Phonetic correlates of the Javanese voicing contrast in stop consonants

Michael J. KENSTOWICZ

Massachusetts Institute of Technology

This paper reports the results of a study of the phonetic correlates to the voicing contrast in the stop consonants of Javanese. The contrast is investigated at three different positions: word-initial, word-medial, and word-final. Previous research has found that the contrast is primarily reflected in the following vowel for the factors of pitch, vowel quality, and voice quality. This study largely replicates these findings. Our results suggest that the vowel quality factors may be swept up into the vowel harmony that characterizes the majority of disyllabic stems in Javanese.

1. Introduction

Contrasts in stop voicing are cross-linguistically common. The two principal phonetic correlates are the presence vs. absence of vocal fold vibration during the phase of oral closure and the presence vs. absence of a significant delay in the onset of voicing (VOT) in a following vowel or sonorant consonant. Paradigm examples of the first are found in Romance languages such as Spanish and of the second in some Germanic languages such as German. In English both closure voicing and aspiration (VOT) play a role in realizing the contrast depending on the segmental and prosodic contexts. In addition to these stop-internal cues, certain properties of an adjacent vowel can help to signal the consonantal voicing contrast. These include the duration of a preceding vowel (English) as well as F0 in the following vowel (English and many other languages). In addition, various spectral properties of the vowel such as breathy phonation are sometimes recruited to express the consonantal voicing contrast. A common path of diachronic development is for the vocalic feature to take over the burden of expressing the consonantal voicing contrast with a concomitant or subsequent minimization or complete loss of the stop-internal cues, as in tonogenesis (see Coetzee et al. 2018 for a recent example from Afrikaans).

Languages also differ in whether or not and if so how the voicing contrast in stop consonants is expressed as a function of context. Word-initial and intervocalic positions are favored sites for the maintenance of the contrast while word-final and internal to a consonant cluster are typical contexts where the voicing contrast is neutralized. The output of neutralization is typically in the direction of voiced stops after sonorants and to voiceless stops word-finally and in a cluster of obstruents. In the latter context the neutralized consonant frequently assimilates its voicing from the adjacent (and typically following) obstruent. According to Steriade (2009), the contextual typology forms an implicational hierarchy as a function of the number and quality of the cues available to signal the voicing contrast: word-medial > word-initial > word-final. This hierarchy has been extended to other laryngeal contrasts including aspiration and glottalization where the features of [+spread gl] and [+constricted gl] are either lost or shifted away from such neutralization sites, as in Icelandic preaspiration or Takelma preglottalization (Golston & Kehrein 2004).

The major languages of Java have added a new dimension to the typology of correlates to the stop-voicing contrast: the height of the following vowel as manifested in the first formant. This acoustic factor is argued to derive from a lowered larynx gesture by Cohn (1993) following a proposal of Trigo (1991) and earlier by Catford (1977). Our goal in this paper is to document and analyze the phonetic correlates to the stop voicing contrast

KENSTOWICZ, Michael J. 2021. 'Phonetic correlates of the Javanese voicing contrast in stop consonants'. NUSA 70: 1-37. Permanent URL: http://hdl.handle.net/10108/100088. doi: https://doi.org/10.15026/100088

in Javanese. Prior research has observed vowel height differences after word-initial or intervocalic Javanese voiced vs. voiceless stops. We add value to this line of study by documenting this factor in combination with other correlates for the voicing contrast for a larger set of Javanese data along with statistical tests of significance. We also examine and compare the expression of the voicing contrast for the three contexts of word-initial, word-medial intervocalic, and word-final positions.

The rest of this paper is organized as follows. In Section 2 we summarize the prior research on the voicing contrast in Bahasa Indonesia, Madurese, and Sundanese before looking into greater detail at the literature on Javanese. Sections 3 and 4 provide an overview of our study and its methods. The following sections then report our results for the expression of the Javanese voicing contrast in word-initial position for the low vowels (Section 5) and for the high and central vowels (Section 6). Section 7 presents the results for word-medial position and Section 8 does the same for word-final position. Section 9 is a brief summary discussion and conclusion.

