INTRUSIVE VOCOIDS AND SYLLABLE STRUCTURE IN GEORGIAN

Caroline Crouch[1], Ioana Chitoran[2], Louis Goldstein[3], Argyro Katsika[4]

[1]Rice University, [2]Université Paris Cité, [3]University of Southern California, [4]University of California, Santa Barbara
cc185@rice.edu; ioana.chitoran@u-paris.fr; louisgol@usc.edu; argyro@ucsb.edu

ABSTRACT

Complex onsets in Georgian are characterized by long lags between consonant constrictions, which differ depending on the sonority shape of the onset cluster. These lags result in relatively open vocal tract configurations where vocoids sometimes intrude. In this study, we examine the phonetic properties of such vocoids and present evidence that, although they emerge from the cluster's articulatory configuration, they may create conditions for resyllabification. Specifically, results show that, although the distribution of vocoids depends on sonority shape (more appear in rises, fewer in falls), their formant structure and duration remain stable and unaffected by the cluster's sonority and the nucleus vowel's quality. They are shorter in duration than Georgian lexical vowels. Yet, preliminary analyses show that vocoids affect the syllable-scale structure of the amplitude envelope in ways that could lead to resyllabification in sonority falls. Implications for Georgian syllabic structure are discussed.

Keywords: vocoids, sonority, syllabic structure, resyllabification, Georgian

1. INTRODUCTION

Georgian syllable structure is typologically unusual: onsets can be up to seven consonants in length, with minimal restrictions on consonant combinations. Complex syllable onsets in Georgian often include transitional, schwa-like elements in between consonants. Word onset clusters are also characterized by long lags between the constrictions of the consonants [1][2]. Interestingly, the duration of these lags differs according to sonority shape, as does the distribution of vocoids [1]. Sonority rises (e.g., /br/) have the longest lags and the highest percentage of vocoid occurrence, while plateaus (e.g., /bg/) have shorter lags, and fewer vocoids, and falls (e.g., /rb/) have the shortest lags and fewest reported vocoids.

The primary goal of this study is to determine whether these vocoids are epenthetic vowels or rather are artifacts of the vocal tract configuration during the open transition between consonants, i.e., during the lag between consonant constrictions. For this purpose, we compare the vocoids to phonological vowels acoustically, in terms of formant structure and acoustic duration. In parallel, we show kinematic data that illustrate the position of major articulators during the open transitions between consonants.

Our second goal is to explore syllable-level motivations for the sonority shape-based differences in the distribution of vocoids, and consequently in the duration of the lag between consonants as well. Syllable-level motivations originate in previous research [1][3], which has shown that constriction lag occurs in all clusters in Georgian, regardless of sonority shape or place of articulation [1][3] (see also [4] on place of articulation). These patterns thus strongly suggest that segment-level perceptual concerns do not motivate differences in degree of lag, as previously proposed [4]. Instead a different proposal is put forward in [1][3], according to which the modulation of consonant-consonant timing in Georgian is due to syllable-level planning: Speakers plan for syllables to have a single, vocalic nucleus, and in onsets with initial approximants (i.e., sonority falls) the sequence of approximant+vocoid could result in an additional large peak in the amplitude envelope that could be reinterpreted as a syllable nucleus, and could lead to syllable misparsing. Here we take a first step towards assessing this hypothesis more directly by examining the structure of the syllable-scale amplitude envelope across syllables with and without audible vocoids, and of all sonority shapes. We use the amplitude envelope, because it has been shown to relate to speech comprehension [5], and thus differences in the amplitude envelope in vocoid-containing syllables of different sonority shapes could reflect different possible syllabification parsings by the listener.

2. METHODS

2.1. Participants and procedure

Three native speakers of Georgian in their twenties (S1, S2, and S3) participated in the experiment. They were recruited through an announcement circulated by a local Georgian cultural association.

