Effects of ejective stops on preceding vowel duration

Gašper Beguš[a)]
Department of Linguistics, Harvard University, Cambridge, Massachusetts 02138, USA

(Received 13 April 2017; revised 8 September 2017; accepted 29 September 2017; published online 19 October 2017)

One of the most widely studied observations in linguistic phonetics is that, all else being equal, vowels are longer before voiced than before voiceless obstruents. The causes of this phonetic generalization are, however, poorly understood and several competing explanations have been proposed. No studies have so far measured vowel duration before stops with yet another laryngeal feature: ejectives. This study fills this gap and presents results from an experiment that measures vowel duration before stops with all three laryngeal features in Georgian and models effects of both closure and voice onset time (VOT) on preceding vowel duration at the same time. The results show that vowels have significantly different durations before all three series of stops, voiced, ejective, and voiceless aspirated, even when closure and VOT durations are controlled for. The results also suggest that closure and VOT durations are inversely correlated with preceding vowel duration. These results combined bear several implications for the discussion of causes of vowel duration differences: the data support the hypotheses that claim that laryngeal gestures, temporal compensation, and closure velocity affect vowel duration. Some explanations, especially perceptual and airflow expenditure explanations, are considerably weakened by the results.
 https://doi.org/10.1121/1.5007728

[ZZ] Pages: 2168–2184

I. INTRODUCTION

That vowels are phonetically longer before voiced obstruents than before their voiceless counterparts, especially in coda position, is a well-established phonetic generalization. The lengthening effect that voiced obstruents have on preceding vowel duration has been reported in various languages across language families: English, French, Russian, Korean (Peterson and Lehiste, 1960; House, 1961; Chen, 1970; Luce and Charles-Luce, 1985; Abdelli-Beruh, 2004; de Jong, 2004), German, Swedish, Icelandic (Port, 1996), Hindi (Maddieson and Gandour, 1976; Lampp and Reklis, 2004; Durvasula and Luo, 2014), Arabic (Port *et al.*, 1980; Jong and Zawaydeh, 2002), Bengali, Hungarian, Italian, Norwegian, Spanish, Danish, Persian, and Dutch (reported with references in Maddieson and Gandour, 1976; Kluender *et al.*, 1988), among others. Following Durvasula and Luo (2014) and other studies, the term *voicing effect* is used to describe this phonetic generalization.

While most studies on the voicing effect focus on vowels in closed syllables, the voicing effect is also reported in open syllables in many languages: among others, for example, in English, Korean, French, Arabic, Spanish, and Norwegian (Port, 1981; Chen, 1970; Abdelli-Beruh, 2004; Port *et al.*, 1980; Fintoft, 1961; Zimmerman and Sapon, 1958). In the absence of evidence to the contrary, it is assumed that the same general mechanism is responsible for vocalic durational differences in both open and closed syllables. The magnitude of the voicing effect, however, *can* differ according to whether the affecting consonant is in coda position (closed syllable) or in the onset position of the following syllable (open syllable). Port (1981) measures vowel durations in English one-, two-, and three-syllable words with the voiced-voiceless consonant pairs in coda and onset positions. The voicing effect in English is considerably greater in monosyllables in which the affecting stop surfaces in the coda compared to di- or trisyllables in which the vowel surfaces in an open syllable before the affecting stop in the onset of the following syllable, but the difference in the conditions there is not only in position within a syllable (coda vs onset), but also in the number of syllables in the word (mono- vs di- and trisyllables). Similar results are reported for English in other studies (Klatt, 1973; Lisker, 1974; Sharf, 1962; Klatt, 1976; and the literature therein). Laeufer (1992) also reports a difference in magnitude of the voicing effect between tautosyllabic and heterosyllabic following obstruents in French using an experimental design that controls for word and syllable structure (speech rate, however, remains a potential confound). On the other hand, no significant differences between coda and onset obstruents were found in a similar experiment in French in Abdelli-Beruh (2004). To the author's knowledge, no detailed studies exist on the causes of the differences in the magnitude of the voicing effect between the coda and onset consonants. The most obvious potential causes for these differences are (i) a higher degree of coarticulatory effects/gestural overlap in tautosyllabic segments compared to heterosyllabic segments (cf. Mok, 2012; Byrd, 1996); (ii) final-lengthening effect in mono- vs polysyllabic words, and (iii) perceptual effects in languages such as English in which the voicing effect is argued to be perceptually enhanced (see discussion below): coda consonants are perceptually less salient, which is why alternative perceptual cues (e.g., vowel duration) have to be enhanced (cf. Wright, 2004). The differences in magnitude of the voicing

[a)]Electronic mail: begus@fas.harvard.edu

 0001-4966/2017/142(4)/2168/17/$30.00

effect between open and closed syllables, however, are beyond the scope of this paper. The present study focuses on open non-final syllables in order to avoid the potential influences of increased coarticulatory or perceptual effects or final lengthening. The open syllable position also allows for more accurate measurements of voice onset time (VOT).

The magnitude of the voicing effect also differs across languages (Chen, 1970; Ohala and Ohala, 1992; Cho, 2015; Laeufer, 1992; Zimmerman and Sapon, 1958). Variation in the magnitude of the effect has led to proposals that claim that vowel duration differences can be either automatic mechanical results of articulatory origin or part of the active phonology, i.e., actively controlled by speakers (Solé, 2007; Cho 2015). English in this respect stands out as featuring an "exaggerated" voicing effect of comparatively great magnitude that is likely phonologized (de Jong 1991, 2004; Solé 2007; Cho, 2015). For most other languages, voicing effect is assumed to originate in some articulatory/physiological or perceptual mechanism that is not part of the phonological grammar, i.e., not actively controlled by speakers [Chen, 1970; Laeufer, 1992; Cho, 2015 and proposals in (1) below].

It has even been proposed that some languages have no voicing effect, i.e., that voicing effect is not a phonetic universal. Keating (1985) reports that vowel duration in medial syllables in Polish and Czech do not differ significantly before a voiced and a voiceless stop. Crucially, in both languages vowels are still longer before voiced than before voiceless stops, but the difference does not reach statistical significance—it is therefore possible that the lack of significance results from lower statistical power, especially because the durations are measured in open syllables where smaller differences are expected. To the author's knowledge, no languages have been reported in which the effect operates in the opposite direction: in which vowels are phonetically shorter before voiced stops compared to voiceless stops.

Voicing effect is thus a robustly attested universal phonetic tendency and while its existence is well-documented, its causes are poorly understood. Several competing proposals have been offered in the literature and there is almost no consensus reached on the matter. Chen (1970), Lisker (1974), Kluender *et al.* (1988), Maddieson (1997), and Durvasula and Luo (2014) summarize the proposals (terminology is from Chen, 1970). Following their work, the explanations can be categorized as follows:

(1) Different proposals for causes of the voicing effect

(a) **Articulatory Energy Expenditure**: Voiceless stops are articulated with greater "physiological" force ("fortis"): "anticipation of greater force" shortens the preceding vowel (Belasco, 1953). In a related proposal, if energy expenditure is constant across the syllable, vowels before voiceless stops that require more energy will be shorter (Meyer, 1903; reported in Lindblom, 1967).

(b) **Compensatory Temporal Adjustment**: Somewhat related to [1(a)], if the timing of syllables/vowel-consonant (VC) sequences tends to be constant, then vowels before voiced stops (which have intrinsically shorter closure and VOT) should be longer and vice versa via "compensatory temporal reorganization of sequential motor commands" (Chen, 1970; Kozhevnikov and Chistovich, 1967; Lindblom, 1967; Port, 1981; de Jong, 1991; Port *et al.*, 1980, 1987).

(c) **Rate of Closure Transition**: Voiceless stops require greater velocity and force of closure due to higher intraoral pressure during their production. Greater force and velocity of closure gesture results in shorter preceding vowel duration (Chen, 1970; Öhman, 1967).

(d) **Laryngeal Adjustment**: Maintaining voicing during closure requires "drastic adjustments in vocal-fold positioning." These adjustments require a longer time interval, which results in longer vowel duration before voiced stops (Halle *et al.*, 1967; Chomsky and Halle, 1968).

(e) **Perceptual Distance**: Vowel duration is exploited by the speaker to maximize one of the main perceptual cues for voicing of the following stop: closure duration. Short closure is perceived as even shorter after a longer vowel (Denes, 1955; Lisker, 1957; Javkin, 1975; Kluender *et al.*, 1988).

Many of these explanations and models extend beyond explaining the differences in vowel duration before voiced–voiceless pairs of obstruents and bear predictions for more general timing effects in VC sequences. For example, the Compensatory Temporal Adjustment hypothesis [1(b)] predicts that vowel duration will be inversely correlated with the duration of the following segment of any kind (for a quantitative model of vowel duration as a function of several parameters, such as voicing, vowel quality, and number of syllables in the word, see Klatt, 1976; Port, 1981). The Perceptual Distance hypothesis [1(e)], for example, predicts that for any segment with salient durational cues, the duration of the preceding vowels would tend to be employed as a perceptual cue, causing an inverse correlation between the durations. Most studies on vowel duration, however, measure vowels before voiced vs voiceless obstruents only.

One of the reasons for the existence of so many competing proposals is precisely the paucity of studies that measure vowel duration before consonants with other laryngeal features and the paucity of studies that model the effects of laryngeal features and closure duration together. Due to the lack of information on these relationships (see the paragraph below), the voice feature is thus assumed to be the only laryngeal feature responsible for differences in vowel duration. More importantly, closure and VOT duration and voice are universally correlated: voiced stops have universally shorter closure and VOT durations (Lisker, 1957; Port, 1981; Luce and Charles-Luce, 1985). In the absence of information on the effects of other laryngeal features and of closure and VOT duration (when laryngeal features are controlled for), it is impossible to discriminate between the three correlated parameters and consequently between different proposals.

Only a subset of studies test vowel duration differences before other series of stops. The effect of another laryngeal feature, [±spread glottis], has been tested in Maddieson and Gandour (1976), Ohala and Ohala (1992), Lampp and Reklis (2004), and Durvasula and Luo (2014). These studies, however, yield somewhat inconsistent results as to whether vowels are longer or shorter before following aspirated obstruents. Most studies argue that vowels are longer before aspirated than before unaspirated stops (Maddieson and Gandour, 1976; Lampp and Reklis, 2004; Durvasula and Luo, 2014). Durvasula and Luo (2014) label this generalization the *aspiration effect*. Ohala and Ohala (1992), on the other hand, argue that there is no consistent pattern: aspiration lengthens the vowel for some places of articulation and shortens it or has no effect for other places. No statistical tests are performed on their results. The inconsistent results are all the more intriguing because all of the studies were performed on speakers of the same language: Hindi. Another confounding factor of these studies is that the experiments included a relatively small number of speakers: from one speaker in Maddieson and Gandour (1976), to three in Ohala and Ohala (1992), and up to seven in Durvasula and Luo (2014). To the author's knowledge, only two studies model closure/consonant duration (but not VOT) together with the voicing and aspiration effects (de Jong, 1991; Durvasula and Luo, 2014). Both studies argue that voicing and aspiration are significant predictors even when closure/consonant duration is modeled, but they yield opposing results on the direction of the effect of the closure duration.

The present study focuses on laryngeal features and two durational properties of stops, closure and VOT, precisely because several proposals in (1) crucially rely on the inverse correlation between closure and preceding vowel duration, or on the effect of laryngeal features on preceding vowel duration. To the author's knowledge, no information about vowel duration before the laryngeal feature [±constricted glottis] is currently available. Yu's (2008) study measures vowel durations in Washo, but no explicit results or statistical tests are mentioned with respect to phonetic vowel length before ejective consonants (Yu, 2008 only reports C/V ratios). The author is unaware of any studies that measure vowel duration before ejective stops compared to the voiced or voiceless series (despite the fact that other acoustic properties of ejective obstruents, such as VOT, closure duration, burst spectra, $F0$, phonation of the following vowel, etc., have received extensive treatments, e.g., Lindau, 1984; Wright *et al.*, 2002; Vicenik, 2010; Grawunder *et al.*, 2010; and literature therein). This paper aims to fill this gap. It presents a phonetic study of vowel duration before stops in Georgian, a Kartvelian language that distinguishes three series of stops: voiced, ejective, and voiceless aspirated. The present study measures vowel duration before stops with all three laryngeal features and also models closure and VOT duration together with the effects of laryngeal features. To the author's knowledge, only one other study (Durvasula and Luo, 2014) has thus far modeled the effects of closure duration together with the effects of aspiration, but no studies exist that model the duration of both closure and VOT. With 12 experimental participants, this study is also one of the largest with respect to the number of speakers and tokens recorded.