2. Background

2.1 Indonesian languages: Bahasa Indonesia, Madurese and Sundanese

In her study of Bahasa Indonesia, Adisasmito-Smith (2004) reports that the stop-voicing contrast is realized with fully voiced vs. voiceless unaspirated stops in word-initial position. While the voiced stops were associated with lower F0 in the following vowel, no consistent spectral differences were found.

For Madurese Cohn (1993) and Cohn & Lockwood (1994) find a three-way [b] vs. [p] vs. [p^{h}] distinction. However, no minimal triples are possible because the voiceless unaspirated series is followed by lower vocalic allophones compared to the higher allophones found after voiced and aspirated consonants, as seen in the data of (1) from Misnadin & Kirby (2017). In these data the first column is a broad phonetic transcription while the second is orthographic. The Madurese voicing contrast is maintained in word-initial and medial (intervocalic) positions but is neutralized word-finally, where stops are typically unreleased.

(1)	ɛ ~ i	pɛɾak	pèrak	‘happy’
		p^{h}iʈak	bhiṭak	‘bird’
		bisa	bisa	‘able’
	a ~ ɤ	padɤ	padâ	‘same’
		p^{h}ɤʈɛ	bhâṭè	‘profit’
		bɤca	bâca	‘read’
	ɔ ~ u	pɔtɛ	potè	‘white’
		p^{h}uta	bhuta	‘giant’
		buta	buta	‘blind’
	ə ~ ɨ	pəsːɛ	pəssè	‘money’
		p^{h}ɨsːɛt	bhessèt	‘scratched’
		bɨsːɛ	bessè	‘iron’

Misnadin & Kirby (2017) review a couple of phonological reflexes of the Madurese vowel height correlate to the voicing contrast. First, when a stem-initial stop is lost via the pan-Austronesian nasal replacement process, the vowels shift to the nonhigh set that is appropriate after a nasal consonant.

(2)	N+[patɛ]	->	matè	[matɛ]	'AV.die'
	N+[pʰɤkta]	->	makta	[makta]	'AV.bring'
	N+[bɤca]	->	maca	[maca]	'AV.read'

Second, as shown by Cohn & Lockwood (1994), there is progressive height harmony across an intervocalic sonorant consonant and glottal stop so that the second vowel of a disyllabic stem assumes a higher vs. lower realization as a function of the voicing of the word-initial stop. This is evident from the data in (3) where the stem vowels in the first group are drawn from the higher set of allophones versus the lower set in the second.

(3)	bɤɾɤ	bârâ	'swell'
	bɤʔɤ	bâ'â	'flood'
	bulu	bulu	'feather'
	kʰɤru	ghâru	'scratch (by hand)'
	kʰulɤ	ghulâ	'sugar'
	t̪ʰɤʔɤr	ḍhâ'âr	'eat'
	lɛʔɛr	lè'èr	'neck'
	paʔaʔ	pa'a'	'chisel'
	pɛlak	pèlak	'kind'
	pɔla	pola	'probably'
	pɔɾak	porak	'cleave'
	raʔa	ra'a	'water germ'

Kulikov (2010) investigated a number of phonetic reflexes of the voicing contrast in Sundanese. His study is based on data obtained from two speakers who produced words in list format for six word-initial stops. The author reports the following reflexes for the voicing contrast. For VOT there was a robust contrast of −58 ms for the voiced series vs. 28 ms for the voiceless with velars showing the cross-linguistically familiar bias towards greater VOT. F0 measurements for the vowel [a] taken at vowel onset were significant for his female speaker with the expected lower values after the voiced series. With regard to spectral properties at the onset of the following [a], Kulikov's study found significantly smaller overall amplitude differences for H1−H2 after voiced stops (−0.6dB) than after voiceless stops (3.2dB) with larger differences in velars (6.2dB) compared to the bilabials (3.6 dB) or dentals (3.3 dB). Thus, for these Sundanese speakers the voiceless series was associated with more breathiness—at least in the low-vowel context. On the other hand, in the higher region of the spectrum measured by H1−A2, the voiced series was significantly associated with more breathy phonation: 16.2 dB (voiced) vs. 13.6 dB (voiceless). Finally, measurements of the first two formants at vowel midpoint found significantly lower F1 and higher F2 after the voiced stops compared to the voiceless ones for the nonhigh vowels, thus following the same general pattern as Madurese, at least with respect to F1.

