Acoustics of the four-way laryngeal contrast in Drenjongke (Bhutia): Observations and implications*

Abstract

Drenjongke is a Tibeto-Burman language spoken in Sikkim, India, whose phonetic properties are understudied. This language is reported to have a four-way laryngeal contrast: aspirated, voiceless, voiced, and "devoiced" (van Driem 2016). An acoustic analysis of twelve Drenjongke speakers shows that in addition to differences in VOT, there are systematic differences in F0 and F1 in the following vowel. Our analysis further suggests that high F1 after devoiced consonants is controlled, rather than being an automatic consequence of long VOT. We conclude that Drenjongke speakers use at least three acoustic dimensions (VOT, F0 and F1) to distinguish the four-way laryngeal contrast.

1 Introduction

Drenjongke is a Tibeto-Burman language spoken in Sikkim, India.[1] Although there is an impressionistic description of this language by van Driem (2016), not much is known about the phonetic nature of this language. This paper therefore examines one aspect of this language in close detail: its four-way laryngeal contrast, which is cross-linguistically not very common.[2] According to van

*This work is supported by Strategic Japanese-Swiss Science and Technology Programme of JSPS and SNSF. We thank Donna Erickson, Fuminobu Nishida, Jeremy Perkins, Julián Villegas, the audience of TAL 2018, and two anonymous reviewers for their comments. We would also like to thank George van Driem and Kunzang Namgyal who helped us find Drenjongke speakers in the Gangtok area, and last but not least, the Drenjongke participants who shared the knowledge of their language. All remaining errors are ours. A much shorter version of this paper is appearing in the proceedings of *ICPhS 2019*.

[1]This language is also known as Bhutia, Dandzongka, Danjongka, Danyouka, Denjong, Denjongkha, Denjongke, Denjongpa, Denjonka, Denjonke, Denzong Ke, Denzongke, Denzongpe Ke, Denzongpeke, Dranjoke, Dranjongke, Drendzongké, Lachengpa, Lachungpa, Sikami, Sikkim Bhotia, Sikkim Bhutia, and Sikkimese, according to ethonologue (`https://www.ethnologue.com/language/sip`, accessed Jan. 2019). We use "Drenjongke" at the request of the informants; "Bhutia" is the official name recognized by the Indian government.

[2]Indo-Iranian languages show up to a five-way laryngeal contrast (Hussain 2018). Dzongkha, the national language of Bhutan which is closely related to Drenjongke, exhibits a four-way laryngeal contrast as well. Its acoustic characteristics are recently examined by Lee & Kawahara (2018).

Driem (2016: 4), this language has aspirated, voiceless, voiced and "devoiced" obstruents (see also Yliniemi 2016 who touches on this contrast).[3] The first two categories are classified as "H-register" consonants, while the last two categories are classified as "L-register" consonants by van Driem. Some minimal quadruplets are shown below in (1), where following van Driem (2016), devoiced consonants are shown with an apostrophe. The data are based on our field notes.

(1) The four-way laryngeal contrast in Drenjongke

a. bilabial: [pʰo] 'stomach' vs. [pø] 'incense' vs. [bu] 'middle' vs. [b'u] 'son'
b. alveolar: [tʰoŋ] 'to see' vs. [tøŋ] 'to show' vs. [do] 'stone' vs. [d'o] 'touch'
c. retroflex: [ʈʰom] 'bazaar' vs. [ʈoŋ] 'to kill' vs. [ɖu] 'dragon' vs. [ɖ'u] 'boat'
d. velar: [kʰi] 'dog' vs. [ki] 'peace' vs. [gæ] 'eight' vs. [g'æ] 'row'

Particularly intriguing is the last category, "devoiced," whose acoustic properties are not clear even from van Driem's (2016) description. The current experiment thus explored how the four types of laryngeal categories are distinguished acoustically, which is our descriptive goal. Moreover, to contribute to current theoretical debates in phonetics, based on the current results, we discuss two issues that are actively debated in the literature: (1) whether F0 perturbation after certain types of stops is controlled or automatic, and (2) how abstract phonological features are with respect to their phonetic realizations.

2 Method

The data reported in the current paper is based on the fieldwork in Sikkim, India, which was conducted in the summer of 2017.

2.1 Speakers

Twelve native speakers of Drenjongke participated in the recording session. They were all school teachers from primary and secondary schools. All the speakers spoke Nepali and English in addition to Drenjongke (there are, unfortunately, no monolingual Drenjongke speakers). The age ranged from 25 years old to 55 years old, most of them being between 36 and 45 years old. Consent forms and demographic information were collected from each speaker before the recording session. Each participant was compensated for their time. Demographic information of these speakers is summarized in Table 1.

[3]The term "devoiced" does not imply the presence of a synchronic process of underlyingly voiced consonants becoming voiceless.

Table 1: The speakers' demographic information.

Speaker 1	East Sikkim, Barapathing	Female
Speaker 2	North Sikkim, Kati Longchok	Female
Speaker 3	East Sikkim, Sibik Susty (?)	Male
Speaker 4	West Sikkim, Tashiding Lumpa	Male
Speaker 5	Gangtok, Lamalen	Male
Speaker 6	East Sikkim, Chandmani Gangtok	Male
Speaker 7	North Sikkim, Lachung	Male
Speaker 8	East Sikkim, Machong	Male
Speaker 9	West Sikkim, Tashiding	Male
Speaker 10	North Sikkim, Tumlong	Male
Speaker 11	Gangtok, Tashiding	Male
Speaker 12	West Sikkim, Tashiding	Male

2.2 Recording

Within each recording session, each speaker read (i) typical syllables that appear in Drenjongke, (ii) words in isolation, and (iii) words in a frame sentence. This paper focuses on the analysis of syllables as read from a syllable transcription, in which all Drenjongke consonants were pronounced with a following [a]. Having always the vowel [a] allows us to control for lexical factors that may affect phonetic implementation patterns, as well as the intrinsic effects of vowel height on F0 and F1 (see e.g. Whalen & Levitt 1995 for the former; Johnson 2003 for the latter).[4] The order of the syllables was randomized, and the speakers repeated the list five times. All recordings were made using a Shure WH30XLR head-worn microphone and a TASCAM recorder (DR100-MK). The stimuli were presented in the Tibetan script using Keynote on a Macintosh computer. The target of the current analysis included stop consonants from four places of articulation (bilabial, alveolar, retroflex, and velar—see (1)), although the current analysis pools data from the different places of articulation.[5]

2.3 Acoustic analysis

Figure 1 provides some representative alveolar tokens of the four-types of laryngeal contrasts. Aspirated consonants are realized with long lag VOT (Voice Onset Time) (Figure 1 (a)), voiceless

[4]While we believe that syllabary reading offers a reasonable dataset in order to take an initial stab at exploring the acoustics of this understudied contrast, we also acknowledge that it is necessary to extend our analysis to the acoustics of the contrast in more naturalistic contexts. It is also necessary to analyze this four-way laryngeal contrast in intervocalic position in order to explore durational cues, which are known to correlate with a laryngeal contrast cross-linguistically (Chen 1970; Kingston & Diehl 1994; Lisker 1957; Port & Dalby 1982).