The results of the experiment show that vowel durations are indeed significantly different before ejective stops compared to positions before voiced and voiceless aspirated stops: they are shorter before ejectives than before voiced stops, and longer before ejectives than before voiceless aspirated stops, even when closure and VOT durations are controlled for. In addition, modeling two durational properties (closure duration and VOT) together with the effects of laryngeal features allows for comparison between different predictors: the study shows that laryngeal features are the most consistent predictors, but that closure and VOT durations also affect preceding vowel duration and that closure duration is more negatively correlated with vowel duration in voiced stops.

The results of the experiment bear several implications for the discussion on causes of vocalic durational differences and provide crucial evidence in favor and against the different proposals in (1). The Laryngeal Adjustment [1(d)], Compensatory Temporal Adjustment [1(b)] and Rate of Closure Transition [1(c)] hypotheses find support in the results, while on the other hand, the Perceptual Distance [1(e)] and Airflow Expenditure hypotheses (discussed in Sec. V D) are considerably weakened by the results. The paper also suggests a potential articulatory mechanism for the Laryngeal Adjustment hypothesis.

II. EXPERIMENT

A. Design

To measure vowel duration before ejective stops in Georgian, V_xC_y sequences were created using three independent variables: (i) vowel quality (Vowel), (ii) place of articulation of the stop (Place), and (iii) laryngeal features of the stop (LF). Georgian features a five-vowel system with no length opposition ([ɑ], [ɛ], [ɔ], [i], [u]; Shosted and Chikovani, 2006), but the vowel quality variable in this experiment included only three vowels: low back [ɑ], mid front [ɛ], and mid back [ɔ]. The place of articulation variable also included three levels: labials, dentals, and velars. Other places were omitted because they do not feature a full three-way opposition in laryngeal features. As previously mentioned, the experiment tested vowel duration before stops with all three laryngeal features: ejective, voiceless aspirated, and voiced (for a detailed description of the phonetic properties of Georgian stops, see Vicenik, 2010). Each of the three independent variables had three levels, resulting in $3 \times 3 \times 3 = 27$ combinations of V_xC_y sequences tested. Table I lists all 27 sequences tested.

Each of the 27 V_xC_y-sequences were embedded in each of the 25 frames, which together with the sequences formed a balanced set of 675 (27×25) nonce words of the structure $CVCV_xC_yV$. In other words, each V_xC_y-sequence was tested in all 25 frames, so that if a particular frame had a coarticulatory effect on vowel duration, the effect would influence all three laryngeal features tested equally (and therefore not influence the overall effect of laryngeal features itself). The frames were created such that the measured vowel V_x would surface in the second syllable of a trisyllabic word in order to avoid any effects of main stress, which in Georgian falls on the initial syllable (Vicenik and Jun, 2014). Moreover,

TABLE I. List of V_xC_y sequences tested in the experiment.

	Bilabial	Dental	Velar
Voiced	ɑb	ɑd	ɑg
Voiceless aspirated	ɑp^h	ɑt^h	ɑk^h
Ejective	ɑp'	ɑt'	ɑk'
Voiced	ɛb	ɛd	ɛg
Voiceless aspirated	ɛp^h	ɛt^h	ɛk^h
Ejective	ɛp'	ɛt'	ɛk'
Voiced	ɔb	ɔd	ɔg
Voiceless aspirated	ɔp^h	ɔt^h	ɔk^h
Ejective	ɔp'	ɔt'	ɔk'

the frames were formed so that the vowel tested would surface in an open syllable (see discussion in Sec. I).

The 25 frames of the shape C_1VC_2-V to which the V_xC_y sequences were inserted were designed so that in 16 cases C_2 was either an [x], [ʃ], or [s]. The fricatives were chosen in order to maximize the accuracy of vowel onset annotation (cf. Turk *et al.*, 2006). In the remaining nine cases, C_2 was composed of stops with each of the three places of articulation and laryngeal features (3 places $\times$ 3 laryngeal features). C_1 consisted of one of the following consonants: [s], [ʃ], [z], [ʒ], [x], [m], [n], [l], or [v]. Note again that each V_xC_y sequence was inserted into each C_1VC_2-V frame in order to balance the design. Altogether 675 (27 $\times$ 25) nonce words were created with no repetitions. In such a large inventory it is almost inevitable that a subset of created words are real words in Georgian: a speaker reported six real words (0.9% of the total 675). Such a small proportion of real words is not expected to alter the results in any significant way.

The nonce words were embedded in a carrier phrase *maiam tkva X ara* "Maya said X, right," where X is the nonce word. The order of the sentences was randomized and printed in Georgian script on sheets of paper.

B. Participants

Twelve speakers of Georgian participated in the experiment: eight females and four males. The mean age of the speakers was 23.5 yr (median = 20, standard deviation = 6.5, range = 18–35). Eleven speakers were from Tbilisi, and one speaker was from Batumi. Ten participants reported that they speak Georgian at home, and the other two reported speaking Georgian at home when in Georgia. All speakers reported proficiency in at least two other languages, mostly English and Russian. All speakers lived in the United States at the time of recording; their stay in the United States ranged from 6 to 84 months (mean = 27, median = 15, standard deviation = 28).

C. Procedure

Speakers were instructed to read the same list of 675 sentences with nonce words and were asked to keep approximately the same pace of reading throughout the experiment. Speakers were also asked to repeat the whole sentence if they thought they had made a mistake. Speakers read from printed sheets of paper; each sheet included seven sentences for a total of 97 pages. A short break was taken after every 10 pages (70 sentences).

The recordings were made at the Department of Linguistics, Harvard University in Cambridge, MA, in a sound-attenuated booth with a Shure KSM 27 cardioid condenser microphone (20–20 000 Hz with 18 dB-per-octave cutoff at 80 Hz) and a Focusrite Scarlett 2i2 pre-amplifier. Recordings were made in the Praat recording device (Boersma and Weenink, 2015) with a 48 kHz sampling rate in 16-bit .wav format.

On average, the total participation time per participant was approximately 2 h. Prior to the recordings, speakers were asked to read and sign the informed consent form and fill out a brief demographics questionnaire. The average recording time excluding breaks per participant was approximately 40 min. Speakers were paid $20/h for participation.

D. Data analysis

Eight undergraduate research assistants and the author annotated the data. All tokens of each speaker were annotated by a single annotator: for no one speaker were there two different annotators in order to ensure within-speaker consistency. Only three annotators annotated more than one speaker. The research assistants were not given the exact purpose of the experiment in order to prevent potential biases in annotations. Vowel duration, closure duration, and VOT were measured in Praat and extracted with a Praat script written by Mietta Lennes (Lennes, 2002).

Vowel onsets were annotated at the beginning of periodic vibration with clear formant structure. Vowel offsets were correspondingly annotated at the end of periodic vibration with clear formant structure and with changes of waveform amplitude (cf. Turk *et al.*, 2006). Closure duration was annotated from the vowel offset to the oral burst of the stop. VOT was annotated from the onset of oral burst to the first periodic vibration with clear formant structure (cf. Turk *et al.*, 2006). Individual annotators had slightly different criteria for annotating vowel onset and offset, but were consistent in their judgments throughout the annotations. As will be shown in Sec. III A, differences between annotators were not significant and relatively minor compared to the observed phonetic effects. The reference work for vowel onset and offset annotation was Chap. 4.5 ("Waveforms and Measurement of Duration") in Ladefoged (2003).

As mentioned above, speakers were encouraged to repeat sentences if they thought they had made a mistake. If the second repetition was correct, the token was measured and included in the analysis. Several tokens, however, were misread or were read with pauses, interruptions or stumbles and were not repeated by the speaker. Such tokens were removed from the database based on auditory analysis. During the initial exploratory data examination, extreme points were also manually examined. If any clear mistakes were detected or if the tokens had signs of stumbling, they were removed. If no obvious mistakes were detected, the tokens were left for analysis. It is possible that some tokens with reading mistakes or stumbles or some mislabeled tokens remain in the analyzed corpus. However, the number of such mistakes is likely small

TABLE II. Sample means and standard deviations of vowel duration (in ms) before ejective, voiced, and voiceless aspirated stops across all 12 speakers.

Place	Vowel	Ejective	SD	Voiced	SD	Voiceless	SD
Labial	e	81.3	21.2	89.2	24.0	76.9	20.9
	o	80.1	21.6	88.8	23.1	74.6	19.7
	a	86.8	22.0	95.4	23.7	82.7	20.3
Dental	e	88.7	22.6	96.9	25.1	84.3	21.7
	o	87.4	22.1	96.6	23.6	82.2	21.3
	a	95.4	22.4	103.5	24.1	91.6	22.8
Velar	e	85.9	23.2	97.1	26.1	81.0	22.1
	o	82.6	21.7	93.5	24.7	77.5	21.4
	a	95.2	23.3	105.5	24.8	89.8	22.4
Acr. pl. & vow.		87.1	22.8	96.3	24.9	82.3	22.1

relative to the large number of total tokens and is not expected to alter the results crucially. No outliers were removed in the process of data analysis.

Each speaker was instructed to read 675 tokens. Altogether, the 12 speakers read 8100 tokens. After the removal of reading mistakes and stumbles, 7917 tokens remained for the analysis. Thus, 183 or 2.3% of tokens in total were excluded from the data analysis. For any individual speaker the highest percentage of excluded tokens was 4.9%.

III. RESULTS

The mean vowel duration for all places of articulation and for all three vowels tested is consistently longest before voiced stops, shorter before ejectives, and shortest before voiceless aspirated stops. Table II summarizes raw sample mean durations with standard deviations across the combinations of V_xC_y sequences. The mean difference across all speakers, places of articulation, and vowels between ejective and voiceless is 4.8 ms, and the mean difference between ejective and voiced is 9.2 ms.

This distribution is also consistent across all 12 speakers, as shown in Fig. 1: vowels are longest before voiced stops, shorter before ejectives, and shortest before voiceless aspirated stops.

A. Simple model

To test the statistical significance of the mean differences, the data were fit to a linear mixed effects model using the *lmer()* function in *lme4* package (Bates *et al.*, 2015). The simple model that tests vowel duration before ejective stops includes only vowel duration as the dependent variable and Laryngeal Features (LF; with three levels: ejective, voiceless aspirated, voiced) as the independent variable of interest. Two additional control independent variables were modeled: Vowel and Place. Laryngeal features were treatment-coded with ejective as the reference level. The two control variables were sum-coded with velar and vowel [ɑ] as reference levels.

The best-fitting model was selected through a step-wise backward model selection technique, starting with a full model with all three predictors, all interactions, and random intercepts for Speaker and Frame and random slopes for LF for both random intercepts. The random effect structure was chosen with step-wise removal of random slopes and intercepts from models with full fixed effects structure based on Akaike Information Criterion (AIC) (fitted with Restricted maximum likelihood, according to Zuur *et al.*, 2009). Random slopes for control predictors and random slopes for interactions with control predictors were not tested. The final model includes crossed random intercepts for Speaker with a by-speaker random slope for LF and a random intercept for Frame. Fixed effects structure was chosen based on the Likelihood Ratio Test (LRT) on models with the chosen random effect structure and fitted with maximum likelihood. Step-wise removal of predictors was performed on a model with all interactions: each higher level interaction term is removed from the full model if it does not significantly improve the fit according to the LRT. This procedure is repeated until all interactions or main effects significantly improve the model. The best fitting fixed effects structure includes all three predictors and the interactions LF × Place and Place × Vowel. Table III summarizes the model. *P*-values and degrees of freedom are calculated using Satterthwaite approximations in the *lmerTest* package (Kuznetsova *et al.*, 2016); effects were extracted from models using the *effects* package (Fox, 2003); and Variance Inflation Factors (VIFs) were calculated using the *vif.mer()* function (Frank, 2011).