2.2 Javanese

The voicing/laryngeal contrast in Javanese stops has received various designations in the scholarly literature including 'light' vs. 'heavy' (Horne 1974), 'clear' vs. 'breathy' (Catford 1977), 'tense' vs. 'lax' (Brunelle 2010), and 'stiff' vs. 'slack' (Hayward 1993, Ladefoged & Maddieson 1996). The latter two are articulatory in nature and refer to the degree of tension in the vocal folds produced by maneuvers of the arytenoid cartilages. The second is acoustic and refers to the distribution of energy in the speech spectrum. The first is more vague and proprioceptive in nature. In this section we review the studies available to us on the phonetic parameters of this phonological contrast, roughly in chronological order.

While not observing any voicing difference during the closure phase, Catford (1964, 1977:203) identified the following correlates to the contrast: [b, d, g] were articulated with a lowered larynx gesture that extended into the following vowel, which was produced with relaxed vocal folds vibrating with an open 'whisper-like chink' that resulted in volume-velocity of airflow four to six times greater than following [p, t, k] and lending them a breathy quality.

Using Horne's light vs. heavy designation, Fagan (1988) explored a number of possible stop-internal as well as external correlates of the Javanese laryngeal contrast with two male native speakers of the Yogyakarta (central Javanese) dialect. He focuses on the realization of the contrast in intervocalic position where laryngeal contrasts are typically the most robust cross-linguistically. Eight words displaying the contrast at the four places of articulation (labial, alveolar, retroflex, velar) were recorded in a frame sentence by each speaker. The preceding and following vowels were held constant as [a]. Measurements of common stop-internal reflexes of voicing contrasts as well as the duration of the surrounding vowels were taken along with the fundamental and formant frequencies of the following vowel. Lastly, voice quality was estimated by observation of possible energy reduction at the onset of the vowel following the stop. Fagan's findings can be summarized as follows. No significant differences were found for the canonical reflexes of a voicing contrast: VOT, closure voicing and duration, as well as the durations of the preceding and following vowels. The most reliable correlate of the contrast proved to be the value of the first formant in the following vowel, which was consistently lower for the heavy (voiced) stops at both the onset as well as the steady state region. Fagan interprets the F1 difference as a reflex of larynx lowering in the heavy stops on the grounds that this gesture lengthens the back cavity between the oral constriction and the vocal folds. The value of the second formant was significantly higher for the heavy stops at vowel onset but not at the steady state position. Fagan notes that this result regarding F2 is puzzling if larynx lowering is taken to be the primary articulatory correlate of the heavy-light stop contrast. Also, the heavy stops were associated with a significantly lower F0 value on the following vowel (measurement point not indicated) except for the velars. Finally, one of the speakers evidences aperiodic energy in the region above F3 at vowel onset as well as an overall reduction of energy for his heavy voiced stops suggesting possible breathy voice as another reflex of the heavy stops.

Thurgood (2004) builds on the results of Fagan's study with a focus on the voice-quality (spectral) reflexes of the Javanese laryngeal stop contrast, investigating whether the voiced pole of the opposition exhibits the common phonetic correlates of breathy voice that have been observed cross-linguistically. She first summarizes a pilot study by Hayward (1995) examining the realization of the vowels [i, a, ɔ, u] after word-initial [p] vs. [b] in 12 words (3 repetitions) produced by two Yogyakarta speakers. Hayward found significant differences in the amplitudes of the first and second harmonics for both speakers as well as a greater VOT for the slack-voiced stops for one of the speakers. Thurgood's own experiment investigated the realization of the contrast for word-initial labials [p] vs. [b] and velars [k] vs. [g] before the back vowels [a, ɔ, u] with 12 near-minimal pairs produced by a single speaker in two repetitions. FFTs were examined for the first two 46 millisecond regions in the vowel following the stop. The results did not evidence the typical profile of breathy voicing and instead showed an increase rather than a decrease in the amplitude of the various peaks in the lower region of the spectrum. On the other hand, Thurgood's subject did evidence significantly lower values following the slack-voiced stops for F1 in the nonhigh vowels [a] and [ɔ] but not for the high vowel [u], which had the opposite profile. In addition, Thurgood reports that the back vowels were fronted in the context following slack-voiced stops as reflected in significantly greater F2 values. Finally, the low vowel became more central (smaller F1 and greater F2) in the second window representing the steady state region of the vowel.