[5]We examined the effects of the different places of articulation, but did not find any interesting differences.

consonants are realized with short lag VOT (Figure 1 (b)), and voiced consonants are realized with prevoicing during closure (i.e. negative VOT), (Figure 1 (c)). Interestingly, devoiced consonants are variably realized with either prevoicing (Figure 1 (d)) or positive VOT (Figure 1 (e)). Measurements were made of the duration of these (negative and positive) VOT—duration of closure voicing for negative VOT and duration between the stop release and the onset of voicing for the following vowel for positive VOT (see Abramson & Whalen 2017). In addition, since F0 and F1 are known to correlate with a laryngeal contrast (Kingston & Diehl 1994, 1995; Lisker 1975, 1986 among many others), a 20 ms analysis window was created from the onset of the following vowel. Average F0 and F1 values were calculated within these analysis windows. These analyses were automated using the scripting function of Praat (Boersma 2001).

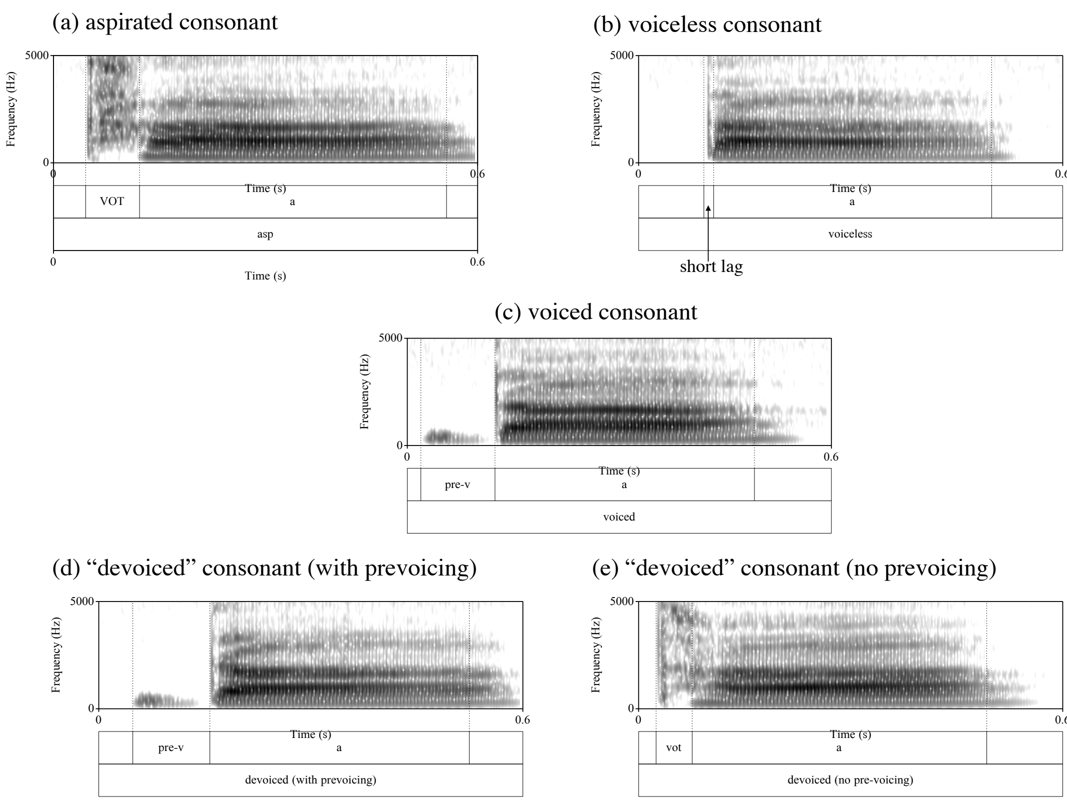


Figure 1: Representative tokens of the four-way laryngeal contrast in Drenjongke. Devoiced consonants can be realized either with prevoicing or positive VOT.

3 Results

Figure 2 is a violin-plot which shows the distribution of positive and negative VOT values for the four-way laryngeal contrast. A violin-plot shows the standardized probability density distributions, and hence is suited to illustrate data with bi-modal distribution.

