaw

Yale University
roslyn.burns@yale.edu jason.shaw@yale.edu

ABSTRACT

Vowel context is a known factor in palatalization, in part because of how co-articulation impacts consonant place cues. While much is known about palatalization triggers and simplex place targets, complex places remain understudied. Here, we investigate the perception of Nupe labial-velars by Yoruba listeners across vowel contexts in an identification task. Results show that labial-velar confusability follows the hierarchy of palatalization triggers established for simplex segments. Labial-velars exhibit two different outcomes in palatalizing environments, one in which the dorsal component is interpreted as a coronal, but also one in which the dorsal component is ignored leading to confusion with labials.

Keywords: labial-velars, palatalization, sound change, Yoruba, Nupe, sound change

1. INTRODUCTION

Vowel context has been shown to influence the perception of velar stops /k/ and /g/ in the environment of /i/, where they become more confusable with the anterior segments /tʃ/ and /ʤ/ respectively. The increased confusability is attributed to CV co-articulation which brings the tongue dorsum forward, thus reducing the perceptual distinctiveness between velars and anterior affricates [1]. Our study is primarily concerned with whether a similar perceptual effect of vowel place can be observed with double closure labial-velar stops, which are understudied with respect to perception.

There are some reasons to suspect that vowel context may impact labial-velar stop perception. First, labial-velar stops are under-attested preceding high front vowels in some West African languages, which is observable in both Igbo and Yoruba dictionaries [2, 3, 4]. Second, co-articulation with a high front vowel could reduce articulatory movements that distinguish labial-velars from other stop places. There are two considerations. First, the labial component of labial-velars can differ from simplex labials in how the jaw contributes to the constriction and release; jaw lowering just before labial release can give labial-velars an implosive quality [5], which may differentiate them from simplex labials. Second, the velar component of labial-velars can differ from simplex velars, with the tongue dorsum being more retracted in labial-velars than simplex velar counterparts [5, 6, 7]. These two articulatory enhancements (jaw lowering and tongue dorsum retraction) may be suppressed by co-articulation with a high front vowel, which requires jaw raising and tongue fronting, possibly leading to increased perceptual confusion with other places.

To evaluate the effect of vowel context on labial-velar stop perception, we ran a cross-language perception study. Native speakers of Yoruba (Niger-Congo: Yoruboid) categorized stops, including labial-velars, produced by a Nupe (Niger-Congo: Nupoid) speaker in different vowel contexts. Both languages have labial-velar stops, but we decided to use Nupe items as stimuli because labial-velars are particularly robust in this language. Nupe exhibits a full six-way contrast in voicing across labial-velar, velar, and labial places of articulation: /k͡p/, /k/, /p/, /g͡b/, /g/, /b/ [8]. Yoruba only contrasts /k͡p/, /k/, /g͡b/, /g/, and /b/ [9]. We used Yoruba as the listener language because of the accessibility of online participants and because Yoruba is the most widely used Niger-Congo language in Nigeria (and globally) with an estimated 42,000,000 native users.

In addition to the /i/ context, we also included /e/, /u/, /a/ contexts in our experiment. Our predictions for the effect of these vowel contexts on stop place identification draw from the literature on sound change. Cross-linguistic surveys show that simplex velars are likely to become either alveolar or palatal in the environment of /i/, i.e., “palatalization” [10, 11, 12]. The vowels /e/ and /u/ can also condition the same sound change but to a lesser extend than /i/. The least likely trigger is /a/. Notably, if vowels other than /i/ trigger palatalization, they are subject to the implicational hierarchy in (1) [10, 11].

(1) a. Low front vowels do not trigger palatalization unless high front vowels do.
b. High back vowels do not trigger palatalization unless high front vowels do.

By including all four vowels, /i/, /e/, /u/, /a/, we test whether perceptual confusions vary across vowel contexts in the way that reflects the likelihood of sound change from velar to a different (more anterior) lingual place of articulation.