Brunelle (2010) investigates the articulatory correlate(s) of the Javanese contrast (termed tense vs. lax in his study). Prior research has postulated glottal opening for the lax stops to explain the breathiness, lower F0, and increased VOT found in some of the earlier studies. But Brunelle observes that this articulatory gesture fails to explain the most consistent finding of decreased F1. As noted above, larynx lowering is the most plausible mechanism to explain this acoustic effect. Brunelle's study explores this hypothesis more directly with the analysis of a fiber optic recording of two Javanese speakers made by Katrina Hayward in 1995. The recording consists of six repetitions of 20 near-minimal pairs of words like *iki pitik* 'this is a chicken' vs. *iki bibit* 'this is a seed' that target the binary tense-lax contrast for stops and affricates. The following vowel was held constant as [i] in Hayward's recording. The video does not permit the larynx to be viewed directly and so the width of the epiglottis (measured in pixels) is taken as an indirect reflection of larynx height. The low time resolution of the equipment used (25 frames per second) does not permit study of the time course of the larynx lowering and so separate measures were made for the stop itself as well as the following vowel. Brunelle's main finding is a significantly greater epiglottal width for the tense stops across most points of articulation for both speakers. This result held for the regions of stop closure as well as the following vowel. In his discussion of this finding, Brunelle notes a Javanese parallel to the phenomenon of "register" in several Southeast Asian languages where diachronically earlier voiced stops have evolved so that the stop contrast is reflected in the following vowel in terms of pitch, vowel quality, and spectral energy distribution. See Brunelle & Kirby (2016) for further typological discussion.

Matthews (2017) investigates the Javanese stiff vs. slack voicing correlates in the following vowel for the factors of voice quality, quantity, and pitch. In his study data was collected from a female speaker of the central dialect who produced six tokens for each of the four word-initial onsets of b(l) vs. p(l) before the point vowels [a], [i], [u]. Each word was recorded in a frame sentence with two repetitions. Complex onsets [bl] and [pl] were included in the study to see whether the contrast was realized on the sonorant lateral as well as carrying over into the following vowel. Measures were taken across the first 25 milliseconds of the vowel and as well as the lateral consonant. Matthews reports that the vowels following the slack-voiced onsets showed the familiar profile of lower values for F0 and F1 and higher values for F2. These differences were also found for the complex onsets as well as in the lateral itself. The stiff vs. slack voicing contrast was also evident in larger differences for the spectral measures of H1−H2 and H−A2 in the slack context for the high vowels showing the greater spectral drop-off that is a characteristic of breathy voicing. However, the low vowel showed a more breathy profile (larger H1−H2) in the stiff voiced context and overlapping distributions for H1–A2 for the stiff vs. slack voiced stops. Matthew's results suggest that the Javanese stiff vs. slack stop voicing contrast can be realized in the entire voiced sonorant region following the onset stop. He speculates that the difference between the high vs. low vowels with regard to the spectral measures might be explained by saying that the tongue body raising and advancement associated with high vowels is articulatorily incompatible with larynx lowering and so instead relies on relaxing glottal constriction resulting in more breathy phonation.

Finally, Perwitasari et al. (2017) investigated the effect of the stiff vs. slack contrast on all six of the Javanese vowel phonemes (as well as the seven vowels of Sundanese) for the parameters vowel quality as reflected in the first and second formants. Four speakers (two male and two female) from each language were recorded pronouncing three repetitions of target words in a frame sentence; the onset consonants were held constant as [b] for the slack voiced context and [h] for the baseline (voiceless). Formant measures were taken at the steady-state midpoint of the following vowel. The Javanese results indicated a significantly lower F1 value for all vowels except schwa; no consistent differences were found for F2. A similar pattern for F1 and F2 was found for Sundanese as well.