The data analysed here were collected as part of a larger series of Electromagnetic Articulography (EMA) experiments that included the simultaneous collection of audio data. EMA sensors were attached at three points along the midsagittal line of the tongue: one on the tongue tip (TT), one on the tongue dorsum (TD), and one on the midpoint between TT and TD, referred to as TB. Two sensors were attached on the lips: one on the upper lip and one on the lower. The Euclidean distance between these sensors at each timepoint was calculated and the resulting variable—Lip Aperture (LA)—was used for labial segments. Sensors were also attached on the upper and lower incisor for reference and jaw movement respectively, and on the bridge of the nose and behind each ear for head correction. Audio was recorded with a Shure SCM262 microphone mixer at a 16 kHz sampling rate, with a Sennheiser shotgun microphone positioned a foot away from the participant’s mouth. Kinematic data was automatically synchronized with the external audio data.

2.2. Stimuli

Stimuli were presented in Georgian orthography on a computer screen approximately four to five feet from the EMA. Test words appeared first in isolation and then in a frame sentence, without the word in isolation disappearing from the screen. Two frame sentences were used. The sentence was changed following S1’s participation in order to facilitate kinematic labelling.

(1) _____. kal-ma _____ mo-m-ts’er-a. [S2 & S3]
_____. woman.ERG ____ PRVB-1SG.OBJ-write-3SG.SUBJ.AOR
_____. The woman wrote _____ to me.

(2) _____. k’idev _____ v-tkv-i. [S1]
_____. again ____ 1SG.SUBJ-say-AOR.
_____. I said ______ again.

Tables 1 and 2 show the test words analysed in the current study. As with the frame sentence, the test words were refined and expanded following S1’s participation. Test words manipulate both sonority shape and order of place of articulation of the consonants in the onset. Order of place of articulation was included as a factor because it has been demonstrated to affect gestural overlap or lag in Georgian [4].

	Front to back	Back to front
Sonority rise	**br**elo ‘chaff’ **p’l**edi ‘rug’	**tm**aze ‘hair.in’ **dm**anisi ‘Dmanisi (town in Georgia)’
Sonority plateau	**mn**axe ‘see me’ **bk’**ich’i ‘raisin’ **bg**era ‘sound’	**t’b**aze ‘lake.in’ **k’b**ili ‘tooth’ **k’b**ena ‘sting’
Sonority fall	**mt**aze ‘mountain.in’ **md**are ‘worthless’	**rb**ena ‘running (n)’ **lp’**eba ‘decaying (n)’

Table 1: Test words for S1

	Front to back	Back to front
Sonority rise	**br**egi ‘mound’ **bn**eli ‘darkness’ **bn**eda ‘epilepsy’ **bl**epi ‘bluff’ **ml**ode ‘waiting’	**tm**aze ‘hair.in’ **dm**anisi ‘Dmanisi (town in Georgia)’ **gr**evi ‘gift’ **gl**exi ‘peasant’
Sonority plateau	**ʃx**ama ‘poison’ **bg**era ‘sound’ **pt**ila ‘hair lock’ **mn**axe ‘see me’	**xʃ**avs ‘shut off’ **gd**eba ‘lie about’ **tb**eba ‘warm up’
Sonority fall	**mt**aze ‘mountain.in’ **md**are ‘worthless’ **rg**eba ‘benefitting’	**rb**ena ‘running (n)’ **lb**eba ‘softening (n)’ **lm**oba ‘feeling sadness’

Table 2: Test words for S2 and S3

Stimuli were randomized and included test words for other experiments alongside those reported here. S1 and S2 recorded each word eight times, while S3 recorded 7 due to time constraints. The underlined words are those that were analysed for the current study. They were selected because a) at least one production had an intrusive vocoid and b) because the vowel following the cluster of interest (bolded in Tables 1 and 2) is either /e/ or /a/. The other phonological vowels of Georgian (/o/, /i/, /u/) are far less frequent in the test words, so in the interest of having a balanced data set, words with those vowels were excluded. Only productions with vocoids were considered in this study. There were a total of 240 vocoid-vowel tautosyllabic pairs across all words, speakers, and phrasal positions (isolation vs. quotative). 58 vocoids precede /a/ and 82 precede /e/.