3. METHODS

3.1 Nupe stimuli design

A native speaker-linguist of Nupe, Ahmadu Ndanusa Kawu, produced 96 disyllabic nonce word targets of the shape C1V1C2V2 where C1 varied in voicing (voiced, voiceless) and place (labial-velar, dorsal, labial, coronal), V1 varied in tone (H, L) and quality ([i], [u], [e], [a]), C2 was always coronal, and V2 was always [a] and matched the tone of V1. Labial-velar-initial words had three variants of C2 ([t], [d], [s]) whereas all other places had only one ([t]).

The speaker produced the CVCV sequences in two Nupe carrier phrases *wun ka* ___ *be* 'he wrote ___ again' and *wun ganan* ___ *be* 'he said ___ again' for a total of 192 utterances. All utterances were recorded on a Zoom Hn4 recorder at a sample rate of 44.1 khz. These utterances were forced aligned with the Montreal Forced Aligner using a Nupe dictionary created by the researchers and an English pronunciation model. Boundaries were checked by hand in Praat [13]. The C1 boundary was moved to the nearest zero-crossing preceding the stop burst for obstruents and to the onset of F2 rise for [w]. The end of the V1 boundary was shifted to the nearest zero crossing at the offset of vowel voicing.

Of the 192 recorded utterances, 96 target items were selected for inclusion in the experiment. Audio stimuli were selected by identifying the most peripheral vowel production for each target. Once we selected a representative token for each item, we excised the initial CV portion of each target word, based on the boundaries described above, to serve as stimuli for the experiment. The stimuli were then normalized to 70dB in Praat. Formant values of the stimuli were extracted at seven evenly spaced intervals using an Inverse Filter Control script [14, 15]. Burst frequencies were extracted in Praat with a 300 hz high pass filter.

3.2 Yoruba participants

49 native speakers of Yoruba who grew up in Nigeria until age 18 were recruited on Prolific.co. Everyone reported speaking, listening, reading, and writing proficiency in Yoruba and proficiency in English. Participants ranged in age from 18 to 59.

3.3 Procedure

As labial-velars across West Africa are phonetically heterogeneous [5] and this is the first study of its kind, we opted for an open response task instead of, e.g, forced choice. Participants were directed to an online experiment written in Yoruba (translation by Oluwaseyi Fasunhan) on the Gorilla platform. Participants were told that they would hear the first part of a word and were instructed to type what they heard as a response. Prior to the 96 experimental trials, there were 7 practice trials of stimuli not used in the actual experiment.

3.3 Response coding

Since we were primarily interested in whether vowel context impacted the perception of labial-velar consonant place of articulation, we coded responses according to this dimension. Place was coded into one of four major place categories, labial, coronal, which collapsed over alveolar and palatal, dorsal, and labial-velar, based on the orthographic representation provided by the participant. In line with previous studies involving non-phonemic palatal contrast [1], we did not distinguish between palatal and alveolar responses since both involve fronting. A fifth category, neither, was coded for non-buccal segments and vowels. As with other perceptual studies, these places represent phonological categories as opposed to exact phonetic places of articulation (i.e. unless the orthography encodes allophones, then allophones are not coded).

Due to variation in how responses were entered (e.g. additional words, providing entire words, etc.) all tokens were coded for place by hand (n= 4705). 83 trials were discarded for failing to respond, containing uninterpretable place properties, and technical errors. Because voiced phonemes in Yoruba exhibit the full set of contrasts across the four investigated places we restricted our analysis to voiced stimuli (n= 2312).

In many Nigerian languages, <kp> represents [k͡p], <p> represents [p], and <gb> represents [g͡b]. Nigerian Yoruba uses <p> for both [k͡p] and [p], which are not contrastive, and lacks <kp>. Yoruba is, however, spoken in countries which impose national orthographies that distinguish between <kp> [k͡p] and <p> [p] [16]. In some cases, participants provided voiceless responses to voiced stimuli. In these cases, <p> was coded as both labial-velar and labial (n = 264) as orthographically it represents both, and <kp> was coded as labial-velar (n = 43).

