Covariation of stop voice onset time across languages: Evidence for a universal constraint on phonetic realization

Eleanor Chodroff,[1,a)] **Alessandra Golden,**[2] **and Colin Wilson**[2]
[1]*Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, Illinois 60208, USA*
[2]*Department of Cognitive Science, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA*
eleanor.chodroff@northwestern.edu, agolden9@jhu.edu, colin.wilson@jhu.edu

Abstract: Stop consonant voice onset time (VOT) was examined in a typological survey of over 100 languages. Within broadly defined laryngeal categories (long-lag, short-lag, and lead voicing), VOT means were found to vary extensively. Importantly, the means for members of the same laryngeal series did not vary independently but instead were highly correlated across languages. The strong linear relations identified here cannot be reduced to previously reported ordinal relations, and provide evidence for a uniformity constraint on phonetic realization: within a language, each laryngeal specification must be realized in approximately the same way across stops of different places of articulation.

[DDO'S]
Date Received: September 11, 2018 **Date Accepted:** January 7, 2019

1. Introduction

While linguistic typology has traditionally focused on the areas of syntax, morphology, and phonology,[1] research over the past several decades has established that languages also vary in the fine-grained details of phonetic realization.[2,3] There are cross-linguistic differences in the release duration of aspirated stops such as [kʰ], the total duration and burst amplitude of ejective stops such as [k′], and the spectra of vowels and nasal consonants, among many other phonetic properties.[4–6] These findings lead naturally to the question of what universal constraints, either absolute or statistical, restrict phonetic realization across languages.

As in other empirical domains, many of the phonetic constraints that have been identified are *relational*. For example, while vowels vary in characteristic fundamental frequency (*f*0) across languages, high vowels (e.g., [i u]) have a higher intrinsic *f*0 than low vowels (e.g., [a]) in all languages studied so far.[7] Relational constraints have also been found for vowel duration (e.g., low vowels being intrinsically longer than high vowels), stop closure duration (e.g., bilabial stops having longer closures than velar stops), and *f*0 at vowel onset (e.g., onset *f*0 being higher after voiceless obstruents than after voiced obstruents).[8,9] In a landmark study of 18 diverse languages, each represented by multiple speakers and analyzed with a common method, Cho and Ladefoged[6] provided evidence for a relational constraint on stop voice onset time (VOT): within a language and laryngeal series, VOT generally increases with more posterior place of articulation (e.g., dorsals have longer values than labials and coronals).[8,10]

We conducted a much larger typological survey of stop VOT, gathering previously collected data from over 100 languages. Like Cho and Ladefoged,[6] we focused on stops in word-initial prevocalic position, but unlike them we included stops that are phonetically voiced (i.e., have lead voicing) in addition to those that are voiceless unaspirated or voiceless aspirated. In this paper, we describe the contents of the survey and present a sequence of analyses that refine our understanding of the relational constraints on this aspect of phonetic realization. We find that the VOT values of stops in the same laryngeal series are *linearly correlated* across languages to a striking degree. These statistical relations are highly unlikely to have arisen by chance and are not reducible to previously identified ordinal rankings (such as dorsal > labial, coronal). Their precise form indicates that, within a language, the fine-grained realization of each laryngeal specification is constrained to be approximately constant for stops of

---

[a)] Author to whom correspondence should be addressed.

different places of articulation—the cross-linguistic counterpart of a relational constraint that we have recently found to hold across speakers of a single language.[11]

2. Typological survey

We collected previously reported VOT values from a variety of sources (articles, theses, grammars, etc.) with the aim of representing every language for which stop VOT has been investigated. In a few cases, colleagues generously provided us with unpublished measurements (see the Acknowledgements). The data subset analyzed here excludes values from child, bilingual, heritage, and second-language speakers. We also excluded stops categorized in the original sources as ejective, glottalized, implosive, breathy ("voiced aspirated"), or emphatic. The full survey, together with the present subset and complete bibliographic information, has been made available to the research community through the Open Science Foundation (https://osf.io/ry3en/).[18]

The languages and language families represented in the survey, after applying exclusionary criteria, are listed in Table 1. Language families were identified from the original source or the World Atlas of Language Structures (WALS).[12] In total, there was data from 147 language varieties (in some cases closely related dialects such as U.S. and Canadian English) belonging to 36 families, including relevant samples from Cho and Ladefoged's survey.