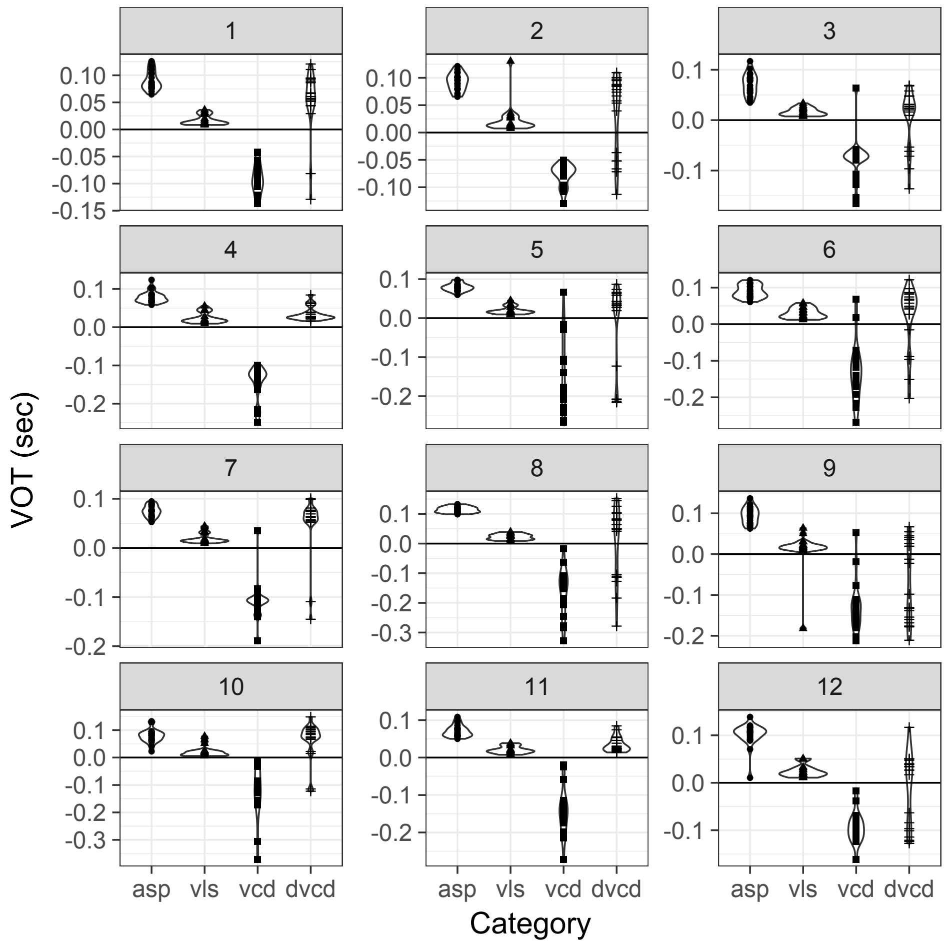


Figure 2: Distributions of negative and positive VOT for the four-way laryngeal contrast. Each dot represents each measurement point. Different panels show the twelve different speakers.

Aspirated consonants have long VOT (distributing around ca. 100 ms). Voiceless consonants have short-lag VOT (distributing around ca. 20 ms). Voiced consonants usually have pre-voicing (i.e. negative VOT). There are a few exceptional tokens of voiced stops with positive VOT (as observed in the data of Speakers 3, 5, 6, 7, 9), perhaps because these tokens were read utterance-initially and it is hard to initiate glottal vibration in this position (Westbury & Keating 1986). Devoiced consonants are realized either with pre-voicing or positive VOT. Speakers 4 and 11 are the only speakers who do not show this variation, who produced all the tokens with positive VOT. It seems to be the case that, given this variability, VOT alone cannot be used as a reliable acoustic cue for devoiced consonants in Drenjongke.[6]

Figure 3 is a boxplot which shows the F0 values in the following [a] after the four different types of consonants; the white circles represent the means in each condition. We observe that aspirated and voiceless consonants generally show higher F0 compared to voiced and devoiced consonants. ANOVA with TukeyHSD multiple comparison tests shows that the difference between aspirated and voiceless consonants and the difference between voiced and devoiced categories are not significant, but all the other comparisons are significant (voiceless vs. voiced: $p < .01$; all other differences: $p < .001$). This result accords well with van Driem's (2016) characterization of these types of consonants: aspirated and voiceless consonants are "H-register" consonants, whereas voiced and devoiced consonants are "L-register" consonants.

[6]We note that Abramson & Whalen (2017), a state-of-the-art summary article on the status of VOT, explicitly declare that VOT is not a magical phonetic measure that successfully distinguishes all laryngeal contrasts. It is therefore not too surprising to find a laryngeal category which cannot be uniquely defined based on VOT.

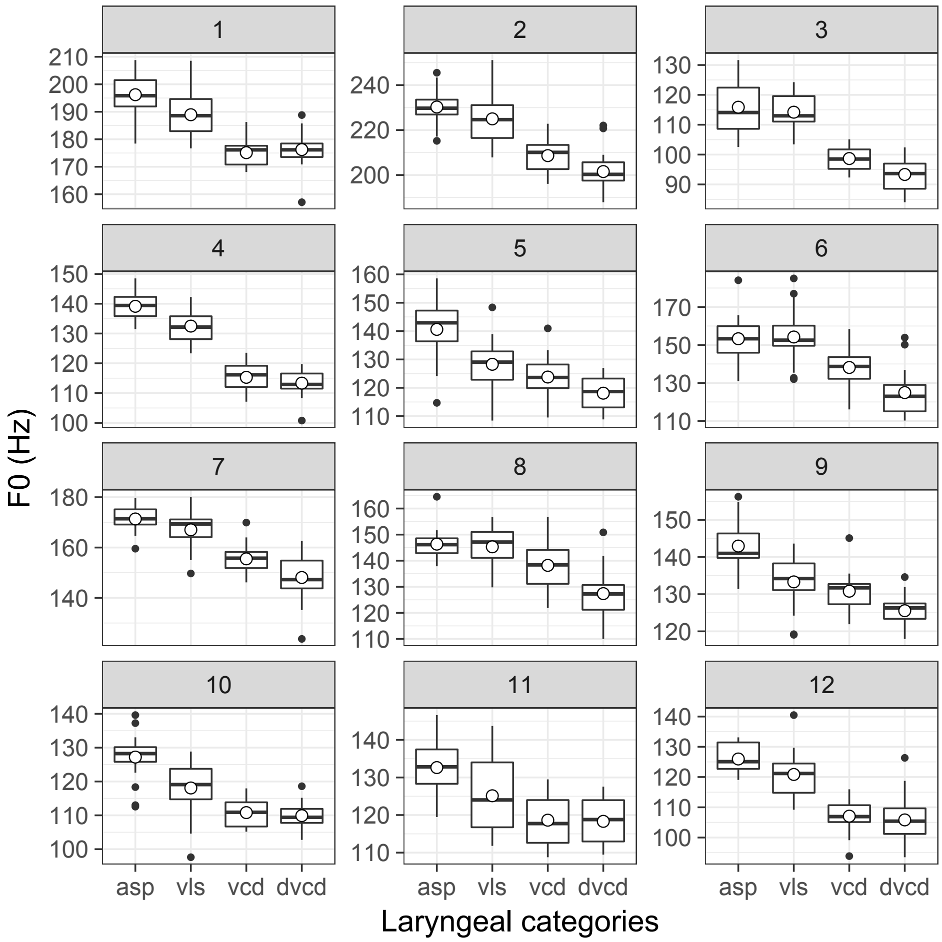


Figure 3: A boxplot representation of the F0 values in the following [a] vowel. The white circles represent the means in each condition. The first two speakers are female.