After coding for place, responses were coded as either “correct” or “incorrect” based on whether the response place matched the original stimuli's coded place. Because the Nupe speaker produced /gu/ as [wu], both <g> and <w> were accepted as correct responses. The reported findings in section 3.2 are made on the basis of incorrect responses.

3. RESULTS

3.1 Nupe stimulus items

Table 1. summarizes the stimuli properties grouped by consonant type (columns) and vowel type (rows). Table 1a on the top shows the highest amplitude frequency of the stop burst, Table 1b in the middle shows the F2 measurement at the first of seven evenly spaced time intervals in the vowel (T1=0%), and Table 1c on the bottom shows the change in F2 between the first and second time intervals (Δ 17%-0%).

	coronal		dorsal		labial		labial-velar	
	mean	sd	mean	sd	mean	sd	mean	sd
a. Maximum amplitude frequency of stop burst								
/a/	1357	355	1389	55	1336	50	1185	239
/e/	3274	3014	2584	870	1637	187	1404	207
/i/	3058	1708	4533	1734	1379	141	1314	165
/u/	3015	3618	1163	87	1120	122	1644	1576
b. F2 at the onset of voicing								
/a/	1671	75	1456	60	1353	49	1283	60
/e/	1656	390	1741	88	1727	65	1602	83
/i/	1594	276	1641	130	1652	322	1551	194
/u/	1479	151	1143	278	1221	171	1143	114
c. F2 change at the onset of the vowel (first 17%)								
/a/	-50	12	28	54	5	50	70	52
/e/	-141	225	-68	193	-39	176	-29	74
/i/	-52	74	-24	88	1	139	-63	118
/u/	-110	66	-81	150	-66	68	-86	68

Table 1: Stimuli properties in hz

Coronal stimuli often have the highest burst frequency followed by dorsals, labials, and then labial-velars. Before /e/, dorsals have higher burst frequencies than coronals. Before /u/, all places except for coronals exhibit low burst frequencies.

On average, labial-velars exhibit the lowest F2 values at T1 in all vowel contexts, except for /u/, while coronals exhibit the highest values.

In the F2 change over the vowel onset, negative numbers indicate a decrease in frequency; positive numbers indicate an increase; and 0 indicates no change. Before /u/, coronals have the greatest F2 decrease followed by labial-velars. Before /i/, F2 increases the most in segments articulated with the lips. Before /e/, coronals exhibit the greatest decrease in F2.

Although burst frequencies are cues that can potentially help listeners distinguish labial-velars from other places (Table 1a), if listeners rely on F2 transition cues (Table 1b-c), labial-velars may be more confusable with other places of articulation in palatalizing contexts.

3.2 Yoruba perception

Participants mostly responded with the correct answer (mean= 0.65) but there is a wide amount of variability (sd= 0.47). Fig. 1. shows the error rate across the 49 participants in different consonant and vowel contexts. The standard error estimates the variability across individual speakers.

As shown in Fig. 1., consonant places exhibit different error patterns across vowels. To evaluate the significance of the error patterns, we fit nested linear mixed effects models (LMER) to the data. The baseline model included consonant place and vowel (fixed) and participant (random intercept). To this baseline, we added a place*vowel interaction. An ANOVA comparison showed that the model with the interaction term was significantly better than the baseline (AIC 2785.9 vs. baseline AIC 2911.7, $\chi 2=144$, $p < 0.05$), indicating that the differential effect of vowel on consonant confusions is reliable.

Labial-velar errors appear to follow the vowel place dependencies outlined in (1). That is, among front vowels error rates were highest for /i/ followed by /e/ then /a/. Among high vowels, error rates were higher for /i/ than for /u/. We evaluated this trend with a LMER model of the labial-velar data predicting accuracy by vowel (fixed) and participant (random intercept). Table 2 shows the fixed effects.

	Estimate	Std. Error	pr(>|t|)
/g͡bi/ (intercept)	0.66	0.04	< 2e-16
/g͡ba/	0.16	0.03	3.33e-07
/g͡be/	0.06	0.03	0.03926
/g͡bu/	0.12	0.03	0.00023

Table 2: Voiced labial-velar response fixed effects.