The model in Table III shows that vowels are significantly shorter before ejectives than before voiced stops and significantly longer before ejectives than before voiceless

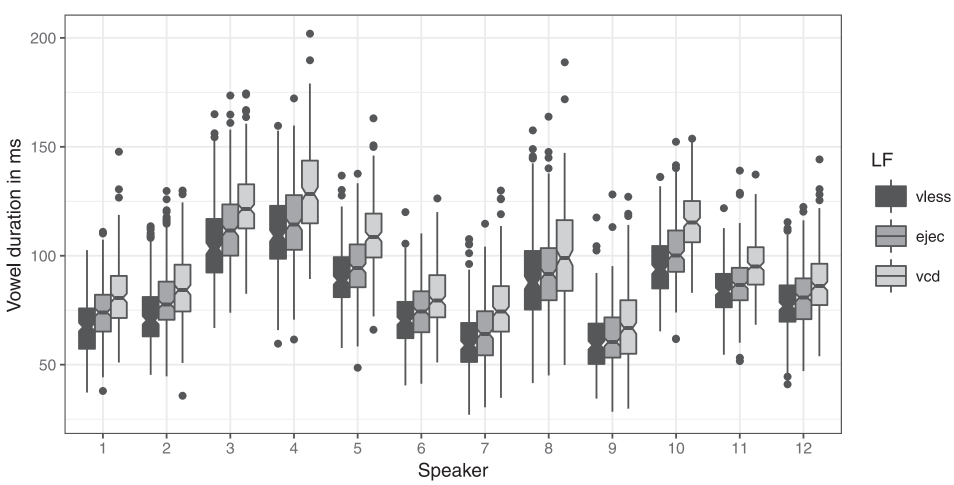

FIG. 1. Boxplot showing vowel duration in ms before stops with different laryngeal features (voiceless, ejective, voiced) in individual speakers.

TABLE III. Linear mixed effects model with three predictors: LF, Vowel, and Place.

	Estimate	SE	df	t value	Pr(>|t|)
(Intercept)	87.04	5.34	14	16.31	0.0000
ejec. vs voiced	9.28	0.88	11	10.55	0.0000
ejec. vs v'less	−4.65	0.57	11	−8.17	0.0000
mean vs lab.	−4.38	0.33	7845	−13.08	0.0000
mean vs alv.	3.45	0.33	7845	10.32	0.0000
mean vs [ɛ]	−1.77	0.19	7845	−9.19	0.0000
mean vs [ɔ]	−3.75	0.19	7845	−19.47	0.0000
voiced:lab.	−0.88	0.47	7845	−1.85	0.0637
v'less:lab.	0.14	0.47	7846	0.29	0.7690
voiced:alv.	−0.59	0.47	7845	−1.25	0.2129
v'less:alv.	0.22	0.47	7845	0.47	0.6377
lab.:[ɛ]	0.19	0.27	7845	0.69	0.4882
alv.:[ɛ]	−0.17	0.27	7845	−0.64	0.5254
lab.:[ɔ]	0.94	0.27	7845	3.45	0.0006
alv.:[ɔ]	0.63	0.27	7845	2.31	0.0208

aspirated stops. Figure 2 illustrates this difference and shows that it holds for all places of articulation. There is a significant interaction of LF × Place: the difference in vowel duration between a voiced and an ejective stop is significantly greater for velar stops compared to the mean across places of articulation ($\beta = 1.5$ ms, $t = 3.1$, df = 7845, $p < 0.01$ in a model with dental as the Place reference level). There is no significant interaction of Place and LF for the opposition ejective vs voiceless aspirated.

The 95% profile[1] confidence intervals (CIs) calculated for the coefficient ejective vs voiced is [7.5 ms, 11.1 ms]; for the coefficient ejective vs voiceless aspirated the CI is [−5.8 ms, −3.5 ms]. The difference in vowel duration before voiced vs ejective stop is thus about twice the size of the difference before voiceless aspirated vs ejective stop.

The observed differences are substantial enough that they are likely phonetically real and not crucially influenced by the annotation process. To test influences of the annotation process, two different annotators were instructed to analyze a number of tokens from a single speaker. Thus, 788 annotations were analyzed (394 identical tokens analyzed by two different annotators). The data were fit to a linear mixed

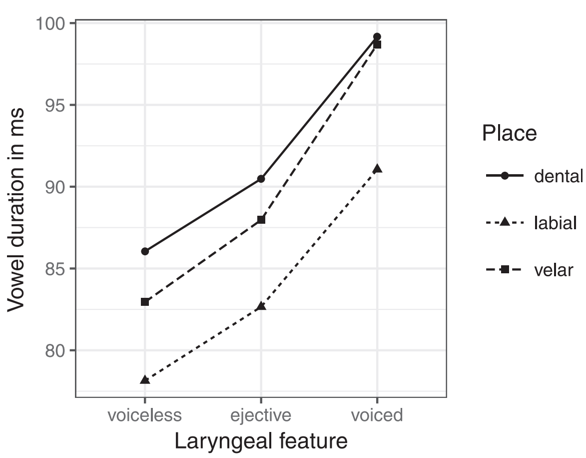

FIG. 2. Estimates of the effects of Laryngeal Features and Place of articulation on preceding vowel duration in ms (from a linear mixed effects model).

effects model with vowel duration as the dependent variable and four predictors: three phonetic predictors, LF, Place, and Vowel (coded as above), and the non-phonetic predictor Annotator (treatment-coded with the two annotators as levels). The random effect structure included a random intercept for Frame. The fixed effect structure (chosen with backwards model selection technique based on LRT) included all three phonetic predictors and interactions LF × Place and Place × Vowel. LRT thus yielded the predictor Annotator and its interactions insignificant [$\chi^2(15) = 5.5, p = 0.99$]. In order to get estimates of the Annotator effect, the final model included interaction between the Annotator and all three phonetic predictors with interactions LF × Place and Place × Vowel. Annotator is insignificant as a main effect ($\beta = -1.5$ ms, $t = -1.1$, df = 734, $p = 0.29$) and none of the interactions with the phonetic predictors are significant either. More specifically, the most relevant interaction between LF (ejective vs voiceless) and Annotator is insignificant ($\beta = -0.6$ ms, $t = -0.3$, df = 734, $p = 0.77$); so is the interaction between LF (ejective vs voiced) and Annotator ($\beta = 1.9$ ms, $t = 0.9$, df = 734, $p = 0.36$). At the same time, the main effect of LF in this subset of the data (the 394 identical tokens from one speaker) is significant in exactly the same way as in the full model described above: vowels are significantly shorter before voiceless aspirated compared to ejectives ($\beta = -5.8$ ms, $t = -4.0$, df = 734, $p < 0.0001$) and significantly longer before voiced than before ejective stops ($\beta = 12.1$ ms, $t = 8.3$, df = 734, $p < 0.0001$). These effects of phonetic predictors are considerably greater than the non-phonetic effect of the annotation process. In sum, the two annotators working on the same data did not yield significantly different results and the estimates of the differences between the two are minor compared to the estimates of the phonetic predictors—the estimates of differences in annotators are about 10%–15% of the estimates of phonetic differences (−5.8 ms vs −0.6 ms and 12.1 ms vs 1.9 ms).

B. Modeling closure duration

As discussed in Sec. I, it has been proposed that closure duration inversely correlates with preceding vowel duration. In fact, closure and preceding vowel duration are assumed to be in causal relationship under several explanations: under the Rate of Closure Transition [1(c)], the Compensatory Temporal Adjustment [1(b)], and the Perceptual Distance [1(e)] hypotheses. These hypotheses can be tested by modeling closure duration together with the effects of laryngeal features. Adding closure as a predictor provides information on whether variation in vowel durations primarily stems from the intrinsic closure duration of stops or from their laryngeal features (or both).

First, the relationship between closure duration and stop laryngeal features needs to be established. Previous literature on Georgian phonetics found no effect of laryngeal features on closure duration: closure duration is reported to not differ significantly across different stop types (Vicenik, 2010). The results from the present experiment, however, suggest a small, but significant effect of LF on closure duration. A linear mixed effects model was fit to data with closure duration as the dependent variable and LF, Place, and VOT (scaled and centered) as independent variables, and random intercepts for Speaker, Frame, and Preceding Vowel ([ɑ], [ɛ], or [ɔ]) (Place and Vowel coded as above) with by-speaker and by-frame random slopes for LF. All two-way interactions were significant. Closure duration is shorter for voiced stops compared to ejective stops ($\beta = -5.0$ ms, $t = -3.3$, df $= 20$, $p < 0.01$) and longer for voiceless aspirated stops than for ejective stops ($\beta = 3.7$ ms, $t = 3.0$, df $= 17$, $p < 0.01$) at the means of other predictors. This distribution is exactly inverse to the duration of vowels before stops with these laryngeal features in the model in Sec. III A (Table III). It is thus possible that vowels have different durations before different laryngeal features primarily because of their intrinsic differences in closure duration.

To test this hypothesis, the data were fit to a linear mixed effects model with four predictors: LF (treatment-coded with ejective as the reference level), Closure (scaled and centered), Place, and Vowel (both sum-coded as above). The best-fitting model was chosen with the step-wise backwards model selection technique described above in Sec. III A. The best-fitting random effects structure includes random intercepts for Speaker and Frame as well as by-speaker random slopes for LF and Closure and by-frame random slope for Closure. The best fitting fixed effects structure includes all four predictors and the interactions LF $\times$ Closure and Place $\times$ Vowel. Despite the correlation between closure duration and laryngeal features that was discussed above, adding closure duration as a predictor does not introduce multicollinearity. Multicollinearity is evaluated by VIFs, a quantified measure of inflation of variance that "is based on the proportion of variance the *i*th independent variable shares with the other independent variables in the model" (O'Brien, 2007) and is indicative of multicollinearity. Rules of thumb exist for thresholds of excessive multicollinearity: they vary from 4 to 10, depending on different scholars (for a discussion, see O'Brien, 2007). In the model in Table IV, VIFs for all coefficients are below 1.8. Table IV summarizes the model.

The model in Table IV shows that vowels are significantly longer before ejective stops than before voiceless aspirated stops and significantly shorter before ejectives than before voiced stops (at the mean of closure duration), just like in the simple model (Table III). Closure duration also has a significant effect on preceding vowel duration: closure duration is inversely correlated with vowel duration for ejective stops. There is a significant interaction between closure duration and LF: closure duration in voiced stops has a significantly more negative effect on vowel duration compared to ejective stops. This interaction remains significant even if a by-speaker random slope for LF $\times$ Closure is added to the model (otherwise unjustified by AIC). There is no significant interaction between voiceless and ejective stops and closure duration. Figure 3 illustrates the relationship of LF and Closure on preceding vowel duration.

TABLE IV. Linear mixed effects model with four predictors: LF, Closure duration, Place, and Vowel.