2.3. Analysis

2.3.1. Acoustic analysis

All vocoids and vowels were segmented in Praat [6]. We define a post-release vocalism as a vocoid, if it has both a voicing bar and a distinct formant structure (see Fig. 3). Onset and offset of both vocoids and vowels were determined by the presence of f2. Formant values at the midpoint of the vowel or vocoid, and acoustic duration, were automatically extracted using a Praat script [7]. Values were checked by hand and corrected where necessary.

2.3.2 *Articulatory analysis*

Articulatory analysis of consonant (C) constriction lag was done semi-automatically using Mview (Mark Tiede, Haskins Laboratories). EMA sensor trajectories were used to label the C constrictions for each onset cluster. Lip aperture was used for labial consonants, tongue tip vertical displacement for coronal, and tongue dorsum displacement for dorsal. Relevant timepoints for the constrictions were labelled using velocity criteria as shown in Figure 1.

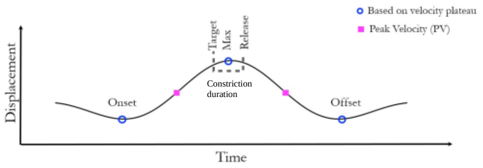

Figure 1: Timepoints labelled for each gesture

The lag between C constrictions is calculated as the time between the first consonant's release and the second consonant's target.

2.3.3 *Amplitude envelope analysis*

The amplitude envelope was extracted following Tilsen and Arvaniti [8] using a Matlab script [9].

2.3.4 *Statistical analysis*

All statistical analysis was done in R [10] with linear mixed effects models [11]. Minimal adequate fixed effects structures were determined using drop1 [10] and minimal adequate random effects structures using rePCA [12]. Post-hoc pairwise comparisons with a Holm correction were done using emmeans [13]. Graphics were made with ggplot2 [14].

3. RESULTS

3.1 Distribution

The distribution of vocoids in the data are not random. Of the 240 vocoids measured, 149 appear in sonority rises, 62 in sonority plateaus, and 29 in falls.

3.2 Duration

Vocoids are significantly shorter than vowels (p<.001). Mean vocoid duration overall is 36.6ms, while mean vowel duration overall is 118.3ms Phrasal position is also highly significant, with both vocoids and vowels produced in the isolation condition being longer than those in the quotative condition (p<.001). Neither sonority shape nor order of place of articulation have an effect on vocoid or vowel duration.

Sonority shape does, however, significantly affect the duration of the lag between consonant constrictions [2]. Sonority falls have shorter lags than rises (p<.001). Across all sonority conditions, lags are on average 100ms, nearly equivalent to the duration of a full vowel. As with vocoid duration, phrasal position significantly affects lag: lags are longer in the isolation condition (p<.001). However, constriction lag duration and vocoid duration are not correlated.

Figure 3 shows the spectrogram and EMA sensor trajectories for /bneda/ with an intrusive vocoid. The colored boxes mark the gestures for /b/, /n/, /e/. The filled boxes represent the constriction plateaus. The horizontal arrow shows the lag between consonant constrictions, while the dotted vertical lines delimit the vocoid. We can see from this example that not only does the vocoid occupy only a fraction of the lag, but also that the tongue dorsum (TD) is not moving towards an articulatory target during the vocoid, instead it is in the process of moving along with the tongue tip (TT) to the following /n/ target.

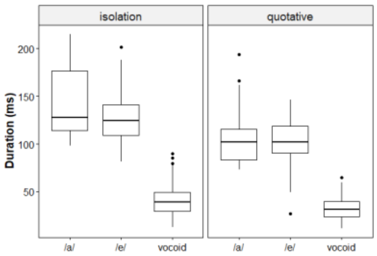

Figure 2: Durations (ms) for vocoids and vowels in the isolation (left) and quotative (right) conditions.

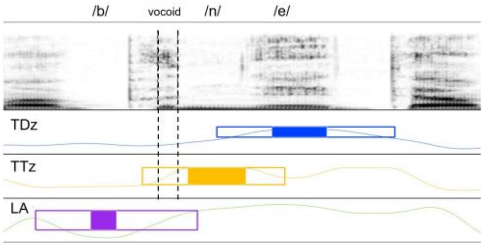

Figure 3: Spectrogram and sensor trajectories for /bneda/, with vocoid (dotted lines) and constriction lag (arrow) marked. TDz stands for tongue dorsum vertical displacement; TTz for tongue tip vertical displacement; LA for lip aperture. Boxes on each trajectory mark the target (filled box) and movement into and out of the target (empty boxes).