Within a language variety, we relied upon the original sources to identify one or more laryngeal series: oral stops that have the same laryngeal setting (e.g., voiceless aspirated) but differ in place of articulation. In doing so we abstracted away from differences in terminology across sources; for example, the Dutch stops variously described as "voiceless" or "fortis" were assigned to the same series. We also categorized each stop as having one of three broad places of articulation: labial (bilabial), coronal (dental/alveolar/retroflex), or dorsal (palatal/velar/uvular). The decision to classify palatals as dorsal may be controversial but is unlikely to affect our results, as these stops made up approximately 2% of the data. We aggregated the VOT values for each reported combination of laryngeal series and place of articulation by averaging multiple measures from the same study and then averaging over studies of the same language variety. This ensured that each variety was represented in subsequent analyses by at most three values per series. (The averaging process did not obscure potentially large differences among sources: for most varieties we had one source, and for the 27 varieties with multiple sources the median standard deviation was 7 ms.)

Finally, we mapped each language-specific laryngeal series to one of three categories by comparing its grand average (i.e., the average of the place-specific means) to thresholds as follows: long-lag ($\overline{\text{VOT}} \geq 35\,\text{ms}$), short-lag ($35\,\text{ms} > \overline{\text{VOT}} \geq 0\,\text{ms}$), or lead ($0\,\text{ms} > \overline{\text{VOT}}$). For example, the Dutch stops [p t k] had averages of $\overline{\text{VOT}}_{\text{labial}} = 30\,\text{ms}$, $\overline{\text{VOT}}_{\text{coronal}} = 44\,\text{ms}$, and $\overline{\text{VOT}}_{\text{dorsal}} = 50\,\text{ms}$, respectively, therefore this series was categorized as long-lag ($\overline{\text{VOT}} = 41$). While other thresholds yielded similar results in subsequent analyses, we selected 35 ms because of its proximity to the relevant auditory boundary in humans and other species and 0 ms as a natural division between phonetically voiced and unvoiced stops.[13]

3. Analysis and results

In line with previous findings, VOT values within each of the laryngeal categories varied extensively across languages (see Fig. 1). For the long-lag series, place-specific averages ranged from 14 to 117 ms (labials), 18 to 130 ms (coronals), and 41 to 154 ms (dorsals). VOT averages for the short-lag series ranged from 0 to 32 ms (labials), 3 to 40 ms (coronals), and 5 to 56 ms (dorsals); for series categorized as lead the ranges were -161 to -20 ms (labials), -177 to -8 ms (coronals), and -144 to 8 ms (dorsals). (The ranges overlap somewhat because laryngeal categories were determined by averaging over place, as described earlier.)

A sequences of analyses was performed to investigate the nature and strength of constraints on this variation. We first calculated the percentage of stop pairs from the same language variety and laryngeal series that conform to the ordinal ranking $\overline{\text{VOT}}_{\text{dorsal}} > \overline{\text{VOT}}_{\text{coronal}} > \overline{\text{VOT}}_{\text{labial}}$.[6,10] Considering all of the categories together, the percentages were 95% (dorsal > labial), 91% (dorsal > coronal), and 72% (coronal > labial). This largely confirms previous observations that dorsal stops have the highest values while the relative order of coronals and labials is more variable.[14] Evidence for the same ordinal relation was also found within each laryngeal category (82% mean conformity), in particular for the short-lag stops (dorsal > labial: 99%, dorsal > coronal: 96%, coronal > labial: 82%).