	Estimate	SE	df	t value	Pr(>|t|)
(Intercept)	87.07	5.63	13	15.46	0.0000
ejec. vs voiced	8.68	1.22	10	7.09	0.0000
ejec. vs v'less	−4.69	0.65	11	−7.19	0.0000
closure	−1.78	0.73	21	−2.44	0.0241
mean vs lab.	−3.73	0.21	7814	−18.13	0.0000
mean vs alv.	3.25	0.19	7833	17.06	0.0000
mean vs [ɛ]	−1.81	0.19	7825	−9.55	0.0000
mean vs [ɔ]	−3.72	0.19	7827	−19.59	0.0000
voiced:closure	−2.01	0.44	180	−4.56	0.0000
v'less:closure	−0.33	0.43	127	−0.76	0.4491
lab.:[ɛ]	0.40	0.27	7825	1.48	0.1379
alv.:[ɛ]	−0.29	0.27	7824	−1.08	0.2820
lab.:[ɔ]	0.64	0.27	7828	2.37	0.0180
alv.:[ɔ]	0.74	0.27	7820	2.76	0.0059

C. Modeling VOT

The models with three and four predictors above (Tables III and IV, respectively) show that vowels are consistently shorter before ejectives than before voiced stops and longer before ejectives than before voiceless aspirated stops, even when closure duration is controlled for. The differences in duration are summarized in (2):

(2) *Vowel duration before stops with different laryngeal features*
voiced > ejective > voiceless aspirated

This distribution corresponds inversely to the VOT duration of these stops. VOT is expected to be longest in voiceless aspirated stops, shorter in ejectives, and shortest in voiced stops. Three studies on Georgian phonetics confirm this distribution, summarized in (3) (Wysocki, 2004; Vicenik, 2010; Grawunder *et al.*, 2010):

(3) *VOT duration before stops with different laryngeal features*
voiceless aspirated > ejective > voiced

Data collected in this experiment also confirm the generalization in (3). In order to test differences in VOT duration between voiced, voiceless aspirated, and ejective stops, the data were fit to a model with VOT as the dependent variable and four independent variables: LF, Closure (scaled

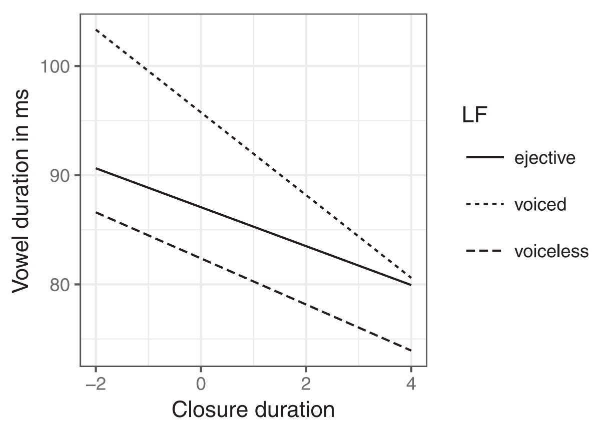

FIG. 3. Estimates of the effects of Laryngeal Features and Closure duration (scaled and centered) on preceding vowel duration in ms (from a linear mixed effects model).

and centered), Vowel, and Place (coded as above), with all interactions[2] and random intercepts for Speaker and Frame with by-speaker random slope for LF. VOT is significantly longer in voiceless aspirated stops compared to ejective stops ($\beta = 36.1$ ms, $t = 9.8$, df $= 11$, $p < 0.0001$) and significantly shorter in voiced stops compared to ejective stops ($\beta = -18.1$ ms, $t = -6.7$, df $= 11$, $p < 0.0001$) at the means of all other predictors.

To the author's knowledge, no existing studies on differences in vowel duration as a function of the following stop type either model duration of VOT or hypothesize that VOT directly affects preceding vowel duration [Port and Rotunno (1979) show that VOT and the *following* vowel duration are in a dependent relation, but they do not measure effects of the preceding vowel duration]. Based on (2) and (3), a hypothesis that VOT inversely affects preceding vowel duration can be added to the collection in (1):

(4) *VOT effect hypothesis*
VOT inversely affects preceding vowel duration

It is possible that differences in vowel durations are primarily due to intrinsic VOT durations of different stop types, rather than due to their different laryngeal features. In other words, it is possible that laryngeal features cease to be a significant predictor once VOT is included in the model, especially because VOT is one of the few robust and consistent predictors of laryngeal features in Georgian (Wysocki, 2004; Vicenik, 2010; Grawunder *et al.*, 2010). To test this hypothesis, the data were fit to a linear mixed effects model with all three predictors of interest: VOT, Closure, and LF. The best fitting model was again chosen according to the procedure described in Sec. III A and Sec. III B. Models with random slopes for interactions between predictors of interest fail to converge, which is why interactions were omitted from the random effects structure. The final model includes random intercepts for Speaker and Frame with three random slopes for both random intercepts: LF, Closure, and VOT duration. The fixed effects structure includes all five predictors (LF, scaled and centered Closure duration, scaled and centered VOT duration, Place, and Vowel, both coded as above) and interactions LF × Closure, Closure × VOT, VOT × Place, VOT × Vowel, Place × Vowel.[3] Table V summarizes the model.

The significance of all predictors of interest from the previous models remains the same when VOT is included in the model. The model in Table V shows that vowels are significantly longer before ejectives than before voiceless aspirated stops and shorter before ejectives than before voiced stops (at the mean of closure duration), even when VOT and Closure duration are included in the model. Closure duration has a negative effect on preceding vowel duration (for ejective stops). There is a significant interaction between voiced stops and closure duration: the effects of closure are significantly more negative for voiced stops than for ejectives. There is no significant interaction between voiceless aspirated and ejective stops and closure duration.

The effects of the new predictor, VOT, are more challenging to interpret. The coefficients of the main effect VOT are $\beta = -1.5$ ms, $t = -2.1$, df $= 15$, $p = 0.052$. VOT as a main effect is thus non-significant at the means of the predictors Closure, Place, and Vowel, but marginally so (note that $|t| > 2.0$). Moreover, adding VOT and its interactions to the model (and keeping the same random effects structure) significantly improves the fit according to the LRT [$\chi^2(6) = 25.2, p < 0.001$].

TABLE V. Linear mixed effects model with five predictors: LF, Closure duration, VOT duration, Place, and Vowel.

	Estimate	SE	df	*t* value	Pr(>|*t*|)
(Intercept)	86.63	5.80	13	14.94	0.0000
ejec. vs voiced	7.74	1.34	10	5.80	0.0001
ejec. vs v'less	−3.30	1.11	15	−2.96	0.0099
closure	−2.12	0.78	19	−2.70	0.0141
VOT	−1.50	0.71	15	−2.11	0.0521
mean vs lab.	−3.93	0.21	7624	−18.80	0.0000
mean vs alv.	3.17	0.19	7709	16.38	0.0000
mean vs [ɛ]	−1.81	0.19	7773	−9.61	0.0000
mean vs [ɔ]	−3.69	0.19	7778	−19.55	0.0000
voiced:closure	−2.40	0.47	181	−5.08	0.0000
v'less:closure	0.52	0.58	193	0.89	0.3749
closure:VOT	−0.76	0.30	383	−2.58	0.0104
VOT:lab.	0.02	0.20	5720	0.11	0.9162
VOT:alv.	0.41	0.19	7803	2.12	0.0341
VOT:[ɛ]	0.34	0.19	7781	1.79	0.0731
VOT:[ɔ]	−0.57	0.19	7787	−2.97	0.0030
lab.:[ɛ]	0.41	0.27	7779	1.52	0.1280
alv.:[ɛ]	−0.22	0.27	7779	−0.82	0.4146
lab.:[ɔ]	0.63	0.27	7781	2.35	0.0187
alv.:[ɔ]	0.62	0.27	7780	2.32	0.0204

Three two-way interactions with VOT are significant. VOT in dentals has a significantly less negative effect on preceding vowel duration compared to the mean across places of articulation. VOT is significantly more negatively correlated with the vowel [ɔ] compared to the mean across vowels. Finally, VOT is significantly more negatively correlated with vowel duration as closure duration increases. Estimates of these interactions are very small (from $\beta = -0.76$ ms to $\beta = 0.41$ ms) and are not of interest to our study. Despite these significant interactions, VOT is, for all places, vowels, and closure durations, inversely correlated with preceding vowel duration.

As expected, including VOT in the model causes multicollinearity. Coefficients for ejective vs voiceless aspirated and VOT are highly correlated. The VIF for VOT in the five-predictor model is 6.70, and the VIF for ejective vs voiceless aspirated is 3.94. It is possible that multicollinearity inflates standard errors. In the model in Table IV with four predictors in which the VIF values are lowest (lower than 1.8 for all predictors), the SE for ejective vs voiceless is 0.7 ms, whereas in the model with five predictors, the SE increases by roughly 50% to 1.1 ms. Estimates for the predictor LF are slightly smaller in the model that includes VOT compared to the four-predictor model (Table IV). In any case, the VIF for ejective vs voiceless is below the "lower" rule of thumb threshold (4) and for VOT is well below the "higher" (10) rule of thumb threshold (O'Brien, 2007). Our data sample is relatively large, which is why the model is probably not crucially affected by multicollinearity and the conclusions from the previous paragraph can be maintained.

D. Closure and VOT as a single predictor

It is possible that the total duration of a stop, that is, closure and VOT together, has a phonetically relevant value that crucially affects preceding vowel duration. To test whether durational differences in vowels result primarily from this combined parameter (Closure + VOT), or whether LF remains a significant predictor even when Closure + VOT is modeled, the data with the new variable, Closure + VOT (ClsVOT), were fit to a linear mixed effects model. The best fitting model was chosen as described above and includes predictors LF (treatment-coded), ClsVOT (scaled and centered), Place, and Vowel (sum-coded as above), interactions LF × ClsVOT, Place × Vowel, and Place × ClsVOT, random intercepts for Speaker and Frame with by-speaker and by-frame random slopes for LF and ClsVOT. The model is summarized in Table VI. The model features much less multicollinearity compared to the model in Table V: VIFs for all coefficients are below 1.80, except for the coefficients ejective vs voiceless aspirated (2.56) and ClsVOT (2.73). The AIC, however, is lower for the model with five predictors (61 983) compared to the model in which Closure and VOT are a single predictor (62 040), which means that from a modeling perspective, treating Closure duration and VOT separately yields a statistically better model.

The results are consistent with results from previous models and do not change if closure and VOT are modeled as a single predictor. Vowels are significantly longer before voiced stops compared to ejectives and significantly shorter before voiceless aspirated stops compared to ejectives at the mean of ClsVOT duration. ClsVOT is significantly negatively correlated with vowel duration for ejectives and while in voiced stops this correlation is significantly more negative compared to ejectives, there is no significant difference in the effect of ClsVOT on preceding vowel duration for the opposition voiceless aspirated vs ejective stops.

TABLE VI. Linear mixed effects model with closure and VOT as a single predictor.

	Estimate	SE	df	t value	Pr(>|t|)
(Intercept)	86.58	5.72	14	15.13	0.0000
ejec. vs voiced	5.80	1.17	13	4.94	0.0003
ejec. vs v'less	−2.75	1.05	15	−2.63	0.0189
ClsVOT	−3.12	1.02	18	−3.05	0.0068
mean vs lab.	−4.38	0.19	7813	−22.79	0.0000
mean vs alv.	3.13	0.19	7819	16.29	0.0000
mean vs [ɛ]	−1.78	0.19	7776	−9.41	0.0000
mean vs [ɔ]	−3.69	0.19	7781	−19.45	0.0000
lab.:[ɛ]	0.24	0.27	7777	0.88	0.3795
alv.:[ɛ]	−0.20	0.27	7774	−0.73	0.4648
lab.:[ɔ]	0.82	0.27	7777	3.07	0.0022
alv.:[ɔ]	0.64	0.27	7776	2.39	0.0171
lab.:ClsVOT	−0.23	0.19	7803	−1.17	0.2410
alv.:ClsVOT	0.58	0.20	7813	2.96	0.0031
voiced:ClsVOT	−2.12	0.64	127	−3.33	0.0011
v'less:ClsVOT	0.72	0.54	101	1.34	0.1848

IV. DISCUSSION

A. The ejection effect

All models consistently show that vowels are significantly longer before ejective stops than before voiceless aspirated stops and shorter before ejectives than before voiced stops, even when closure duration and VOT duration are controlled for. This phonetic generalization is called the *ejection effect* henceforth.

(5) *Ejection effect*
Vowel duration differs significantly before ejective stops compared to voiced and voiceless aspirated stops when closure duration and VOT duration are controlled for. Vowels are longest before voiced stops, shorter before ejective stops, and shortest before voiceless aspirated stops.