3.3 Formants

Vocoids are distinct from both /a/ and /e/ in both first and second formants (p<.001). Moreover, the quality of the following vowel does not affect the formant values of the vocoid. Figure 4 shows the f1 by f2 plot. Three clouds are visible: one for the /a/ vowel, one for the /e/ vowels and one for the vocoids, regardless of the quality of the following vowel.

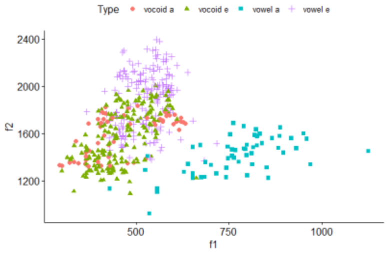

Figure 4: f1 and f2 values for /a/, /e/, and vocoids preceding either /a/ or /e/.

3.4 Amplitude envelope

Figure 5 shows the waveform and corresponding amplitude envelope per sonority shape.

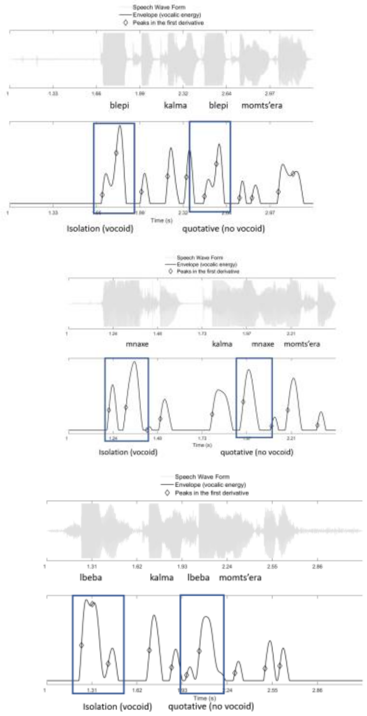

Figure 5: Waveform and amplitude envelope for sonority rise /blepi/ (5a), sonority plateau /mnaxe/ (5b), and sonority fall /lbeba/ (5c) with and without a vocoid, as produced by S2. Regions of interest are outlined; the box on the left of each pair marks the production with a vocoid, and the box on the right marks the one without.

In these examples, isolation productions have a vocoid, while the quotative ones do not. In sonority rises (Fig. 5a), the vocoid and non-vocoid productions have the same amplitude envelope shape, with the same number of amplitude peaks identified. The vocoid peak is smaller in magnitude and peak velocity than the vowel peak. In a sonority plateau like /mnaxe/, we see different shapes in the envelope for the vocoid and non-vocoid productions (Fig. 5b). The production with a vocoid has two distinct peaks, while the non-vocoid production has one. We see something similar for the sonority fall /lbeba/ in Figure 5c. The production with the vocoid has three peaks in the first derivative identified, while the non-vocoid production has only two. The first peak in the vocoid production, which coincides with the vocoid itself, is greater in magnitude than many lexical vowels in the utterance.

4. DISCUSSION

These results indicate that intrusive vocoids in Georgian are articulatory artifacts that emerge due to the long lag between consonant constrictions. According to [3], this lag is created because C constriction gestures in Georgian onsets are in anti-phase coordination with each other, with the prenuclear C gesture also anti-phase with the nucleus V gesture. This means that each successive gesture is initiated only after the previous gesture has reached its target. As articulators move from one C to the other, a relatively neutral configuration is created in the vocal tract. When the second consonant is voiced, voicing may occur during some part of this transition resulting in the vocoid, as illustrated in Fig. 3. Since the overlap between lag and voicing is brief and the oral tract is target-less during this interval—hence the non-effect of sonority on vocoid quality—the resulting vocoid has a schwa-like quality (see Fig. 4; F1 ~ 500 Hz and F2 ~ 1500 Hz for the vocoids). The fact that the quality of the vocoid is unaffected by the following vowel is consistent with the account of the nucleus V gesture being anti-phase with the onset. If the vowel gesture were in-phase with the onset, the lag between consonant gestures would briefly ‘unmask’ the V gesture, and the formant structure of the resultant vocoid would be expected to be affected by the quality of the phonological vowel.