Figure 4 shows the F1 values of the following vowels. We observe that voiceless and voiced consonants show generally low F1, whereas aspirated and devoiced show high F1. Devoiced consonants, whose VOT profiles are rather variable (Figure 2), may thus instead be characterized as consonants with low F0 and high F1. ANOVA with TukeyHSD multiple comparisons shows that all the differences are significant ($p < .001$), which means that aspirated consonants overall show even higher F1 than devoiced consonants, and voiced consonants show even lower F1 than voiceless consonants. We note, however, there are speakers (e.g. Speakers 8 and 10) for whom F1 is higher after devoiced consonants than after aspirated consonants—there thus seems to be inter-speaker differences in this regard.

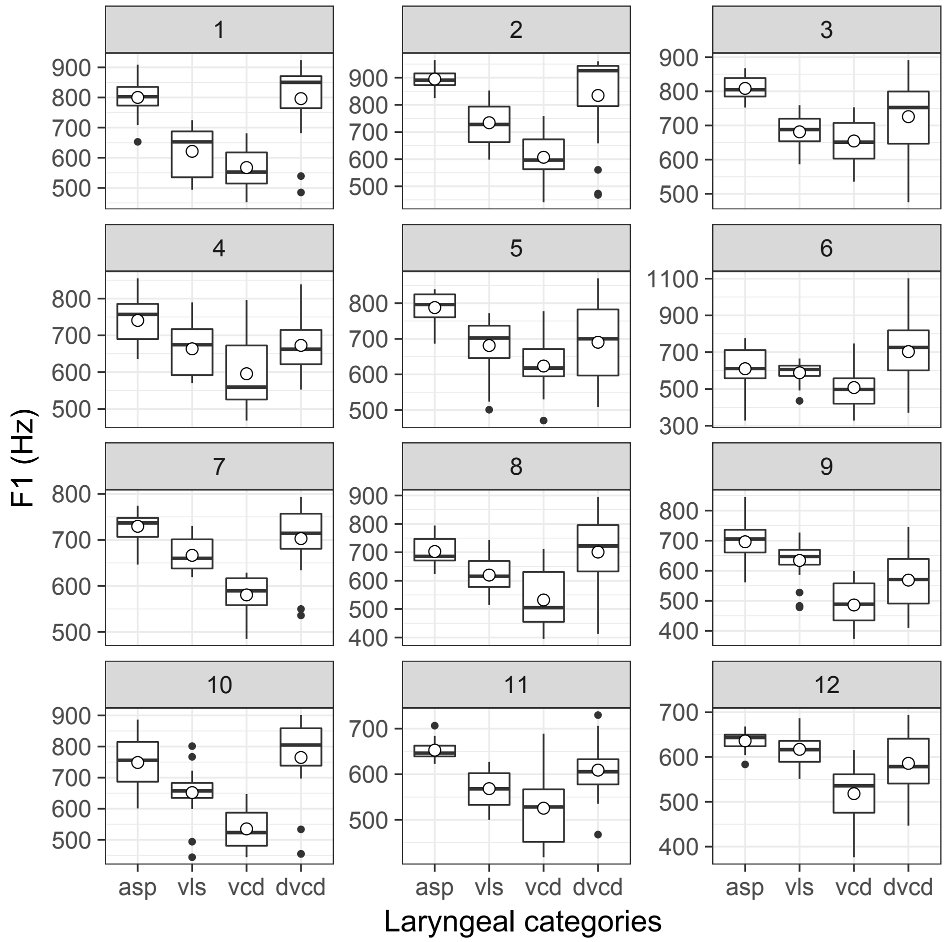


Figure 4: A boxplot representation of the F1 values in the following [a] vowel. The white circles represent the means in each condition.

To the extent that high F1 provides an important acoustic cue for "devoiced" consonants, one question that arises is whether high F1 values after devoiced consonants are intended (or controlled) or consequences of long VOT (which are exhibited by some tokens). Since F1 is correlated with the openness of the oral cavity (Johnson 2003; Kawahara et al. 2017; van Summers 1987), consonants with long VOT can show higher F1, because by the time F1 becomes measurable, the oral cavity is open more widely. To address this possibility, Figure 5 shows, for devoiced consonants, the correlation between F1 and VOT values, separately analyzed by whether VOT values are negative or positive.