This experiment provides evidence in favor of the ejection effect in Georgian. Further experiments on other languages should be conducted to confirm the universality of the ejection effect. However, based on the results from this study, it is reasonable to assume that the ejection effect is replicable for languages with similar ejective stop realizations as Georgian. There are three main arguments that support this assumption. First, all 12 Georgian speakers as well as all 25 nonce-word frames consistently show the ejection effect, both in raw means (Fig. 1) and in random slopes. Except for one speaker and one frame in the five-predictor model, random slopes of all 12 speakers and 25 frames are consistent with the ejection effect in the models with three, four, and five predictors (Secs. III A, III B, and III C). Second, the experiment was conducted using nonce words, which means that Georgian-internal word frequency effects were controlled for. Neighborhood density effects that could also influence durations were not specifically controlled for, but the high number of nonce word frames and the balanced design in this study should reduce potential effects of neighborhood density (for a study that suggests neighborhood density does not affect vowel durations, see Munson and Solomon, 2004). Finally, a pilot experiment was conducted with the same experimental design as for the Georgian experiment (described in Sec. II C), but with a speaker of Mingrelian, another Kartvelian language with voiced, voiceless aspirated, and ejective stops. The confounds of this pilot experiment are that the experiment involves only one speaker and that the speaker's first language was Georgian. The speaker was nevertheless fluent in Mingrelian and spoke it in his grandparents' village with them and other relatives. The carrier phrase was in Mingrelian to prompt the speaker away from Georgian influences. That the speaker used Mingrelian in the experiment is suggested by very short VOT durations in voiceless aspirated stops (shorter in raw means than in any Georgian speaker). Despite this difference, the Mingrelian speaker also shows the ejection effect. In a model with five predictors, vowels are significantly longer before voiced compared to ejective stops ($\beta = 28.4$ ms,

$t = 8.4$, df $= 613$, $p < 0.0001$) and significantly shorter before voiceless aspirated compared to ejective stops ($\beta = -5.2$ ms, $t = -2.1$, df $= 623$, $p < 0.05$). This pilot study with only one speaker cannot contribute any firm conclusions, but it does suggest that the ejection effect might be found in other languages as well. Again, experiments with further speakers and languages are needed to confirm the universality of the ejection effect.

B. Closure

The results also show that closure duration is negatively correlated with preceding vowel duration and more negatively correlated for voiced stops compared to ejectives.

The observed negative correlation between vowel and closure duration could be phonetically real: it is possible that closure duration affects preceding vowel duration when laryngeal features are controlled for. However, the correlation between vowel and closure durations could also result from other influences. There are two main problems with interpreting the effects of closure duration: the correlation between vowel and closure duration can be influenced by (i) the dependence of the data annotation process and (ii) by speech rate. These two problems are addressed in this section.

1. Annotation bias

As previously mentioned, it is possible that the inverse correlation between vowel and closure duration results from the fact that the boundary point between vowel and closure is not annotated independently, but rather with a single boundary point. The decision to annotate the data in one direction or the other can result in an inverse correlation between the two variables. This confound of modeling closure duration is inevitable in such studies and is not easy to control for.

For the purpose of testing influences of annotation, two different annotators were instructed to annotate a number of tokens from a single speaker: 394 annotations from each annotator were analyzed. To model the effects of annotation on the correlation between vowel and closure duration, two combinations of vowel duration and closure duration measurements were created: one in which vowel and closure were annotated by a single annotator (dependent), and one in which vowel duration measurements from one annotator were combined with closure duration measurements from the other annotator (the measurements are therefore independent). The data were fit to a linear model with vowel duration as the dependent variable and only Closure duration (scaled and centered) and Dependence (treatment-coded with dependent as the reference) as independent variables with an interaction between the two. Closure duration is significantly inversely correlated with preceding vowel duration for dependent annotations ($\beta = -2.35$ ms, $t = -2.32$, $p = 0.02$). However, the interaction between Closure and Dependence is not significant ($\beta = 0.78$ ms, $t = 0.54$, $p = 0.59$). In other words, the slope for Vowel duration $\sim$ Closure duration is not significantly less negative in independent annotations compared to dependent annotations. Also, the estimates of difference in slope between dependent and independent annotations is 0.78 ms, considerably smaller compared to the effect of closure in the model with five predictors (Table V): $\beta = -2.14$ ms. These results suggest that the inverse effects of closure duration are most likely phonetically real and not due to annotation bias.

2. Speech rate

Speech rate is another possible factor that influences correlation between vowel duration and closure duration. Longer closure duration is expected to positively correlate with longer vowel duration for slower speech and vice-versa for faster speech. The speech rate effect can also reverse correlation from being underlyingly negative to surfacing as positive: it is possible that in a language with a weak underlying negative correlation between vowel and closure durations (due, for example, to some articulatory reason), a strong effect of the speech rate turns an underlying negative correlation to an observed positive correlation. If, for example, the effect of speech rate (positive correlation) is twice the size of the potential articulatory effect that causes negative correlation, the observed correlation between vowel and closure duration would be positive, because the two effects are cumulative. However, because the Georgian data show a negative correlation between vowel and closure duration, effects of speech rate most likely do not crucially influence the results. The experimental design in this paper also avoids repetition of individual tokens, which reduces the effects of speech rate. Moreover, by-speaker and by-frame random intercepts and slopes for Closure duration in the models account for at least some amount of inter-speaker variance in speech rate.

Other studies known to the author yield opposing results with respect to the effect of closure or consonant duration on the preceding vowel. de Jong (1991) models the voicing effect together with consonant duration (along with other predictors) in English. He too reports an inverse relation between vowel and consonant duration, but the effect is inconsistent across speakers and contexts. de Jong (1991) also reports that the voicing effect is more robust than the effect of consonant duration and concludes that the inverse correlation between vowel and consonant duration might be speaker-specific. Only two speakers were tested in the experiment, which does not allow for any firm conclusions. The pilot study on Mingrelian with the same experimental design as described in this paper, on the other hand, yields a positive correlation between closure and vowel duration, but the results there are unreliable as well because only one speaker participated in the experiment.

Durvasula and Luo (2014) is the only other study known to the author that models the effects of closure along with measurements of vowel duration before stops with different laryngeal features. They report a significant positive, but weak correlation of vowel duration and closure duration. Their study involves a high enough number of speakers (seven) and is modeled with random intercepts for subject and token, which means that speech rate is less likely to affect the results. The experiment in Durvasula and Luo (2014) does, however, include repetitions of tokens: each of the 12 unique words is repeated ten times per speaker. Experimental design involving repetitions is generally more prone to speech rate effects. Despite the order of stimuli being randomized, repetition could cause gradual reduction and consequently shorter vowel and consonant durations (Shields and Balota, 1991; Umeda, 1975). The increased reduction over repetitions and increasingly shorter durations could cause positive correlation between vowel and closure durations, but the question remains whether the effect of speech rate could be substantial enough to reverse the correlation. In the Georgian experiment presented in this paper, none of the tokens are repeated.

Based on the presented facts, two possibilities exist. It is possible that effects of closure duration are phonetically real, but weak and language-specific. In Hindi, the effect would be positive, in Georgian (and, with confounds, in English) inverse. However, it is also possible that correlation between closure duration and preceding vowel duration is indeed inverse and universal, as suggested by the Georgian data (and the English data, with confounds, in de Jong, 1991). This conclusion would be supported by the fact that the inverse correlation between vowel and closure duration cannot be influenced by speech rate and that it is relatively consistent in Georgian: only two speakers (of 12 tested) show a weak positive rather than negative random slope in the five-predictor model. Under this approach, the positive correlation in Hindi and in the two speakers with a positive correlation would have to be interpreted as influenced by speech rate. Further experiments are needed for more conclusive evidence in favor or against one or the other possibility.

C. VOT

The experiment in this paper also brings new information on the effects of VOT when modeling Laryngeal Features and Closure at the same time. To the author's knowledge, no study thus far models all three parameters (LF, VOT, and Closure) together. VOT measurements are free of annotation bias: they are independent of vowel annotations. The problem with modeling VOT is that it introduces some multicollinearity to the model (see Sec. III C). It can nevertheless be concluded that VOT affects vowel duration inversely in Georgian when closure duration and laryngeal features are controlled for. Adding VOT and its interactions to the model improves fit significantly. VOT is marginally non-significant as a main effect with significant interactions for Closure, Place, and Vowel. Despite these interactions, the effect of VOT is in all cases inverse. The inverse effect remains significant when both closure duration and VOT are modeled as a single predictor–total stop duration.

In sum, both VOT and closure duration effects appear to be weak predictors of preceding vowel duration (except if modeled as a single predictor), show significant interactions, and yield somewhat inconsistent results across languages. The Georgian data suggest that there is an inverse correlation between both closure and VOT and preceding vowel duration. At the very least, an inverse correlation between VOT and preceding vowel duration has to be phonetically real. The correlation between closure and vowel duration is also likely phonetically real; the tests of annotation bias presented above suggest that the results are likely not crucially affected. Moreover, because the correlation is inverse, speech rate likely does not crucially affect the results in Georgian. The Georgian data also consistently show a significantly greater inverse effect of closure duration for voiced stops compared to ejectives. Firmer conclusions on the universality of the inverse correlation between closure and preceding vowel duration will only be possible with further data, especially because other studies yield contradictory results.

V. IMPLICATIONS

As discussed above, little consensus has been reached on the question of causes of durational differences in vowels. For virtually every proposal in (1), contradictory evidence has been presented. In the interest of space, this paper cannot summarize all arguments; for a more comprehensive overview of the discussion, see Chen (1970), Kluender *et al.* (1988), Fowler (1992), and Maddieson (1997).

The main advantage of the experimental design in this paper is that it measures vowel durations before stops with all three laryngeal features and models both closure and VOT durations at the same time. The results presented in Sec. III thus provide new information on the effects of stops on preceding vowel duration and bear several implications for the general discussion on the causes of these durational differences.

A. Laryngeal Adjustment

Experimental results in this paper suggest that differences in vowel durations before stops with different laryngeal features are not due to their intrinsic closure or VOT duration: even if closure and VOT are modeled, laryngeal features are still significant predictors. Moreover, it is shown that the feature [±voice] is not the only laryngeal feature responsible for durational differences in vowels. Voiceless aspirated and ejective stops are both voiceless (the percent of voicing into closure is slightly but not significantly higher for ejectives according to Vicenik, 2010), but vowel durations before the two series are nevertheless significantly different, which means that other laryngeal features have to affect preceding vowel duration as well. This result aligns well with results from experiments that tested the effect of aspiration on preceding vowel duration: aspiration, or [+spread glottis], according to most studies, is another laryngeal feature that affects vowel duration besides the voice feature (Durvasula and, Luo 2014).

Because laryngeal features remain significant predictors even when closure and VOT durations are modeled and because the effects of LF remain consistent across models, speakers, and languages, not just in this paper but also in others that tested the aspiration effect, it can be concluded that laryngeal features affect preceding vowel duration, even when most other phonetic parameters are controlled for. This generalization, that combines previously established voicing and aspiration effects with the ejection effect, is called the *Laryngeal Features Effect* henceforth.

(6) *Laryngeal Features Effect (LFE)*
Different laryngeal features ([±voice], [±spread glottis], [±constricted glottis]) affect preceding vowel duration, even when closure duration and VOT are controlled for.

This conclusion supports the body of work that assumes laryngeal gestures cause differences in preceding vowel duration [(1)d]: the Laryngeal Adjustment hypothesis (Halle *et al.*, 1967; Chomsky and Halle, 1968). The results from this study suggest that each laryngeal feature, [±voice], [±spread glottis], and [±constricted glottis], affect preceding vowel duration, whereby vowels are longest before stops with the feature [+voice], shorter before stops with the feature [+constricted glottis], and shortest before stops with the feature [+spread glottis].

While some scholars argue against the Laryngeal Adjustment hypothesis (for an overview of arguments, see Kluender *et al.*, 1988), none of the evidence against it is in fact fatal. Chen (1970) himself argues that his counterarguments "[a]dmittedly [⋯] do not entirely rule out the plausibility of the laryngeal adjustment hypothesis." The presence of the voicing effect in whispered and esophageal speech (Sharf, 1964; Gandour *et al.*, 1980) does not necessarily argue against the Laryngeal Adjustment hypothesis either, because the differences are reported for English where the voicing effect is likely phonologized and, more generally, laryngeal gestures can be transferred from non-whispered to whispered speech.

The explanation of vowel duration differences due to laryngeal adjustment in Halle *et al.* (1967) works for effects of the [±voice] feature, but the additional information on LFE from this paper as well as information on the aspiration effect in Durvasula and Luo (2014) requires further explanation as to why vowels also differ before voiceless aspirated and ejective stops. An articulatory explanation of these differences is a desideratum, but beyond the scope of this paper (for speculations and directions of future work, see Sec. VI).