Table 1. Languages and number of data points (after averaging within place of articulation) for each language family, with families ordered by decreasing size in the survey.

Family	Languages	N
Indo-European	Afrikaans, Armenian (Eastern), Assamese, Bengali, Catalan, Croatian, Danish, Dutch, English, French, Gaelic (Scots), German, Greek (Modern), Hindi, Icelandic, Italian, Kurmanji, Marathi, Nepali, Norwegian, Pahari, Panjabi, Pashto, Persian, Polish, Portuguese (Brazilian), Portuguese (European), Russian, Serbian, Sindhi, Spanish, Swedish, Welsh	337
Sino-Tibetan	Bunun, Burmese, Cantonese, Fukienese, Galo, Hakha Lai, Hakka, Hokkien, Karen (Sgaw), Khonoma Angami, Kurtop, Mandarin, Stau, Wu (Shanghainese)	84
Niger-Congo	Bowiri, Igbo, Shekgalagari, Swati, Tswana, Xhosa, Zulu	39
Austronesian	Belep, Madurese, Malay, Tsou, Yapese	21
Afro-Asiatic	Amharic, Arabic, Dahalo, Hebrew (Modern), Musey	20
Na-Dene	Apache (Western), Hupa, Navajo, Tlingit	19
Quechuan	Quechua (Bolivian), Quechua (Cuzco), Quichua	15
Mayan	Itzaj Maya, Mam (Southern), Mopan Maya, Tz'utujil, Yukateko Maya	14
Altaic	Azerbaijani, Turkish	12
Dravidian	Tamil, Telugu	12
Tai-Kadai	Tai Khamti, Thai	12
Austro-Asiatic	Pnar, Remo	11
Oto-Manguean	Mazatec (Jalapa), Zapotec (Yalalag)	10
Uralic	Finnish, Hungarian	9
Uto-Aztecan	Paiute (Northern), Shoshoni, Ute	9
Tupian	Arara, Munduruku	8
Burushaski	Burushaski	6
Japanese	Japanese	6
Kartvelian	Georgian	6
Kordofanian	Moro	6
Northwest Caucasian	Kabardian	6
Pama-Nyungan	Warlpiri, Yan-Nhangu	6
Ticuna	Ticuna	6
Wakashan	Kwak'wala	6
Eskimo-Aleut	Aleut (Eastern), Aleut (Western)	4
Tucanoan	Waimaha	4
Algic	Ojibwe	3
Chapacura-Wanham	Wari'	3
Creole	Hawaiian Creole	3
Ijoid	Defaka	3
Korean	Korean	3
Muskogean	Chickasaw	3
Nakh-Daghestanian	Udi	3
Salishan	Montana Salish	3
Tangkic	Kayardild	3
Arauan	Banawa	2

Ordinal ranking is a familiar but relatively imprecise relation. For example, $\overline{\text{VOT}}_{\text{dorsal}} > \overline{\text{VOT}}_{\text{labial}}$ is consistent with a typology in which there are both very small (~5 ms) and very large (~75 ms) differences between these two place-specific means. Therefore, we repeated the comparisons above with the Pearson correlation coefficient, which evaluates the degree to which VOT means are linearly related across languages.

Overall, the correlations among stops from the same laryngeal series were extremely high and significant (adjusted *ps* < 0.004): 0.99 (labial–coronal), 0.98 (coronal–dorsal), and 0.98 (labial–dorsal); Fig. 1 provides correlations within each of the laryngeal categories. For long-lag stops, the correlations found across language varieties were comparable to those reported previously across speakers of American English, which ranged from 0.95 to 0.96 in isolated speech and 0.77 to 0.83 in connected speech.[11] Correlations among stops with lead VOT were also quite strong.