The following table summarizes the various factors investigated in the previous literature on the phonetic correlates of the Javanese stop-voicing contrast. N refers to the number of items.

Table 1. Synopsis of prior studies on Javanese

study	subjects	context	stops	vowels	N	Parameters
Fagan 1988	2 (M)	inter-vocalic	p, t, th, k; b, d, dh, g	a	16	F0, F1, F2 and others
Thurgood 2004	1 (M)	initial	p, k, b, g	a, ɔ, u	70	F0, F1, F2; H1–H2, H1–A1, H1–A2
Brunelle 2010	2 (1F)	initial	p, t, th, c, k; b, d, dh, j, g	i	12	epiglottal width
Matthews 2017	1 (F)	initial	p(l) vs. b(l)	a, i, u	72	F0,F1,F2,H1–H2, H1–A2
Perwitasari 2017	4 (2F)	initial	b vs. h	i, u, e, ə, o, a	72	F1,F2

The most reliable correlates of the contrast are lower F0 and F1 in the following vowel for the Javanese voiced stops. More variable is the effect on F2 with some studies reporting an increase for back vowels. Spectral reflexes are also variable with some drop-off reported by Fagan but not by Thurgood and a a difference as a function of vowel height by Matthews. In terms of articulation, larynx lowering best explains the F1 differences and was indirectly established by Brunelle. F0 and breathy phonation implicate the laryngeal setting for open quotient—a relaxation of vocal fold tension for the voiced stops. Both of these gestures can help to sustain vibration of the vocal folds in the face of oral closure and hence are best explained as fossilized remnants of an earlier stop voicing/VOT contrast that is still found in Bahasa Indonesia and other Austronesian languages. The F2 differences are more variable; it is unclear whether and if so how they can be connected with either glottal opening or glottal lowering and appear to be an additional and independent reflex of the contrast.

3. Overview of this study

The goal of our study is to broaden the range of the phonetic correlates of the Javanese stop contrast to be investigated as well as its phonological status. Specifically, while earlier studies either examined a limited number of consonants in a wider range of vowels or vice versa (see Table 1), we broaden the data set by fully crossing the vowels and stop consonants as well as comparing them to a more neutral baseline sonorant when appropriate. Second, whereas earlier studies restricted the context primarily to word-initial position, we examine its implementation in all three contexts where stops are phonotactically permitted in Javanese: word-initial, word-medial, and word-final. Third, taking a cue from Madurese, we investigate the domain of the realization of the Javanese voicing contrast by asking whether the word-initial opposition is reflected in the second syllable of a CVRVC stem. Fourth, while almost all prior studies have focused on the correlates to the voicing contrast that are found in the following vowel, we also examine the status of the vowel preceding the stop in intervocalic and word-final positions. Finally, consistent tests of statistical significance are reported for all phonetic correlates investigated. Like the previous studies of Javanese reviewed above, our data are based on the speech of a small number of speakers (in our case just one). However, due to the fact that in large part our results replicate earlier findings, we can place some confidence in the data obtained from the broader range of contexts and parameters studied here.

4. Methods

Our investigation of the phonetic correlates to the voicing contrast in Javanese was conducted as follows. Our speaker is an educated female in her forties from the central dialect area. She uses Javanese on a daily basis with her family and Javanese friends. The recordings were made in a sound-insulated booth with a head-mounted Shure SM10A Unidirectional Head-Worn Dynamic Microphone and a USB Pre 2 Preamp at a sampling rate of 44.1 kHz, 16 bits. The data of interest were recorded in randomized word lists consisting of the English gloss followed by the Javanese lexical item. The words were taken primarily from Horne (1974) but were checked for familiarity to our speaker. Five repetitions were made for each list, with the speaker going through the entire list once and then four more times. The sound files were analyzed with Praat textgrids (Boersma & Weenink 1992–2017) with segmentation based on visual inspection of the spectrogram, waveform, and its auditory properties. The measurements of interest were gathered by Praat scripts for all phonetic correlates except the phonation factors of H1−H2, H1−A1, and H1−A2. The latter were collected by a Matlab script in VoiceSauce (Shue et al. 2011). Charts and statistical tests were made in R version 2.11.1 (Bates & Maechler 2010, R Development Core Team 2011) or in simple cases in Excel. Mixed-effects linear regression tests were run with word (item) and trial (repetition) as random intercepts; random slopes by word were set for the fixed effects. In cases where the tests did not converge, the random slope was dropped. A t-value greater than 2.0 was taken to be significant. All statistical models are included in the appendixes along with the test words.