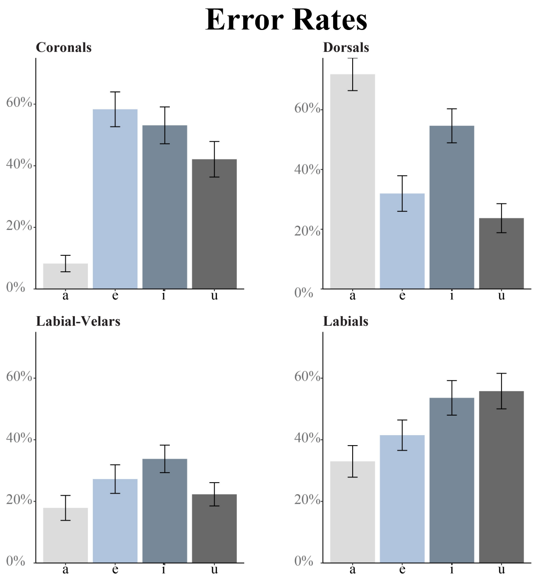

Figure 1: Error rate averaged across subjects

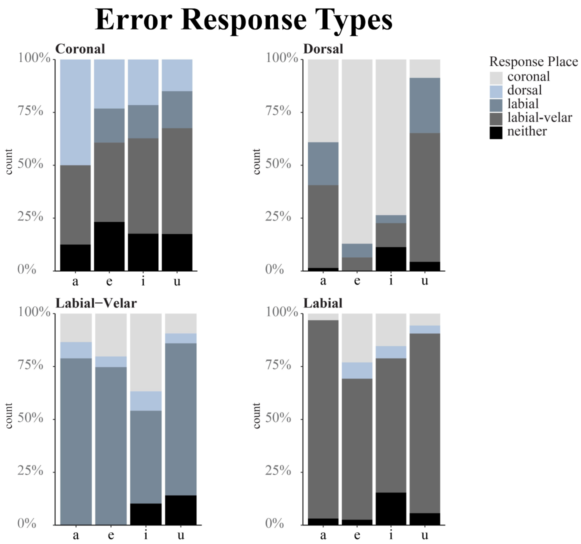

Figure 2: Error response types

Each vowel significantly differs from /i/ (intercept) and accuracy increases from /i/ to /e/, /u/, and then /a/. To directly compare /e/ and /a/, we made /a/ the intercept [17]. In this model, the /e/ coefficient was -0.1 ($p < 0.05$). Thus, the trends corresponding to (1a) and (1b) among labial-velars are statistically supported.

Fig. 2. shows the distribution of erroneous responses by consonant and vowel. As shown in Fig. 2., /g͡b/ is mostly mistaken as labial and /b/ is mostly mistaken as labial-velar. This is likely because, as shown in Table 1, acoustic cues for labial-velars and labials are highly similar to each other. In palatalizing contexts, /g͡b/ is increasingly mistaken as coronal following the hierarchy in (1). Notably, /b/ lacks increased confusability with coronals in this same context implicating the dorsal constriction as the likely source of the increased coronal confusability for labial-velars.

Comparison between /g͡b/ and /g/ suggests that the dorsal closures have similar confusability profiles save for the high error rate of /ga/ in an apparent violation of (1a). Fig. 2. reveals that /ga/ compounds errors involving labial closures (simplex and complex) and those involving coronal closures. Focusing on just the misperceptions of /g/ as coronal, /ga/ has lower error rates than both /gi/ and /ge/ which conforms to (1a); labial-velars show the same pattern. One exception is that /ge/ has a higher error rate than /gi/. This possibly because velars in the context of /e/ have higher F2 values than velars in the context of /i/, which may be an artifact of our stimuli. The above exceptions notwithstanding, the confusions for both /g͡b/ and /g/ are relatable to the palatalization hierarchy, but the confusions for /b/ are not.