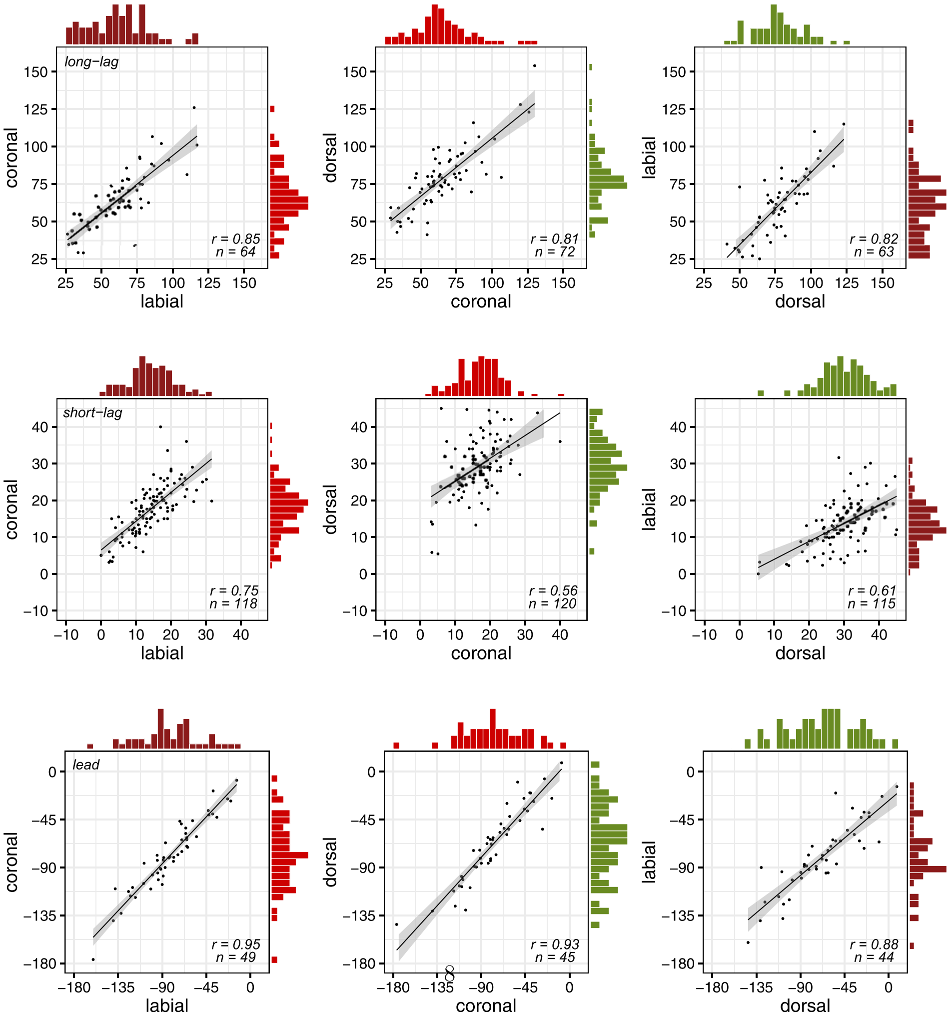

Fig. 1. (Color online) Variation and covariation of VOT means (ms) across languages for long-lag stops (top row), short-lag stops (middle row), and stops with lead voicing (bottom row). Figure axes were extended 10 ms from the 0 and 35 ms thresholds. Marginal histograms reflect the range of variation in VOT means for a given place of articulation; these differ slightly across plots due to inventory gaps. Gray shading indicates the local confidence interval around the best-fit linear regression line. The Pearson correlation coefficient and number of data points are reported in each subfigure.

Correlations among short-lag stops were weaker than those for the other two categories, but nevertheless moderate to moderately strong. The following additional analyses investigated the nature of the linear relations that underlie the correlations, the source of the relative weakness of relations in the short-lag category, and whether the correlations are reducible to ordinal rankings.