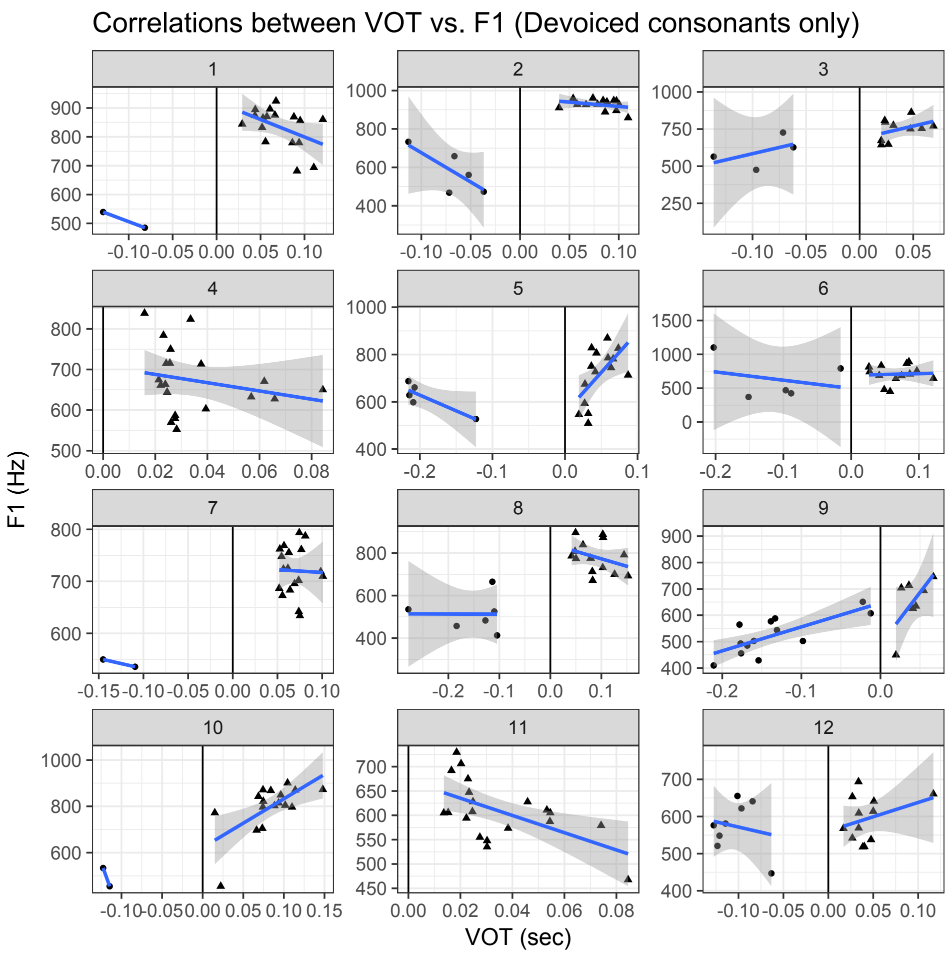


Figure 5: Correlation between F1 and VOT, separately analyzed depending on whether VOT is negative or positive (devoiced consonants only).

If high F1 is an automatic consequence of long VOT covering the opening phase of the oral cavity, there should be a positive correlation between F1 and VOT. However, this is true only for Speakers 5, 9 and 10. (The positive slope observed for Speaker 12 is most likely a spurious correlation arising from one outlier.) The rest of the speakers show either negative or no correlations, as summarized in Table 2. We thus suspect that high F1 after devoiced consonants is a consequence of an intended articulatory gesture; high F1 is particularly important to distinguish devoiced consonants from voiced consonants, both of which have low F0 in the following vowels.

One remaining question is how the high F1 values after "devoiced consonants" are achieved in this language.[7] One possibility is jaw movement: the Drenjongke speakers may be opening their

[7]Donna Erickson reminded us that anatomical connection is such that more opened jaw is associated with lower larynx, which in turn results in lowered F0 (Erickson et al. 2017), but more open jaw is associated with higher F1 (Johnson 2003; Kawahara et al. 2017; van Summers 1987). This may mean that there may be an additional articulatory mechanism for voiced consonants to achieve low F0 and low F1.

Table 2: Pearson correlation coefficients between F1 values and VOT. Devoiced consonants with positive VOT values only.

Speakers	*r*	Speakers	*r*
1	-0.46	7	-0.03
2	-0.31	8	-0.34
3	0.01	9	0.64
4	-0.23	10	0.64
5	0.59	11	-0.58
6	0.05	12	0.32

mouth more quickly after devoiced consonants, as English speakers do for voiceless consonants (van Summers 1987). Alternatively, devoiced consonants may be accompanied by pharyngeal constriction: while constriction in the oral cavity generally lowers F1 (Johnson 2003), Al-Tamimi (2017) shows that vowels in pharyngealized context show higher F1 in Jordianian and Moroccan Arabic. Tongue root retraction, which is often associated with pharyngeal constriction, also shortens the oral cavity behind the constriction, which would raise Helmholtz resonation, i.e., F1 (Johnson 2003; Kawahara et al. 2017). Overall, the question of how F1 is achieved for devoiced consonants should be addressed by a future articulatory study.

4 Discussion

In addition to revealing the acoustic correlates of the Drenjongke laryngeal contrast, the current data allows us to address two theoretical questions that are currently debated. One is whether effects of onset consonants on the F0 of the following vowels (Figure 3) are based on discrete phonetic/phonological categories or continuous VOT values (Chodroff & Wilson 2018; Clayards 2018; Dmitrieva et al. 2015; Kirby 2018; Kirby & Ladd 2016). This question is important as it bears on the question of how automatic/controlled F0 perturbation is. If F0 perturbation is an automatic consequence of laryngeal gestures associated with aspiration/voicelessness (see e.g. Hombert et al. 1979), stronger degrees of aspiration should show larger F0 perturbation. On the other hand, if F0 perturbation is under speakers' control, then there should be F0 targets for each laryngeal category, and hence we do not predict such correlation. To address this question, Figure 6 plots the correlations between VOT and F0 of the following vowels, in which the regression lines are calculated within each laryngeal category. (Figures showing these correlations separately for each laryngeal category are provided as supplementary materials.)