5. Word-initial position: low vowels

The correlates of the Javanese stop voicing contrast for word-initial position were investigated with two sets of data. The first examined the effect of the contrast on the low vowel [a] and its rounded counterpart [ɔ]. The later is derived from underlying /a/ by two regular morphophonemic processes (Dudas 1976). The first changes /a/ to /ɔ/ in word-final position and the second spreads the height and rounding of this derived /ɔ/ to a preceding /a/ in an open syllable: cf. [med³ɔ] 'table', [med³a-ne] def. and [bɔsɔ] 'language', [basa-ne] def. There are 31 words of the CaCaC shape and 25 CaCa (= [CɔCɔ]) in the list. The medial consonant is held constant as a sonorant to provide a more neutral baseline compared to a voiced or voiceless stop and the initial consonant is a stop that varies its place of articulation (labial, coronal, velar) as well as voicing (voiced, voiceless). The CaCaC and CaCa lists were merged and then randomized. See Appendix A for the complete list. Several measurements were made for this data set: VOT, the first and second formants as well as the F0 of the following vowel, and the spectral factors of H1−H2, H1−A1, and H1−A2.

For VOT measures, the results as a function of onset voicing and place of articulation are indicated in the boxplots below. They exhibit the cross-linguistically familiar velar > coronal > labial VOT hierarchy for place. But the voiced stops have systematically greater VOT than the corresponding voiceless ones. This is contrary to cross-linguistic expectation and reflects the breathy slack voicing of the voiced stops in Javanese noted by Ladefoged & Maddieson (1996) and others, as mentioned above.

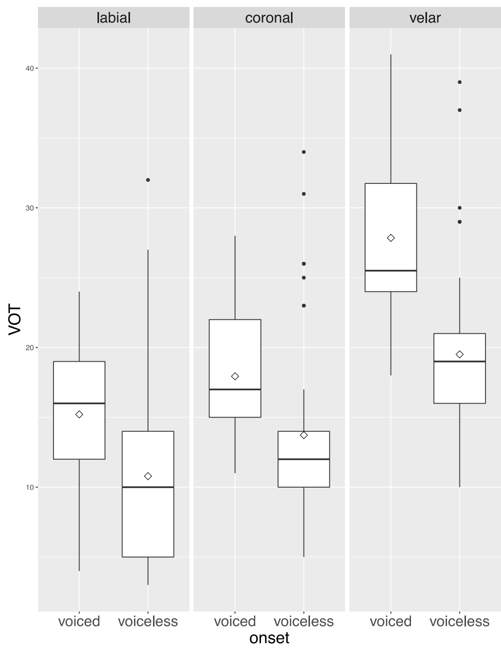


Figure 1. VOT (ms) word-initial stops

Mixed-effects linear regression with voicing and place as fixed effects found both factors to be significant (Table 2). The interaction of voice and place was also checked but proved to not be significant and was dropped from the model. In this test, the data were treatment coded with coronal and voice as the baselines for the fixed effects.

Table 2. Mixed-effects linear regression model for VOT

	Estimate	Std. Error	t value
(Intercept)	18.906	1.177	16.065
Onset:voiceless	–5.070	1.104	–4.593
Place:labial	–3.083	1.393	–2.212
Place:velar	7.310	1.376	5.314

Measurements of the formants were taken at the midpoint of the initial vowel. The results are reported in the boxplots and relevant regression tests below. For both [a] and [ɔ] F1 was significantly lower after the voiced stops compared to the voiceless stops; but for F2 no significant difference was found for either vowel.