The correlations among place-specific VOT means within a laryngeal series could, in principle, reflect approximately constant differences across language varieties (e.g., $\overline{\mathrm{VOT}}_{\mathrm{dorsal}} \sim 15\,\mathrm{ms} + \overline{\mathrm{VOT}}_{\mathrm{labial}}$). We evaluated the hypothesis of constant differences by comparing nested linear regression models with the likelihood ratio test. The constant difference hypothesis corresponds to a model with one free parameter, the intercept: $\overline{\mathrm{VOT}}_{\mathrm{place1}} \sim \beta_0 + 1 * \overline{\mathrm{VOT}}_{\mathrm{place2}}$. The alternative model had both an intercept and a scaling factor, allowing the difference to vary linearly with one of the VOT means: $\overline{\mathrm{VOT}}_{\mathrm{place1}} \sim \beta_0 + \beta_1 * \overline{\mathrm{VOT}}_{\mathrm{place2}}$. When all of the laryngeal categories were analyzed together, the constant difference model was a better fit ($ps > 0.15$). The estimated differences were 3 ms between coronals and labials, 13 ms between dorsals and coronals, and 16 ms between dorsals and labials.

The same result was found for stops in the lead-voicing series, with similar estimated differences of 3 ms (coronal–labial), 10 ms (coronal–dorsal), and 12 ms (dorsal–labial). However, relationships among VOT means within the long-lag category were better fit by the more complex model (adjusted $ps < 0.004$); specifically, the results indicated that differences tend to decrease at the upper end of the VOT range, perhaps reflecting physiological limits on place-specific modulation. For short-lag stops, the relationship between labials and coronals was also a combination of difference and scaling ($p < 0.004$), but the intercept-only model was superior for the other comparisons (with estimated differences of 14 ms for coronal–dorsal and 17 ms for dorsal–labial). In summary, the hypothesis of constant differences can account for many of the VOT correlations found in our survey, but some require scaling as well.

Turning now to the weaker correlations among short-lag stops, one hypothesis is that these reflect laryngeal underspecification (i.e., absence of a phonetic target could give rise to less systematic realization).[15] Alternatively, they could be artifacts of the smaller VOT range covered by the short-lag category relative to the long-lag and lead categories. Consistent with the latter alternative, the correlations in the other two categories were appreciably reduced when computed over a 35 ms subrange centered on their median values (rs for long-lag subrange: labial–coronal = 0.51, coronal–dorsal = 0.48, labial–dorsal = 0.47; rs for lead subrange: labial–coronal = 0.75, coronal–dorsal = 0.73, labial–dorsal = 0.55). Relatedly, the short-lag correlations involving dorsals were noticeably weaker than the labial–coronal relation. This could indicate a ceiling effect that specifically impacts the short-lag stops with the longest VOTs (though the findings reported here were robust to minor changes in the short-lag/long-lag boundary). Therefore, while underspecification remains a potential explanation for the finding that stops in this category are less systematically related across languages, the present results are equivocal.

Finally, we were interested in whether the linear correlations, regardless of their strength, could be reduced to the ordinal relations discussed earlier. We tested this by repeatedly resampling VOT averages for each place pair within each laryngeal category, subject to the constraint that the relevant canonical ranking was obeyed, and using the samples to compute a distribution of hypothetical correlations (over $R = 10\,000$ replications). All of the empirical correlations were higher than would be expected from these resampling distributions (adjusted ps < 0.004). Thus, rather than focusing exclusively on coarse-grained ordering, future research into constraints on phonetic variation should consider detailed linear relations among VOT values and other measurements.