B. Compensatory Temporal Adjustment

Results from this experiment also provide some support for the Compensatory Temporal Adjustment hypothesis [(1)b]. Most studies thus far (except Durvasula and Luo, 2014 and de Jong, 1991) do not control for the effects of LF when modeling the effects of closure duration on preceding vowel duration and, to the author's knowledge, no studies measure the effects of VOT on preceding vowel duration. As already mentioned, this is problematic because laryngeal features and closure and VOT durations are correlated, especially if only two series of stops, voiced and voiceless, are modeled. This study suggests that even when laryngeal features are controlled for, closure duration and VOT have a significant inverse effect on preceding vowel duration (especially when modeled as a single predictor) that is likely not influenced by annotation bias or speech rate.

The fact that both durational differences, closure duration as well as VOT, in the Georgian experiment show inverse correlation with preceding vowel duration supports the assumption that temporal dimension tends to be constant across VC-sequences in the Compensatory Temporal Adjustment hypothesis (Chen, 1970). According to the hypothesis, longer closure and VOT have to result in shorter vowel duration. This is exactly what is found in Georgian, but universality of this hypothesis is yet to be confirmed (the experiment in Durvasula and Luo, 2014 yields exactly the opposite results). For a discussion on results from other languages, see Sec. IV B: it is possible that Compensatory Temporal Adjustment is language-specific (as argued in de Jong, 1991).

While the present study does not measure the correlation between vowel and closure/VOT duration in tautosyllabic, but rather in heterosyllabic VC sequences, the results are nevertheless informative: the effects of Compensatory Temporal Adjustment are not limited in scope to syllables (Port *et al.*, 1980, 1987). It is reasonable to assume that a longer inherent closure and VOT duration in heterosyllabic (V.C) sequences would cause adjustment in vowel duration, just as it would in tautosyllabic (VC) sequences, because the voicing effect is confirmed in both sequences and the effect presumably has the same underlying mechanism in both conditions (see discussion in Sec. I). In addition, temporal adjustment has been shown to operate across syllables in other contexts and conditions as well (Port *et al.*, 1980, 1987).

The results in this paper are in opposition to Chen's (1970) argument against the Compensatory Temporal Adjustment. Chen (1970) presents an experiment that might speak against this hypothesis if constant timing is understood absolutely. Chen (1970) reports that the total syllable duration in words like [paik] vs [paikt] is not constant, as would be expected under the more radical approach to the Compensatory Temporal Adjustment hypothesis, and that vowel durations differ minimally in the two words. Klatt (1976) and Port (1981) have argued, however, that various factors influence vowel durations proportionally, not absolutely, which means that Compensatory Temporal Adjustment can have a minimal effect. The results in this paper are consistent with this latter approach, but it is true that further research is a desideratum because studies that measure the correlation between vowel and closure durations and control for laryngeal features at the same time are rare and yield somewhat inconsistent results.

C. Rate of Closure Transition

A somewhat surprising but consistent and significant observation from the data is that the inverse correlation of vowel and closure duration is significantly greater for voiced stops compared to ejective and voiceless aspirated stops. This generalization is called the *Voiced Closure Effect* henceforth.

(7) *Voiced Closure Effect (VCE)*
Closure is significantly more inversely correlated with preceding vowel duration in the voiced series of stops.

As for any inverse correlation between vowel and closure duration, it is possible that the VCE results from nonphonetic factors. The transition between vowel and consonant is less clear for voiced stops than for voiceless stops, which is why increased inverse correlation in voiced stops could be due to annotation bias. To test the phonetic reality of VCE, the data with the combined dependent and independent 394 annotations (the combination of measurements was created as in Sec. IV B 1 above) were fit to a linear model with vowel duration as the dependent variable and Closure (scaled and centered), Dependence (treatment-coded with dependent as reference), and LF (treatment-coded with ejective as reference) as independent variables. The three-way interaction of Closure:Dependent:LF is not significant for voiced vs ejective ($\beta = -1.66$, $t = -0.45$, $p = 0.65$), although the closure is also not significantly more negative for voiced stops compared to ejectives. In other words, the slope for Vowel duration $\sim$ Closure duration in voiced stops compared to ejectives is not significantly different between the independent and the dependent annotations. These results suggest that the VCE does not result from annotation bias.

The most compelling phonetic cause for the VCE would be Rate of Closure Transition. It has been argued that closure gesture is reached with greater articulatory force and with greater velocity for voiceless stops compared to voiced stops (Chen, 1970). To the author's knowledge, no empirical studies exist on closure velocity for ejective stops, but it is reasonable to assume that they pattern together with voiceless aspirated stops: resistance to higher intraoral pressure is assumed to be one of the factors that leads to greater closure velocity/force (Chen, 1970). Intraoral pressure of ejective stops is more likely to resemble intraoral pressure in voiceless aspirated stops (Shosted and Rose, 2011) than to resemble the pressure in voiced stops. The Rate of Closure Transition hypothesis would constitute a plausible cause for the VCE; smaller velocity and force of closure gesture in voiced stops would allow even greater negative correlation between vowel and closure duration: the slower it takes to reach full closure, the longer the vowel can be. However, at the current stage, this is only a speculation that requires further acoustic and articulatory research.

D. Airflow Expenditure

An absence of information about vowel duration before ejective stops could lead to the assumption that greater oral airflow in the VOT of voiceless aspirated stops causes shorter vowel duration under the hypothesis that airflow tends to be constant across syllables/VC-sequences (for an experiment that tests the relationship between airflow expenditure and the following vowel quality, see Hamann and Velkov, 2005). In other words, it would be reasonable to assume that greater oral airflow expenditure in the stop production would be compensated for by shorter vowels that require less airflow. The data in this paper clearly contradict such an assumption. A qualitative study of oral airflow of Georgian stops in Shosted and Chikovani (2006) shows that the amount of airflow in Georgian is greatest in voiceless aspirated stops, smaller in voiced stops, and smallest in ejectives, as summarized in (8). While voiceless unaspirated stops sometimes feature less airflow than voiced unaspirated stops (Nihalani, 1975), it is expected for voiceless *aspirated* stops to have greater airflow than voiced stops (as it is reported for Georgian in Shosted and Chikovani, 2006 and for other languages in Trullinger and Emanuel, 1983; Gilbert, 1973).

(8) *Scale of oral airflow expenditure across stops with different LF*
voiceless aspirated < voiced < ejective

Vowel duration differences do not align according to this scale: vowels are longest before voiced stops, shorter before ejectives, and shortest before voiceless aspirated stops. Due to this mismatch, the amount of airflow can therefore be neither inversely nor positively correlated with preceding vowel duration. The airflow expenditure explanation also fails to account for the aspiration effect in Durvasula and Luo (2014), where vowels before aspirated stops are longer than before plain stops. Under the airflow approach, the opposite distribution would be expected. It can therefore be concluded that airflow expenditure during VOT does not affect preceding vowel duration in any significant way.

E. Perceptual Distance

The results of the present study also considerably weaken the Perceptual Distance hypothesis [(1)e]. The Perceptual Distance approach adopts the claim that closure duration is a prominent cue of stop voicing. Vocalic differences are argued to arise in order to maximize this perceptual contrast: a longer vowel makes the following shorter closure of a voiced stop perceptually even shorter, thus enhancing the perceptual cue (Kluender *et al.*, 1988) [(1)e]. However, the durational differences found in this study are small: the 95% profile CIs for the coefficient ejective vs voiceless aspirated is [−5.8 ms, −3.5 ms]. Perception is most likely not able to capture such small differences in duration, especially under the assumption that perceptual enhancement *causes* these durational differences to arise, i.e., that differences originate in perceptual enhancement and not that perception builds on some already present articulatory distribution.

This objection to the Perceptual Distance explanation has been raised already in Chen (1970). Chen's study shows that the ratio of vowel duration differences in languages other than English fall below the Just-Noticeable Difference (JND) threshold, established for tonal durational differences at the time of the study. Chen (1970) operates with Weber $\Delta T/T$ JND ratios that range from 0.120 to 0.196 (established by Stott, 1935 and Henry, 1948). Other studies report similar JND ratios (e.g., approximately 0.125 in Abel, 1972; cf. Nooteboom and Doodeman, 1980; for closure and burst stimuli, see Huggins, 1968; Abel, 1972), but the substantially lower JNDs from Ruhm *et al.* (1966) are rejected as not reflecting realistic speech situations (Lehiste, 1970; Chen, 1970)—the experimental design in Ruhm *et al.* (1966) involves "extensive training" (Nooteboom and Doodeman, 1980). A later study, however, reports a considerably smaller Weber $\Delta T/T$ JND ratio based on *vocalic* stimuli that should reflect realistic speech situation: the measured $\Delta T/T$ JND ratio in Dutch is 0.055 (Nooteboom and Doodeman, 1980).

The JNDs are not the perfect metrics for estimating the minimal threshold for perceptual distinctness: JNDs are based on conscious judgments and it is possible that a phonetic difference is subconsciously perceptually distinct even if overt judgments do not capture it. Such covert distinctness might be invoked, for example, in near mergers unless near mergers are analyzed as being influenced by orthography, morphophonology, or coarticulation [e.g., in the case of vocalic differences before final (de)voiced stops, see below]. Warner *et al.* (2004) measure vowel duration differences before incompletely neutralized final voiceless and underlyingly voiced stops in Dutch. They find a very small difference in vowel duration (3.5 ms) before the two series, which is, besides burst duration after long vowels, the only significant difference they find. This, however, does not mean that other parameters in the production of underlying /t/ and /d/ are not different: in fact, other studies on the Dutch near merger found parameters other than vowel duration to be significant (Baumann, 1995; Ernestus and Baayen, 2007). In other words, vowel duration could be influenced by the production of the following stop. In the perception part of their study, speakers were asked to distinguish between minimal pairs with incompletely neutralized final stops. This task is perhaps the closest approximation to testing subconscious perceptual distinctness. Speakers were able to identify underlying voicedness better than chance, but only for two speakers—the one who had "a relatively large difference in vowel duration, and the largest difference of any speaker for burst duration," and the one who had the "largest effect on vowel duration of any speaker, but a relatively small effect on burst duration." Unfortunately, the authors do not reveal the mean difference of vowel duration for the two speakers, but it is, based on their description, higher than the average for all speakers (3.5 ms).

Other studies that measure vowel duration in the production of incompletely neutralized final stops find much greater differences: 8–15 ms (Port and O'Dell, 1985; Slowiaczek and Dinnsen, 1985; Smith *et al.*, 2009; Roettger *et al.*, 2014). van Rooy *et al.* (2003) report the smallest difference at 3 ms in one subcategory, but other categories feature substantially larger differences. Pye (1986) (via Dmitrieva *et al.*, 2010) and Dmitrieva *et al.* (2010) report the difference at 5–20 ms for Russian. Some studies find no significant differences in vowel duration at all (Shrager, 2012; Dinnsen and Charles-Luce, 1984; Piroth and Janker, 2004 find a significant difference for only one speaker). For an overview of the literature, see Shrager (2012) and Kharlamov (2015). Port and O'Dell (1985) report that in German the difference is 15 ms, and even with such a large difference, listeners were able to correctly identify only approximately 60% of the stimuli (with similar results also in Roettger *et al.*, 2014). The author is unaware of any studies of near mergers that would involve length distinction and would control for the cues and articulatory influence of the following stop (e.g., near merger of long and short vowels in a given environment).

Note that the differences in the present experiment fall slightly below the lowest reported JND ratio. The $\Delta T/T$ for vocalic differences between different stop types can be calculated based on estimates from the simple model (Sec. III A) that includes only LF as the predictor of interest. If vowel duration before the ejective is taken at the means of other predictors (Intercept in the model in Table III), the $\Delta T/T$ ratio of the difference between vowel duration before ejective vs voiceless aspirated stops compared to vowel duration before ejectives is 4.7 ms/87.0 ms = 0.054. While this is just slightly below the Weber $\Delta T/T$ JND ratio in Nooteboom and Doodeman (1980), it falls well below the Weber $\Delta T/T$ JND ratios obtained in other studies. With 4.7 ms in absolute duration, the differences in vowel durations between voiceless and ejective stops are also close to the smallest measured near merger in vowel duration: 3.5 ms in Dutch (Warner *et al.*, 2004) and well below the 8–15 ms difference measured in the majority of other studies (Port and O'Dell, 1985). It is clear that the 3.5 ms difference was not perceived as distinct in a perceptual experiment that requires subjects to differentiate between minimal pairs (Warner *et al.*, 2004).