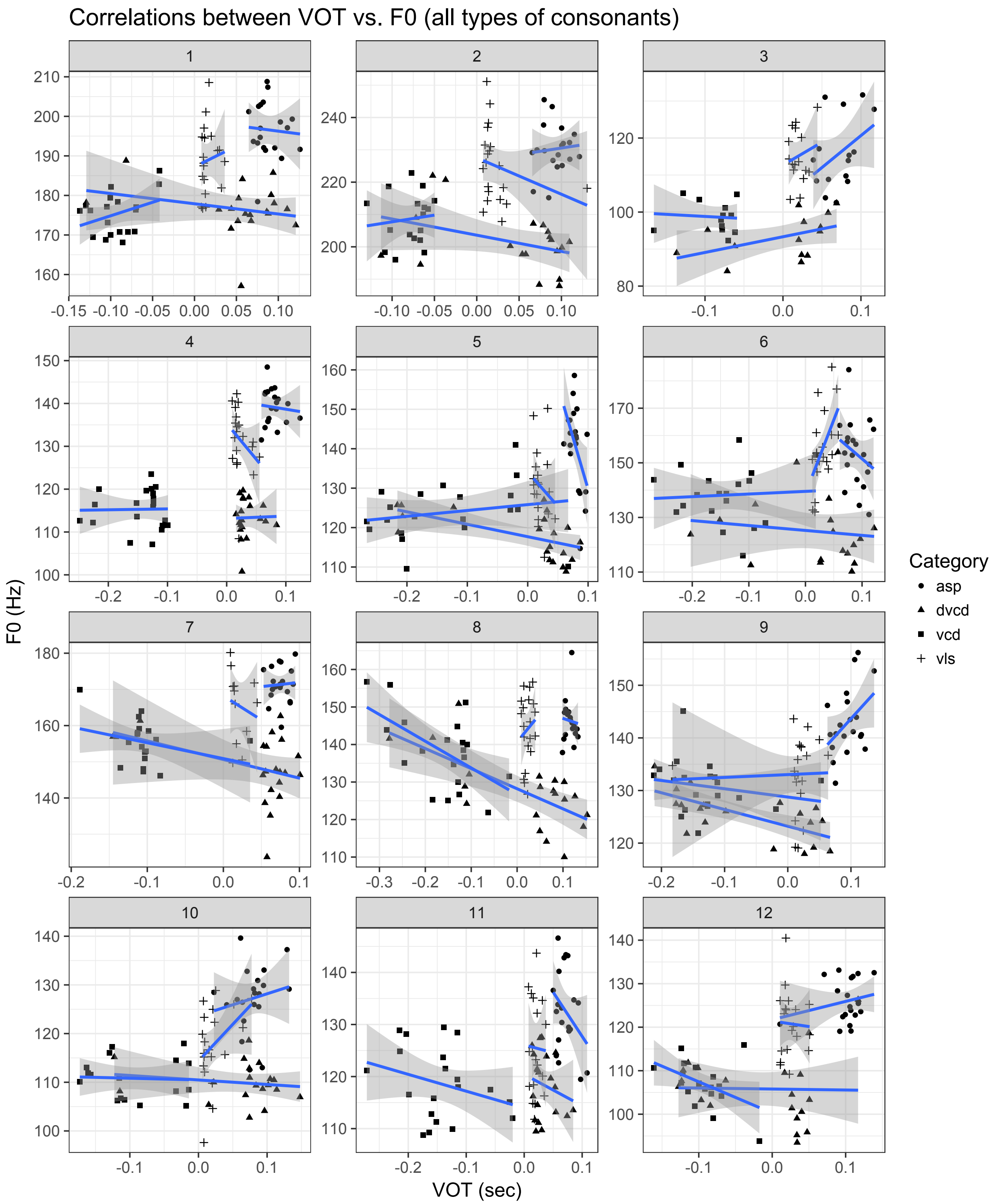


Figure 6: Correlation between F0 and VOT (all categories). Regression lines are calculated for each laryngeal category.

We observe that the correlations generally do not exist or are negative. There are only a few cases in which we observe clear positive correlations (e.g. voiceless consonants for Speakers 3 and 6 and aspirated consonants for Speakers 3 and 9). We conclude from this data that generally speaking, it is the phonological category, rather than raw phonetic values, that determine the F0 values of the following vowel (see in particular Clayards 2018; Dmitrieva et al. 2015; Kirby 2018 who reached a similar conclusion). It further implies that differences in F0 after different laryngeal categories are consequences of intended articulatory gestures rather than automatic consequence of laryngeal configurations (e.g. Kingston & Diehl 1994 vs. Hombert et al. 1979). In other words, each laryngeal category type looks to have a particular F0 target range.

The second question that our current study bears on is the relationship between phonological features and the way they are realized phonetically. One theory, Laryngeal Realism, argues that the relationship is very direct: the [+voice] feature should be realized as glottal vibration during closure, whereas [+spread glottis] should be realized as aspiration (Iverson & Salmons 1995, 2003; Jessen & Ringen 2002). For example, Jessen & Ringen (2002: 119-120) state that "[w]e take sounds specified as [spread glottis] to be implemented with an active glottal opening gesture...On the acoustic/auditory level, [spread glottis] in stop production is implemented primarily by aspiration...We take the feature [voice] in stop production to be prototypically implemented with voicing during closure."

Others point out, on the other hand, that the relationship between phonological features and their acoustic realizations is not always as direct. A series of work by Kingston and his colleagues (Kingston & Diehl 1994, 1995; Kingston et al. 2008) in particular argues that both F0 and F1 play a crucial role in the realization of the laryngeal contrast in English, and therefore, the laryngeal contrast in English—whether it is represented as a [voice] contrast or [spread glottis] contrast—is not solely realized in terms of differences in laryngeal gestures. They thus argue that the relationship between phonological features and their phonetic manifestations must be more abstract.[8]

The current data further corroborates the second view—in particular, "devoiced" consonants are characterized by low F0 and high F1, the latter of which is achieved via some supralaryngeal gesture (either by quick jaw opening or pharyngeal constriction/tongue retraction). In this sense, whatever the phonological feature that defines this class of sounds, its articulatory commands need to be "distributed" to some laryngeal gesture as well as to supralaryngeal gesture (cf. Fujimura 2000).

[8]Substance Free Phonology pushes this position to its extreme and posits that there are no intrinsic relationships between phonetics and phonology (Hale & Reiss 2000; Reiss 2017).

5 Conclusion

The current paper set out to explore how the four-way laryngeal contrast in Drenjongke is acoustically realized. This was important because (1) there has not been instrumental studies examining how this contrast is acoustically realized, and (2) a four-way laryngeal contrast is cross-linguistically rare. Our finding is summarized in Table 3.

Table 3: A summary of how the four-way laryngeal contrast is distinguished acoustically in Drenjongke.

	aspirated	voiceless	voiced	devoiced
VOT	long	short	negative	variable
F0	high	high	low	low
F1	high	low	low	high

We admit that this may not be the complete picture of the laryngeal contrast in Drenjongke. Since our analysis is based on CV-tokens, we were unable to measure other acoustic correlates that are known to signal laryngeal contrasts, such as preceding vowel duration and consonant duration (Chen 1970; Kingston & Diehl 1994; Lisker 1957; Port & Dalby 1982). More generally, it is known that the actual manifestations of a laryngeal contrast vary substantially across context (Kingston & Diehl 1994). Our future work will therefore examine the acoustic correlates of the four-way laryngeal contrast in different contexts, especially in the VCV context. With this limitation in mind, however, our data allows us to conclude that Drenjongke speakers use at least three acoustic dimensions—VOT, F0 and F1—to distinguish the four-way laryngeal contrast.

References

Abramson, Arthur S. & Douglas H Whalen. 2017. Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics* 63. 75–86.

Al-Tamimi, Jalal. 2017. Revisiting acoustic correlates of pharyngealization in Jordanian and Moroccan Arabic: Implications for formal representations. *Laboratory Phonology* 8(1). 1–40.

Boersma, Paul. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5(9/10). 341–345.

Chen, Matthew. 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22. 129–159.