4. Discussion

In a sample of over 100 languages from diverse families, variation in stop VOT was found to be both extensive and highly structured. As has been observed previously, place-specific VOT means largely adhere to an ordinal relationship: within a language and laryngeal series, values generally increase with more posterior places of articulation. We have shown that an even tighter quantitative relation holds: the VOT means of stops in the same laryngeal series are linearly correlated across languages, to a striking degree within the long-lag and lead categories and more modestly but significantly within the short-lag category. These findings parallel patterns of VOT covariation within American English, such that speakers with long mean VOTs for [kʰ] also have long means for [pʰ] and [tʰ].[11] Close investigation of the linear relations found in our typological survey indicates that differences among place-specific means are roughly constant across languages for the short-lag and lead categories, whereas the differences decrease somewhat at the upper range of long-lag stops.

The specific relations identified here point to a *uniformity* constraint that applies to the phonetic realization of stop laryngeal distinctions. Within each language-specific series, stops are constrained to have approximately the same phonetic targets for their shared laryngeal specification (e.g., the relative timing of oral and glottal gestures, the duration/magnitude of any glottal opening gestures, vocal fold tension and airflow rate, and other relevant properties).[16] On this account, consistent VOT differences across place of articulation must largely reflect aerodynamic and other independent factors, rather than implicating substantial differences in phonetic targets.[6,8]

We should emphasize that the uniformity constraint reins in cross-linguistic (and indeed cross-speaker) variation that might otherwise occur freely. For instance, Cho and Ladefoged[6] consider the possibility that "the value [aspirated] would correspond to one VOT target in the context [velar] and another when it is in the context [labial]" (Cho and Ladefoged, p. 227).[6] Indeed, our findings indicate that it would be physically possible for one language/speaker to have respective VOT means for [pʰ] and [kʰ] of 50 and 80 ms, and another language/speaker to have the reverse pattern, yet we have found this type of departure from uniformity to be quite rare.

We acknowledge that the VOT relations identified here, while quite strong, are not perfect. Deviations from uniformity may reflect purely methodological or properly linguistic factors. The studies surveyed here differed in the materials collected, speaker age and gender characteristics, typical speaking rate, and presumably VOT segmentation methods. These factors could potentially exaggerate the cross-linguistic variation in VOT means and, more generally, could impact the assessment of covariation in multiple ways. To give one example, the VOT averages for Marathi long-lag stops appear to reverse the typical ordering of labial and dorsal: [p^h] = 110 ms, [t^h] = 81 ms, [k^h] = 103 ms. However, several of our sources for this language happened to exclude [p^h]: the apparent reversal may be due entirely to computing averages from different speakers, rates, etc., across unbalanced samples.

The extent to which VOT means adhere to the best-fit regression lines may also be modulated by linguistic factors, and some languages may flout the constraint for reasons that are currently unknown. For instance, Mopan Maya has reported VOT means of 73 ms for its labial stop, 34 ms for the coronal, and 50 ms for its dorsal.[17] The ranking of the labial stops with respect to the others is typologically unexpected, and the relevant points fall far from the regression lines. Intriguingly, this language does not have a pure voicing or aspiration contrast (its other stops are ejectives and implosives), and this could plausibly license greater deviation from uniformity. However, several other languages that lack such a contrast do conform to the expected ordinal and linear relations (e.g., Itzaj Maya, Tsou, Tz'utujil, among others).

The relationship of VOT among place of articulation is just one instance of a relational universal. Additional relational universals include the relations between *f*0 and vowel height,[7] *f*0 and preceding stop voicing, vowel duration and vowel height, vowel duration and following stop voicing, among others.[8] As with VOT place differences, these relationships may be defined linearly, which would inherently encode the ordinal rank. Indeed, across languages and talkers, the *f*0 means of high vowels [i] and [u] are not only consistently higher than that of the low vowel [a], but also almost perfectly correlated [rs > 0.98; as calculated from the means reported by Whalen and Levitt, 1995 (Ref. 7)]. Such non-arbitrary differences in phonetic realization can be attributed to articulatory or aerodynamic factors but only under the assumption that speakers have uniform targets for the relevant phonetic properties.