These results suggest that the differences in vowel duration before voiceless aspirated stops and ejective stops are too small to be perceptible. Precisely because the observed vocalic differences are likely not perceptible, the significant differences in vowel durations before voiced, voiceless aspirated, and ejective stops are likely caused by articulatory rather than by perceptual factors. Further research on exactly which durational differences can be perceived subconsciously is needed before final conclusions that rule out perception can be drawn.

It has to be noted that while perception is likely not the factor that *causes* durational differences in vowels, it most likely *does* contribute to durational differences in their magnitude: speakers can perceptually enhance an existing articulatory generalization (in contexts with greater magnitude of the effect where it is perceptible) in order to maximize perceptual cues. For example, the fact that the coda voicing effect in English is much greater in magnitude compared to other languages is likely due to perceptual factors: English speakers employed an existing phonetic generalization and used it to enhance perceptual cues (cf. de Jong, 1991, 2004; Solé, 2007).

VI. CONCLUSION AND FUTURE WORK

This paper presents evidence in favor of the *ejection effect*: vowels are longer before ejectives than before voiceless aspirated stops and shorter before ejectives than before voiced stops when most other phonetic parameters (including closure duration and VOT) are controlled for. The ejection effect is confirmed for Georgian, but further studies are required to confirm the universality of these results. Because it is consistent across speakers, nonce-words, and models, and we expect the ejection effect to be replicable for languages with similar ejective stops realization as in Georgian. Based on the results from this paper as well as other work on the effects of aspiration, it is proposed that all three laryngeal features affect preceding vowel durations when most other phonetic parameters are controlled for (LFE).

Closure and VOT duration are argued to inversely affect preceding vowel duration in Georgian, although this generalization is not as consistent across languages. The paper also presents techniques for testing annotation bias (by combining data from different annotators): the tests suggest that the ejection effect as well as the inverse effect of closure in Georgian are phonetically real. Moreover, this paper argues that the Georgian experiment better controls for speech rate effects than other studies do, and that positive correlation elsewhere might be due precisely to influences of speech rate, while at the same time acknowledges that the inverse correlation between closure and preceding vowel duration can also be language-specific. Further research is needed for more conclusive results. The experiment also shows that closure affects vowel duration more for voiced stops (VCE) and suggests that this effect is also phonetically real.

The Laryngeal Adjustment, Compensatory Temporal Adjustment, and Rate of Closure Transition hypotheses receive support from these results. It is thus likely that more than a single factor influences vowel duration before stops with different laryngeal features. The Perceptual Distance and Airflow Expenditure hypotheses, on the other hand, are considerably weakened by the results.

The paper also outlines why testing vowel duration before only voiceless and voiced obstruents can lead to misleading conclusions: laryngeal features are often correlated with closure duration and VOT. The study thus aims to be a step towards expanding measurements of vowel duration differences to positions before stops beyond the voiced-voiceless or aspirated-unaspirated distinction. The new information on vowel duration before ejective stops yields implications for the discussion in (1) and provides grounds for more comprehensive models of vowel durations and timing relations in phonetics. Further information on vowel durations before non-obstruent segments (such as nasals and laterals, modeled together with effects of all laryngeal features) or non-egressive segments (such as implosives or clicks) would undoubtedly shed further light on the proposals in (1) and yield implications for the general models of timing in speech (Klatt, 1976; Port, 1981). However, non-obstruent or non-egressive segments introduce further complications as their articulation differs from egressive obstruents in many respects (see Ladefoged and Maddieson, 1996), which means many further factors would have to be controlled for. Measuring vowel durations before ejective, voiced, and voiceless aspirated stops involves minimal changes: all three series of obstruents involve a closure gesture, differ minimally in airstream mechanism, and allow for accurate measurements, which means that most articulatory parameters other than laryngeal features are controlled for.

The results allow further speculations on the exact mechanisms and articulatory explanations for durational differences and suggest future directions that research on the effects of obstruents on preceding vowel duration should take. The three laryngeal features (voice, spread glottis, and constricted glottis) require three very different gestures that involve both tongue root movement and laryngeal rising and lowering (Westbury, 1983; Hong *et al.*, 2002; Kingston, 1985; Moisik and Esling, 2014; Hirose and Gay, 1972; Hirose, 1977; and literature therein). It can be speculated that these different gestural targets require different times to be achieved, which would result in different durations of the preceding vowel. According to the results in this paper, spreading of the glottis for voiceless aspirated stops would require the shortest period of time and the laryngeal gestures necessary for voicing the longest, while adjustment for ejection—constriction of the glottis—would take longer than spreading, but shorter than the adjustment for voicing. According to the line of reasoning in Halle *et al.* (1967), laryngeal gestures required for ejective stops—closure of the glottis—would take a longer time compared to laryngeal gestures for spread glottis because pressure buildup during closure facilitates the latter but not the former. Evaluation of these claims is beyond the scope of the present paper. Future studies on the issue should involve articulatory data and should explore an articulatory basis for the LFE hypothesis.

ACKNOWLEDGMENTS

I would like to thank Edward Flemming, Kevin Ryan, Patrick Jones, and Maria Polinsky, and the audience at the 172nd ASA meeting, as well as the editor and reviewers for their useful comments. All mistakes are my own. Special thanks go to all thirteen informants and nine undergraduate research assistants at Harvard who helped with data annotations. The project was funded by the Mind Brain Behavior Interfaculty Initiative at Harvard University and the Abby and George O'Neill Research Grant at Davis Center at Harvard University to the author, as well as the NSF grants BCS-1144223 and BCS-1619857 to Maria Polinsky.

[1]Profile CIs are calculated by "computing a likelihood profile and finding the appropriate cutoffs based on the likelihood ratio test" (Bates *et al.*, 2017). Profile CIs on mixed effects models are chosen over Wald CIs because they require a weaker assumption (for a detailed discussion, see Bates *et al.*, 2015); the interpretation of both types of CIs is essentially the same. For the purpose of comparison, Wald and parametric bootstrap (with 500 simulation) CIs were also calculated: the Wald, profile, and bootstrap CIs match almost completely, with the greatest difference between the three being smaller than 0.3 ms.

[2]The model with all interactions was chosen because models without four-way interactions or without some three-way interactions had problems with convergence. This model shows traces of heteroskedasticity in the residual plot. Several strategies were employed to address this issue: the data was fit to a Gamma regression, inverse Gaussian regression, log-transformed, and modeled with different variances in groups with *lme()* function from *nlme* package (Pinheiro *et al.*, 2016). The latter method helps reduce heteroskedasticity, but the model had to be fit without the crossed random intercept for Frame, which introduces new problems. In all these attempts, significance of coefficients remains the same, which is why the initial model is kept with this acknowledgement.

[3]Because the interactions VOT × Place and VOT × Vowel were only marginally non-significant during the step-wise backwards model selection procedure, they were added to the model together after the pre-final model was selected. Based on LRT and AIC, they significantly improve fit, which is why they are included in the final model.

Abdelli-Beruh, N. B. (**2004**). "The stop voicing contrast in French sentences: Contextual sensitivity of vowel duration, closure duration, voice onset time, stop release and closure voicing," Phonetica **61**(4), 201–219.

Abel, S. M. (**1972**). "Duration discrimination of noise and tone bursts," J. Acoust. Soc. Am. **51**(4B), 1219–1223.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**(1), 1–48.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., and Green, P. (**2017**). "Package 'lme4':Linear Mixed-Effects Models using 'Eigen' and S4," https://cran.r-project.org/web/packages/lme4/lme4.pdf (Last viewed October 12, 2017).

Baumann, M. (**1995**). *The Production of Syllables in Connected Speech* (Benda Drukkers, Nijmegen, the Netherlands), p. 198.

Belasco, S. (**1953**). "The influence of force of articulation of consonants on vowel duration," J. Acoust. Soc. Am. **25**(5), 1015–1016.

Boersma, P., and Weenink, D. (**2015**). "Praat: Doing phonetics by computer [Computer program]. Version 5.4.06," from http://www.praat.org/ (Last viewed February 21, 2015).

Byrd, D. (**1996**). "Influences on articulatory timing in consonant sequences," J. Phon. **24**(2), 209–244.

Chen, M. (**1970**). "Vowel length variation as a function of the voicing of the consonant environment," Phonetica **22**(3), 129–159.

Cho, T. (**2015**). "Language effects on timing at the segmental and suprasegmental levels," in *The Handbook of Speech Production*, edited by M. A. Redford (Wiley-Blackwell, Hoboken, NJ), pp. 505–529.

Chomsky, N., and Halle, M. (**1968**). *The Sound Pattern of English* (Harper & Row, New York), Chap. 7, pp. 293–329.

de Jong, K. (**1991**). "An articulatory study of consonant-induced vowel duration changes in English," Phonetica **48**(1), 1–17.

de Jong, K. (**2004**). "Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration," J. Phon. **32**(4), 493–516.

Denes, P. (**1955**). "Effect of duration on the perception of voicing," J. Acoust. Soc. Am. **27**(4), 761–764.

Dinnsen, D., and Charles-Luce, J. (**1984**). "Phonological neutralization, phonetic implementation and individual differences," J. Phon. **12**(1), 49–60.

Dmitrieva, O., Jongman, A., and Sereno, J. (**2010**). "Phonological neutralization by native and non-native speakers: The case of Russian final devoicing," J. Phon. **38**(3), 483–492.

Durvasula, K., and Luo, Q. (**2014**). "Voicing, aspiration, and vowel duration in Hindi," Proc. Mtgs. Acoust. **18**(1), 060009.

Ernestus, M., and Baayen, H. (**2007**). "Auditory perception of duration of speech and non-speech stimuli," in *Voicing in Dutch: (De)voicing—Phonology, Phonetics, and Psycholinguistics*, edited by J. van de Weijer and E. J. van der Torre (John Benjamins, Amsterdam, the Netherlands), pp. 153–174.

Fintoft, K. (**1961**). "The duration of some Norwegian speech sounds," Phonetica **7**(1), 19–39.

Fowler, C. A. (**1992**). "Vowel duration and closure duration in voiced and unvoiced stops: There are no contrast effects here," J. Phon. **20**(1), 143–165.

Fox, J. (**2003**). "Effect displays in R for generalised linear models," J. Stat. Softw. **8**(15), 1–27.

Frank, A. F. (**2011**). "R-hacks/mer-utils.R," github.com/aufrank/R-hacks/blob/master/mer-utils.R (Last accessed October 12, 2017).

Gandour, J., Weinberg, B., and Rutkowski, D. (**1980**). "Influence of postvocalic consonants on vowel duration in esophageal speech," Lang. Speech **23**(2), 149–158.

Gilbert, H. R. (**1973**). "Oral airflow during stop consonant production," Folia Phoniatr. Logop. **25**(4), 288–301.

Grawunder, S., Simpson, A., and Khalilov, M. (**2010**). "Phonetic characteristics of ejectives—samples from Caucasian languages," in *Turbulent Sounds*, edited by S. Fuchs, M. Toda, and M. Zygis (De Gruyter Mouton, Berlin), pp. 209–224.

Halle, M., Stevens, K. N., and Oppenheim, A. V. (**1967**). "On the mechanism of glottal vibration for vowels and consonants," Massachusetts Institute of Technology, Research Laboratory of Electronics, Quarterly Progress Report 85, 267–277, available at http://hdl.handle.net/1721.1/55578.

Hamann, S., and Velkov, H. (**2005**). "Airflow in stop-vowel sequences of German," ZAS Papers Ling. **42**, 1–12, available at http://www.zas.gwz-berlin.de/190.html?&L=1%2527%252band%252bchar%28124%29%25252Buser%25252Bchar%28124%29%25253D0%252band%252b%2527%2527%25253D%2527.

Henry, F. M. (**1948**). "Discrimination of the duration of a sound," J. Exp. Psychol. **38**(6), 734–743.