Chodroff, Eleanor & Colin Wilson. 2018. Predictability of stop consonant phonetics across talkers: Between-category and within-category dependencies among cues for place and voice. *Linguistics Vanguard* 4(S2).

Clayards, Meghan. 2018. Individual talker and token variability in multiple cues to stop voicing. *Phonetica* 75(1). 1–28.

Dmitrieva, Olga, Fernando Llanos, Amanda A. Shultz & Alexander L. Francis. 2015. Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics* 49. 77–95.

van Driem, George. 2016. The phonology of Dränjoke: Experimental development of Roman Dränjoke and Phonological Dränjoke. Ms. University of Bern.

Erickson, Donna, Kiyoshi Honda & Shigeto Kawahara. 2017. Interaction of jaw displacement and f0 peak in syllables produced with contrastive emphasis. *Acoustical Science and Technology* 38(3). 137–146.

Fujimura, Osamu. 2000. The C/D model and prosodic control of articulatory behavior. *Phonetica* 57. 128–138.

Hale, Mark & Charles Reiss. 2000. "Substance abuse" and "dysfunctionalism": Current trends in phonology. *Linguistic Inquiry* 31. 157–169.

Hombert, Jean-Marie, John Ohala & William G. Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55. 37–58.

Hussain, Qandeel. 2018. A typological study of Voice Onset Time (VOT) in Indo-Iranian languages. *Journal of Phonetics* 71. 284–305.

Iverson, Gregory & Joseph Salmons. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12(3). 369–396.

Iverson, Gregory & Joseph Salmons. 2003. Laryngeal enhancement in early Germanic. *Phonology* 20(1). 43–74.

Jessen, Michael & Catherine Ringen. 2002. Laryngeal features in German. *Phonology* 19(2). 189–218.

Johnson, Keith. 2003. *Acoustic and auditory phonetics: 2nd edition*. Malden and Oxford: Blackwell.

Kawahara, Shigeto, Donna Erickson & Atsuo Suemitsu. 2017. The phonetics of jaw displacement in Japanese vowels. *Acoustical Science and Technology* 38(2). 99–107.

Kingston, John & Randy Diehl. 1994. Phonetic knowledge. *Language* 70. 419–454.

Kingston, John & Randy Diehl. 1995. Intermediate properties in the perception of distinctive feature values. In Bruce Connell & Amalia Arvaniti (eds.), *Papers in laboratory phonology IV: Phonology and phonetic evidence*, 7–27. Cambridge: Cambridge University Press.

Kingston, John, Aditi Lahiri & Randy L. Diehl. 2008. Voice. Ms. University of Massachusetts, Amherst.

Kirby, James. 2018. Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages. *Journal of Phonetics* 71. 326–354.

Kirby, James & D Robert Ladd. 2016. Effects of obstruent voicing on vowel F0: evidence from "true voicing" languages. *Journal of the Acoustical Society of America* 140(1). 2400–2411.

Lee, Seunghun & Shigeto Kawahara. 2018. The phonetic structure of Dzongkha: A preliminary study. *Journal of the Phonetic Society of Japan* 22(1). 13–20.

Lisker, Leigh. 1957. Closure duration and the intervocalic voiced-voiceless distinction in English. *Language* 33. 42–49.

Lisker, Leigh. 1975. Is it VOT or a first formant transition detector? *Journal of the Acoustical Society of America* 57. 1547–1551.

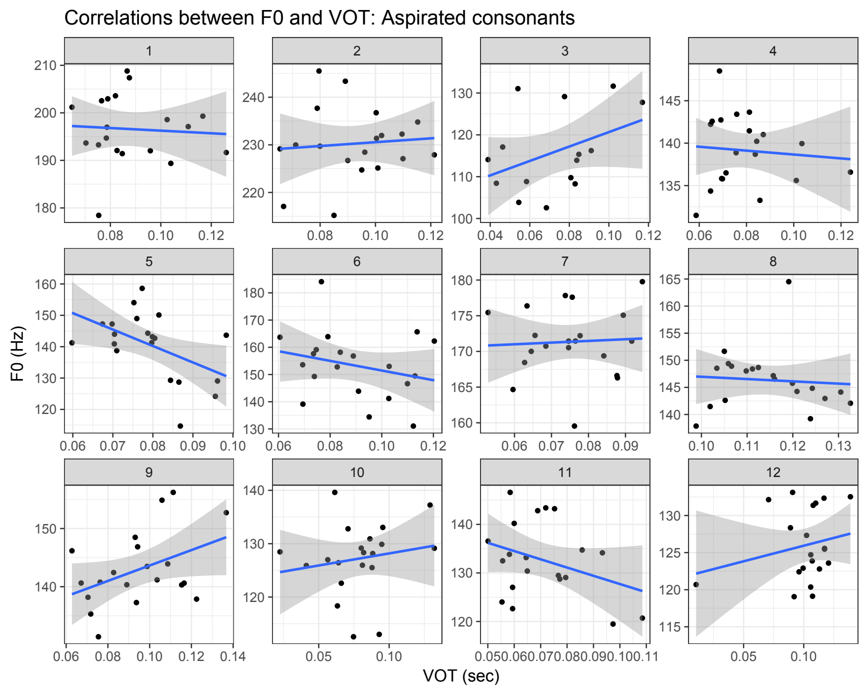
Lisker, Leigh. 1986. "Voicing" in English: A catalog of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29. 3–11.