In this paper, we have marshaled a large body of data in support of a relational universal in phonetics, and have provided new quantitative measures of the nature and strength of this universal. The phonetic realizations of stops in the same laryngeal series, as measured by the acoustic parameter of VOT, do not vary independently across languages. They are yoked together in a pattern of linear covariation that suggests an underlying uniformity constraint. Further research is required to investigate whether this pattern holds across an even larger sample of diverse languages, whether it is modulated by language family or phonetic inventory, whether it is found across speakers within languages other than American English, and whether the uniformity constraint applies to other types of phonetic variation.

Acknowledgments

We are grateful to Andries Coetzee, Karee Garvin, Paul Morris, and Priscilla Shin for providing us with unpublished VOT data, and to Matthew Faytak, Keith Johnson, Eric Raimy, Paul Smolensky, Doug Whalen, and participants of the Berkeley Phorum, Northwestern Phonatics group, and LSA 2018 Annual Meeting for helpful comments and suggestions. This research was partially supported by a JHU Distinguished Science of Learning Pre-Doctoral Fellowship to E.C.

References and links

[1] J. H. Greenberg, "Some universals of grammar with particular reference to the order of meaningful elements," in *Universals of Language* (MIT Press, Cambridge, MA, 1963), Vol. 2, pp. 73–113.

[2] P. Ladefoged, *Preliminaries to Linguistic Phonetics* (The University of Chicago Press, Chicago, IL, 1971).

[3] P. A. Keating, "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, 1985), pp. 115–132.

[4] M. Lindau and P. Wood, "Acoustic vowel spaces," UCLA Work. Papers Phonetics **38**, 41–48 (1977).

[5] M. Lindau, "Phonetic differences in glottalic consonants," J. Phonetics **12**(2), 147–155 (1984).

[6] T. Cho and P. Ladefoged, "Variation and universals in VOT: Evidence from 18 languages," J. Phonetics **27**(2), 207–229 (1999).

[7] D. H. Whalen and A. G. Levitt, "The universality of intrinsic f0 of vowels," J. Phonetics **23**(3), 349–366 (1995).

[8] I. Maddieson, "Phonetic universals," in *The Handbook of Phonetic Sciences*, edited by W. J. Hardcastle and J. Laver (Blackwell Publishers, Oxford, 1997), pp. 619–639.

[9] A. S. House and G. Fairbanks, "The influence of consonant environment upon the secondary acoustical characteristics of vowels," J. Acoust. Soc. Am. **25**(1), 105–113 (1953).

[10]E. Fischer-Jørgensen, "Acoustic analysis of stop consonants," Le Maître Phonétique **32**(69), 42–59 (1954), available at http://www.jstor.org/stable/44705403.

[11]E. Chodroff and C. Wilson, "Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English," J. Phonetics **61**, 30–47 (2017).

[12]M. S. Dreyer and M. Haspelmath, eds., *WALS Online* (Max Planck Institute for Evolutionary Anthropology, Leipzig, 2013).

[13]P. K. Kuhl and J. D. Miller, "Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants," Science **190**(4209), 69–72 (1975).

[14]G. Docherty, *The Timing of Voicing in British English Obstruents* (Walter de Gruyter, Berlin, 1992).

[15]J. Beckman, M. Jessen, and C. Ringen, "Empirical evidence for laryngeal features: Aspirating vs. true voice languages," J. Linguistics **49**(2), 259–284 (2013).

[16]A. Löfqvist, "Acoustic and aerodynamic effects of interarticulator timing in voiceless consonants," Lang. Speech **35**(1–2), 15–28 (1992).

[17]A. Burnett-Deas, "VOT of ejectives, implosives, and plain stops in Yukateko, Mopan, and Itzaj Maya," Master's thesis, Southern Illinois University, Carbondale, 2009.

[18]E. Chodroff, A. Golden, and C. Wilson, "Cross-linguistic voice onset time," https://osf.io/ry3en/ (Last viewed January 18, 2019).