Hirose, H. (**1977**). "Laryngeal adjustments in consonant production," Phonetica **34**, 289–294.

Hirose, H., and Gay, T. (**1972**). "The activity of the intrinsic laryngeal muscles in voicing control," Phonetica **25**, 140–164.

Hong, K. H., Kim, H. K., and Niimi, S. (**2002**). "Laryngeal gestures during stop production using high-speed digital images," J. Voice **16**(2), 207–214.

House, A. S. (**1961**). "On vowel duration in English," J. Acoust. Soc. Am. **33**(9), 1174–1178.

Huggins, A. W. F. (**1968**). "The perception of the timing in natural speech I: Compensation within the syllable," Lang. Speech **11**(1), 1–11.

Javkin, H. (**1975**). "Perceptual basis of vowel duration differences," J. Acoust. Soc. Am. **58**(S1), S58.

Jong, K., and Zawaydeh, B. (**2002**). "Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration," J. Phon. **30**(1), 53–75.

Keating, P. (**1985**). "Universal phonetics and the organization of grammars," in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, edited by V. A. Fromkin (Academic Press, Orlando, FL), pp. 115–132.

Kharlamov, V. (**2015**). "Perception of incompletely neutralized voicing cues in word-final obstruents: The role of differences in production context," Lab. Phon. **6**(2), 147–165.

Kingston, J. (**1985**). "The phonetics and phonology of the timing of oral and glottal events," Ph.D. thesis, University of California, Berkeley, Berkeley, CA.

Klatt, D. H. (**1973**). "Interaction between two factors that influence vowel duration," J. Acoust. Soc. Am. **54**(1), 313.

Klatt, D. H. (**1976**). "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," J. Acoust. Soc. Am. **59**(5), 1208–1221.

Kluender, K. R., Diehl, R. L., and Wright, B. A. (**1988**). "Vowel-length differences before voiced and voiceless consonants: An auditory explanation," J. Phon. **16**(2), 153–169.

Kozhevnikov, V. A., and Chistovich, L. A. (**1967**). *Speech Articulation and Perception* (Joint Publications Research Service, U.S. Dept. of Commerce, Washington, DC), 272 pp.

Kuznetsova, A., Bruun Brockhoff, P., and Haubo Bojesen Christensen, R. (**2016**). "lmerTest: Tests in linear mixed effects models," R package version 2.0-33.

Ladefoged, P. (**2003**). *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques* (Blackwell, Malden, MA), Chap. 4, pp. 75–103.

Ladefoged, P., and Maddieson, I. (**1996**). *The Sounds of the World's Languages* (Blackwell, Oxford, UK), 425 pp.

Laeufer, C. (**1992**). "Patterns of voicing-conditioned vowel duration in French and English," J. Phon. **20**(4), 411–440.

Lampp, C., and Reklis, H. (**2004**). "Effects of coda voicing and aspiration on Hindi vowels," J. Acoust. Soc. Am. **115**(5), 2540.

Lehiste, I. (**1970**). *Suprasegmentals* (MIT Press, Cambridge, MA), Chap. 2, pp. 6–53.

Lennes, M. (**2002**). "Total-duration-of-labeled-segments Praat script," http://www.linguistics.ucla.edu/faciliti/facilities/acoustic/praat.html (Last viewed September 16, 2015).

Lindau, M. (**1984**). "Phonetic differences in glottalic consonants," J. Phon. **12**, 147–155.

Lindblom, B. (**1967**). "Vowel duration and a model of lip-mandible coordination," Speech Transm. Lab. Q. Prog. Status Rep. **4**, 1–29.

Lisker, L. (**1957**). "Closure duration and the intervocalic voiced-voiceless distinction in English," Language **33**(1), 42–49.

Lisker, L. (**1974**). "On 'explaining' vowel duration variation," Glossa **8**(2), 233–246.

Luce, P. A., and Charles-Luce, J. (**1985**). "Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production," J. Acoust. Soc. Am. **78**(6), 1949–1957.

Maddieson, I. (**1997**). "Phonetic universals," in *The Handbook of Phonetic Sciences*, edited by W. J. Hardcastle and J. Laver (Blackwell, Cambridge, MA), Chap. 20, pp. 619–639.

Maddieson, I., and Gandour, J. (**1976**). "On 'explaining' vowel duration variation," UCLA Work. Papers Phon. **31**, 46–52, available at http://escholarship.org/uc/item/31f5j8m7.

Meyer, E. A. (**1903**). *Englische Lautdauer: Eine Experimentalphonetische Untersuchung* (*Duration of English Sounds: An Experimental Phonetic Study*) (Akademiska Bokhandeln, Uppsala, Sweden), 111 pp.

Moisik, S. R., and Esling, J. H. (**2014**). "Modeling the biomechanical influence of epilaryngeal stricture on the vocal folds: A low-dimensional model of vocal-ventricular fold coupling," J. Speech Lang. Hear. Res. **57**(2), S687–S704.

Mok, P. P. K. (**2012**). "Effects of consonant cluster syllabification on vowel-to-vowel coarticulation in English," Speech Commun. **54**(8), 946–956.

Munson, B., and Solomon, N. P. (**2004**). "The effect of phonological neighborhood density on vowel articulation," J. Speech Lang. Hear. Res. **47**(5), 1048–1058.

Nihalani, P. (**1975**). "Air flow rate in the production of stops in Sindhi," Phonetica **31**(3–4), 198–205.

Nooteboom, S. G., and Doodeman, G. J. N. (**1980**). "Production and perception of vowel length in spoken sentences," J. Acoust. Soc. Am. **67**(1), 276–287.

O'Brien, R. M. (**2007**). "A caution regarding rules of thumb for variance inflation factors," Qual. Quant. **41**(5), 673–690.

Ohala, M., and Ohala, J. J. (**1992**). "Phonetics universals and Hindi segment duration," in *Proceedings of the International Conference on Spoken Language Processing*, October 12–16, Banff, Alberta, Canada, pp. 831–834.

Öhman, S. (**1967**). "Peripheral motor commands in labial articulation," Speech Transm. Lab. Q. Prog. Status Rep. **8**, 30–63.

Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**(6), 693–703.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team. (**2016**). "nlme: Linear and nonlinear mixed effects models," R package version 3.1-128.
Piroth, H. G., and Janker, P. M. (**2004**). "Speaker-dependent differences in voicing and devoicing of German obstruents," J. Phon. **32**(1), 81–109.
Port, R. F. (**1981**). "Linguistic timing factors in combination," J. Acoust. Soc. Am. **69**(1), 262–274.
Port, R. F. (**1996**). "The discreteness of phonetic elements and formal linguistics: Response to A. Manaster Ramer," J. Phon. **24**(4), 491–511.
Port, R. F., Al-Ani, S., and Maeda, S. (**1980**). "Temporal compensation and universal phonetics," Phonetica **37**, 235–252.
Port, R. F., Dalby, J., and O'Dell, M. (**1987**). "Evidence for mora timing in Japanese," J. Acoust. Soc. Am. **81**(5), 1574–1585.
Port, R. F., and O'Dell, M. (**1985**). "Neutralization of syllable-final voicing in German," J. Phon. **13**(4), 455–471.
Port, R. F., and Rotunno, R. (**1979**). "Relation between voice onset time and vowel duration," J. Acoust. Soc. Am. **66**(3), 654–662.
Pye, S. (**1986**). "Word-final devoicing of obstruents in Russian," Cambridge Papers Phon. Exp. Ling. **5**, 1–10, available at http://www.worldcat.org/title/cambridge-papers-in-phonetics-and-experimental-linguistics/oclc/12291164.
Roettger, T. B., Winter, B., Grawunder, S., Kirby, J., and Grice, M. (**2014**). "Assessing incomplete neutralization of final devoicing in German," J. Phon. **43**, 11–25.
Ruhm, H. B., Mencke, E. O., Milburn, B., Cooper, W. A., Jr., and Rose, D. E. (**1966**). "Differential sensitivity to duration of acoustic signals," J. Speech Lang. Hear. Res. **9**(3), 371–384.
Sharf, D. J. (**1962**). "Duration of post-stress intervocalic stops and preceding vowels," Lang. Speech **5**(1), 26–30.
Sharf, D. J. (**1964**). "Vowel duration in whispered and in normal speech," Lang. Speech **7**(2), 89–97.
Shields, L. W., and Balota, D. A. (**1991**). "Repetition and associative context effects in speech production," Lang. Speech **34**(1), 47–55.
Shosted, R. K., and Chikovani, V. (**2006**). "Standard Georgian," J. Int. Phon. Assoc. **36**(2), 255–264.
Shosted, R. K., and Rose, S. (**2011**). "Affricating ejective fricatives: The case of Tigrinya," J. Int. Phon. Assoc. **41**(1), 41–65.
Shrager, M. (**2012**). "Neutralization of word-final voicing in Russian," J. Slav. Ling. **20**(1), 71–99.
Slowiaczek, L., and Dinnsen, D. (**1985**). "On the neutralizing status of Polish word-final devoicing," J. Phon. **13**(3), 325–341.
Smith, B. L., Hayes-Harb, R., Bruss, M., and Harker, A. (**2009**). "Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German," J. Phon. **37**(3), 257–275.
Solé, M.-J. (**2007**). "Controlled and mechanical properties in speech: A review of the literature," in *Experimental Approaches to Phonology*, edited by M.-J. Solé, P. Beddor, and M. Ohala (Oxford University Press, Oxford, UK), pp. 302–321.
Stott, L. H. (**1935**). "Time-order errors in the discrimination of short tonal durations," J. Exp. Psychol. **18**(6), 741–766.
Trullinger, R. W., and Emanuel, F. W. (**1983**). "Airflow characteristics of stop-plosive consonant productions of normal-speaking children," J. Speech Lang. Hear. Res. **26**(2), 202–208.
Turk, A., Nakai, S., and Sugahara, M. (**2006**). "Acoustic segment durations in prosodic research: A practical guide," in *Methods in Empirical Prosody Research*, edited by S. Sudhoff, D. Lenertova, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, and J. Schließer (De Gruyter, Berlin, Germany), pp. 1–28.
Umeda, N. (**1975**). "Vowel duration in American English," J. Acoust. Soc. Am. **58**(2), 434–445.
van Rooy, B., Wissing, D., and Paschall, D. D. (**2003**). "Demystifying incomplete neutralisation during final devoicing," South. Afr. Ling. Appl. Lang. Stud. **21**(1–2), 49–66.
Vicenik, C. (**2010**). "An acoustic study of Georgian stop consonants," J. Int. Phon. Assoc. **40**(1), 59–92.
Vicenik, C., and Jun, S.-A. (**2014**). "An autosegmental-metrical analysis of Georgian intonation," in *Prosodic Typology II: The Phonology of Intonation and Phrasing*, edited by S.-A. Jun (Oxford University Press, Oxford, UK), Chap. 6, pp. 154–186.
Warner, N., Jongman, A., Sereno, J., and Kemps, R. (**2004**). "Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch," J. Phon. **32**(2), 251–276.
Westbury, J. R. (**1983**). "Enlargement of the supraglottal cavity and its relation to stop consonant voicing," J. Acoust. Soc. Am. **73**(4), 1322–1336.
Wright, R. (**2004**). "A review of perceptual cues and cue robustness," in *Phonetically Based Phonology*, edited by B. Hayes, R. Kirchner, and D. Steriade (Cambridge University Press, Cambridge, UK), pp. 34–57.
Wright, R., Hargus, S., and Davis, K. (**2002**). "On the categorization of ejectives: Data from Witsuwit'en," J. Int. Phon. Assoc. **32**(1), 43–77.
Wysocki, T. (**2004**). "Acoustic analysis of Georgian stop consonants and stop clusters," Ph.D. thesis, University of Chicago, Chicago, IL.
Yu, A. C. L. (**2008**). "The phonetics of quantity alternation in Washo," J. Phon. **36**(3), 508–520.
Zimmerman, S. A., and Sapon, S. M. (**1958**). "Note on vowel duration seen crosslinguistically," J. Acoust. Soc. Am. **30**(2), 152–153.
Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., and Smith, G. M. (**2009**). *Mixed Effects Models and Extensions in Ecology with R* (Springer, New York), 574 pp.