In line with cross-linguistics studies of sound change [10, 11, 12], velars in our speech perception task were more likely to be misidentified as coronals in the contexts identified in (1) but notably labial-velars were also more likely to be identified as coronal in the same contexts. Unlike velars, however, labial-velars are also highly confusable with labials. Despite the increased confusability with coronals in palatalizing contexts, labial-velars are still frequently (mis)identified as labials, but notably labials do not follow the same vowel-induced pattern of confusability with coronals. These findings have implications for theorizing about sound changes. When the cues to labial-velar stops are compromised in palatalizing contexts, labial-velars may be perceived as velars. If language users prioritize the maintenance of cues which distinguish labial-velars from velars, labial-velars may undergo simplification by losing the dorsal component in palatalizing contexts (i.e. labial-velar > labial) thus leading to sound change.

4. CONCLUSION

Our study found an effect of vowel context on the perception of labial-velar stops. The likelihood of being misperceived as coronal follows the palatalization hierarchy outlined in previous work on simplex velars. This is likely due to the fact that in palatalizing contexts, the cues which make labial-velars unique deteriorate due to co-articulation with the neighboring vowel. This suggests the potential for diachronic instability of labial-velars in palatalizing contexts.

5. REFERENCES

[1] Guion, S. 1998. The role of perception in the sound change of velar palatalization. *Phonetica* 55(1–2): 18–52.

[2] Echeruo, M. J. C. 1998. *Igbo-English dictionary : a comprehensive dictionary of the Igbo language with an English-Igbo index*. New Haven: Yale University Press.

[3] Igwe, G. E. 1999. *Igbo-English Dictionary.* Ibadan: University Press.

[4] 2001. *A dictionary of the Yoruba language*. Ibadan: University Press.

[5] Ladefoged, P. 1968. *A Phonetic Study of West African Languages: an auditory-instrumental survey.* Cambridge: Cambridge University Press.

[6] Connell, B. 1994. The structure of labial-velar stops. *Journal of Phonetics* 22(4): 441–476.

[7] Hudu, F., Miller, A., & Pulleyblank, D. 2009. Ultrasound imaging and theories of tongue root phenomena in African languages. In Austin, P. K., Bond, O., Charette, M., Nathan, D., & Sells, P. (eds.), *Proceedings of Conference on Language Documentation & Linguistic Theory* 2, 153–163. London: SOAS.

[8] Pulleyblank, D. 1984. Yoruba. In Comrie, B. (ed.), *The World's Major Languages 2nd edition*, 866–882. London: Routledge.

[9] Moran, S. & McCloy, D (eds.) 2019. PHOIBLE 2.0. Jena: Max Planck Institute for the Science of Human History. http://phoible.org

[10] Bateman, N. 2007. *A Crosslinguistic Investigation of Palatalization.* Dissertation: University of California San Diego.

[11] Bateman, N. 2011. On the typology of palatalization. *Language and linguistics compass* 5(8): 588–602.

[12] Bhat, D. NS. 1978. A general study of palatalization. In Greenberg, J. (ed.), *Universals of Human Language* vol. 2, 47–92. Stanford: Stanford University Press.

[13] Boersma, P. & Weenink, D. 2019. *Praat: doing phonetics by computer* v. 6.0.46. praat.org

[14] Sprouse, R. & Johnson, K. 2016. The Berkeley phonetics machine. *UC Berkeley PhonLab Annual Report* 12(1).

[15] Ueda, Y., Hamakawa, T., Sakata, T., Hario, S., & Watanabe, A. 2007. A real-time formant tracker based on the inverse filter control method. *Acoustical science and technology* 28(4): 271–274.

[16] Tchitchi, T. & Hazoumé, M. L. 1983. Le dévelopment d'un alphabet nacional dans une communauté multilingue. In Winter, W. (ed.), *Writing in Focu*s, 187–208. Berlin: Mouton.

[17] Baayen, R.H. 2008. *Analyzing Linguistic Data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.