Sonority shape does affect the duration of the lag and the distribution of vocoids. It has been hypothesized that this may avoid resyllabification [3]. Preliminary examination of the syllable-scale amplitude envelope supports this hypothesis, since it shows that the presence of a vocoid in a sonority plateau or fall (Figures 5b-c) results in a distinct additional peak, which could be interpreted as an additional syllable. Conversely, in a sonority rise, the vocoid is less distinct from the nuclear amplitude peak (Figure 5a). The distributional asymmetry of intrusive vocoids across onsets of different sonority shapes thus mirrors the potential disruptiveness of the vocoid. In those onsets where a vocoid may result in resyllabification, we find fewer vocoids and shorter lags between consonants, as speakers plan for a syllable with a single nucleus.

5. ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation (#1551428) to Argyro Katsika (PI) and National Institutes of Health (NIDCD-DC-002717) to Douglas Whalen (PI).

6. REFERENCES

[1] C. Crouch, A. Katsika, and I. Chitoran, “Sonority sequencing and its relationship to articulatory timing in Georgian,” Journal of the International Phonetic Association, pp. 1–24, 2023.

[2] M. Pouplier et al., “Language and cluster-specific effects in the timing of onset consonant sequences in seven languages,” Journal of Phonetics, vol. 93, p. 101153, Jul. 2022, doi: 10.1016/j.wocn.2022.101153.

[5] Y. Oganian and E. F. Chang, “A speech envelope landmark for syllable encoding in human superior temporal gyrus,” Science Advances, vol. 5, no. 11, p. eaay6279, Nov. 2019, doi: 10.1126/sciadv.aay6279.

[3] C. Crouch, “Postcards from the syllable edge: sonority and articulatory timing in complex onsets in Georgian,” UC Santa Barbara, 2022. Accessed: Jan. 04, 2023. [Online]. Available: https://escholarship.org/uc/item/5w18167d

[4] I. Chitoran, L. Goldstein, and D. Byrd, “Gestural overlap and recoverability: Articulatory evidence from Georgian,” Laboratory phonology, vol. 7, no. 4–1, pp. 419–447, 2002.

[6] P. Boersma, and D. Weenink, “Praat: doing phonetics by computer [Computer program],” Version 6.3.03, http://www.praat.org/

[7] Lennes, Mietta, “SpeCT - Speech Corpus Toolkit for Praat (v1.0.0). First release on GitHub”. Zenodo, Mar. 08, 2017. doi: 10.5281/zenodo.375923.

[8] S. Tilsen and A. Arvaniti, “Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages,” The Journal of the Acoustical Society of America, vol. 134, no. 1, pp. 628–639, 2013.

[9] A. MacIntyre, “AcousticLandmarks” Accessed: 4 January 2023 [Matlab script] https://github.com/alexisdmacintyre/AcousticLandmarks

[10] R Core Team, “R: A language and environment for statistical computing.” R foundation for statistical computing Vienna, Austria, 2018.

[11] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, “lmerTest Package: Tests in Linear Mixed Effects Models,” Journal of Statistical Software, vol. 82, no. 13, pp. 1–26, 2017, doi: 10.18637/jss.v082.i13.

[12] D. Bates, M. Mächler, B. Bolker, and S. Walker, “Fitting Linear Mixed-Effects Models Using lme4,” Journal of Statistical Software, vol. 67, no. 1, pp. 1–48, 2015, doi: 10.18637/jss.v067.i01.

[13]R. Lenth, J. Love, and M. Herve, “emmeans: Estimated Marginal Means, aka Least-Squares Means.” Jan. 10, 2018. Accessed: Jan. 04, 2023. [Online]. Available: https://CRAN.R-project.org/package=emmeans

[14] H. Wickham, ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016. [Online]. Available: https://ggplot2.tidyverse.org