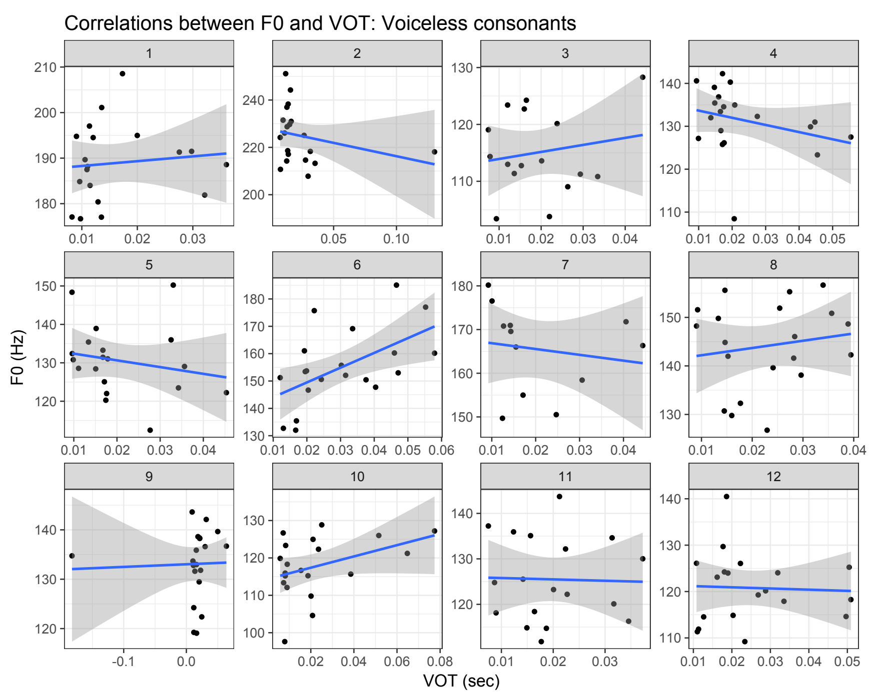
Port, Robert & Johnathan Dalby. 1982. Consonant/vowel ratio as a cue for voicing in English.

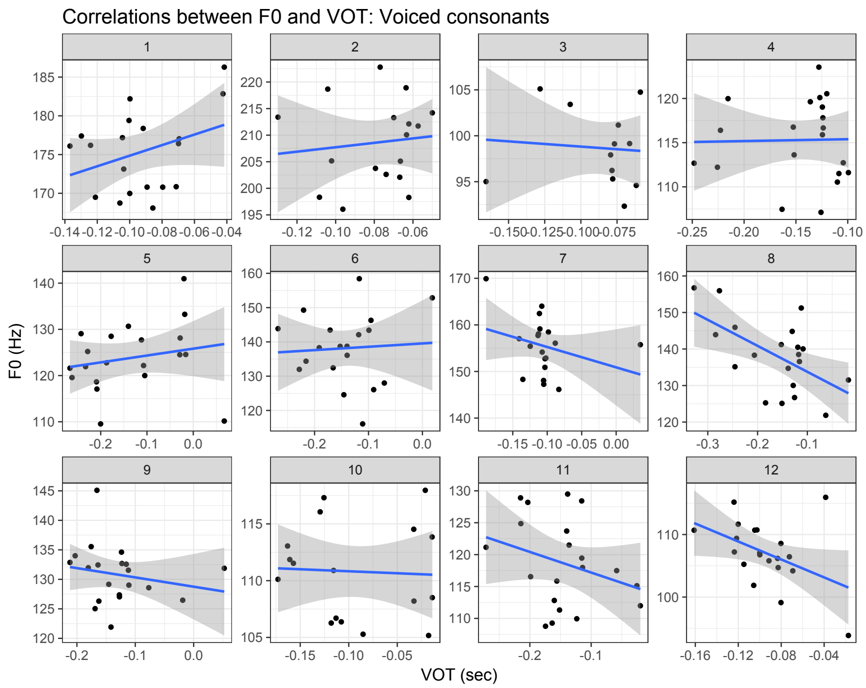
Perception & Psychophysics 32. 141–152.
Reiss, Charles. 2017. Substance free phonology. In S. J. Hannahs & Anna Bosch (eds.), *Routledge handbook of phonological theory*, Routledge.
van Summers, W. 1987. Effects of stress and final-consonant voicing on vowel production: articulatory and acoustic analyses. *Journal of the Acoustical Society of America* 82. 847–863.
Westbury, John R. & Patricia Keating. 1986. On the naturalness of stop consonant voicing. *Journal of Linguistics* 22. 145–166.
Whalen, Douglas H & Andrea G. Levitt. 1995. The universality of intrinsic F0 of vowels. *Journal of Phonetics* 23. 349–366.
Yliniemi, Juha. 2016. Attention marker =ɕo in Denjongke (Sikkimese Bhutia). *Linguistics of the Tibeto- Burman Area* 39(1). 105–160.

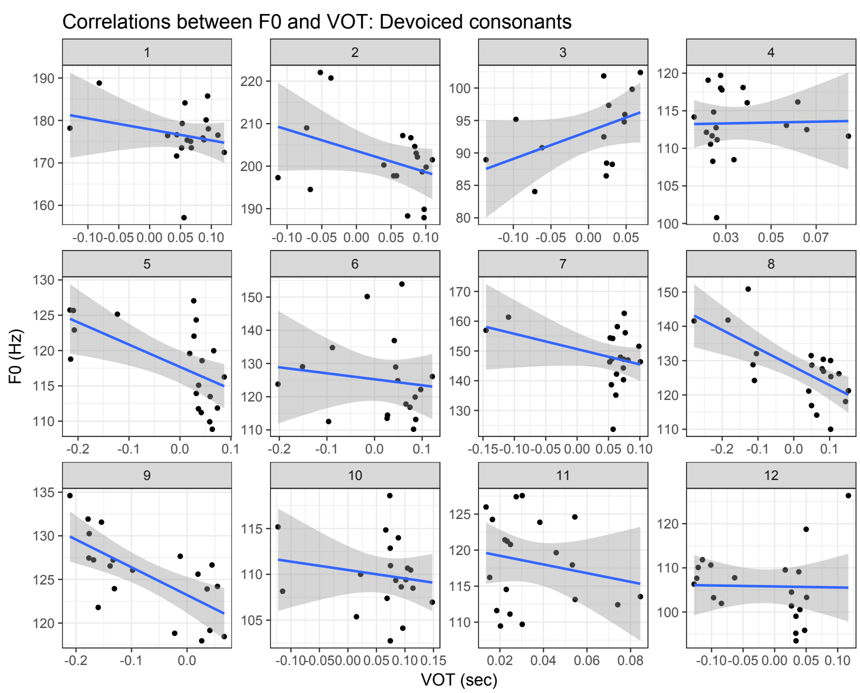
Supplementary figures: Correlations between F0 and VOT

[xxx Note to the editor: These are not to be included in the main text. Please use these as supplementary materials]






Correlations between F0 and VOT: Voiced consonants
1
2
3
4
5
6
7
8
9
10
11
12
F0 (Hz)
VOT (sec)


Correlations between F0 and VOT: Devoiced consonants
1
2
3
4
5
6
7
8
9
10
11
12
F0 (Hz)
VOT (sec)