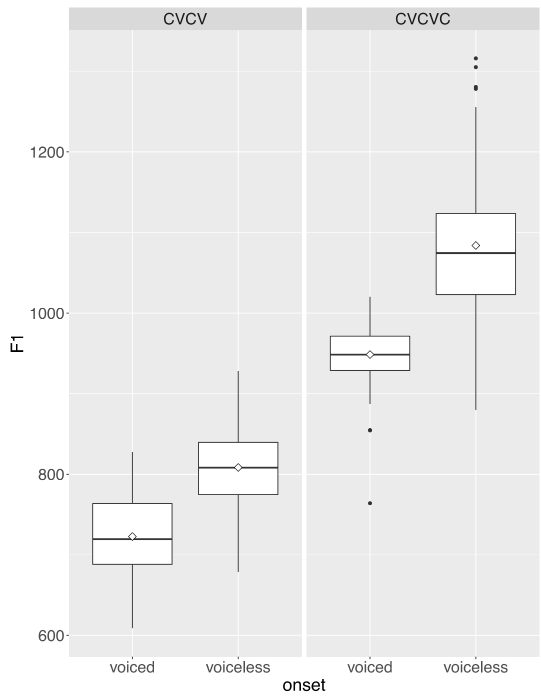


Figure 2. F1(Hz) V1; CɔCɔ and CaCaC

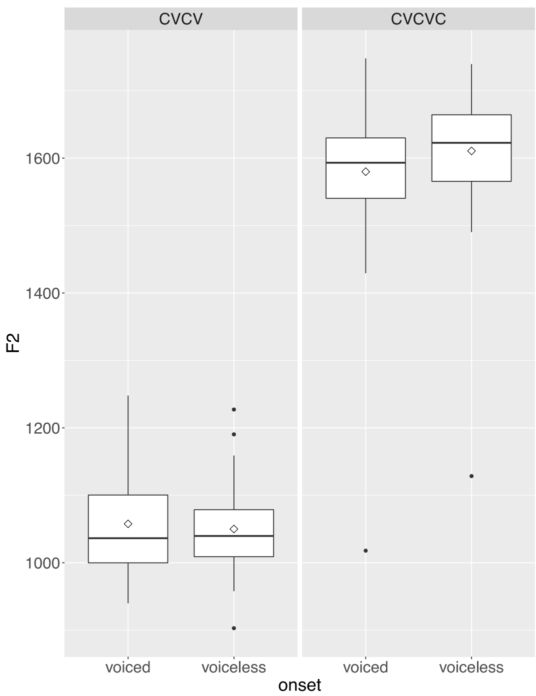
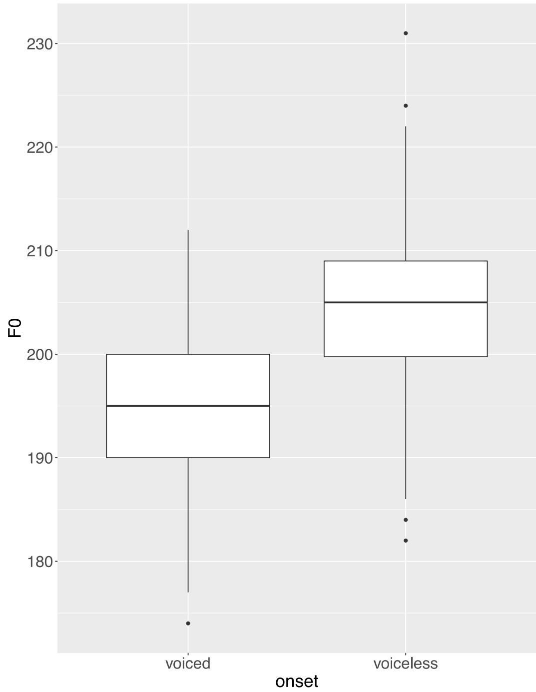


Figure 3. F2(Hz) V1; CɔCɔ and CaCaC

We also calculated the F0 and duration values for the vowels following the word-initial stops. Both measures were taken across the entire vowel and are seen in the boxplots in Figures 4 and 5 below.



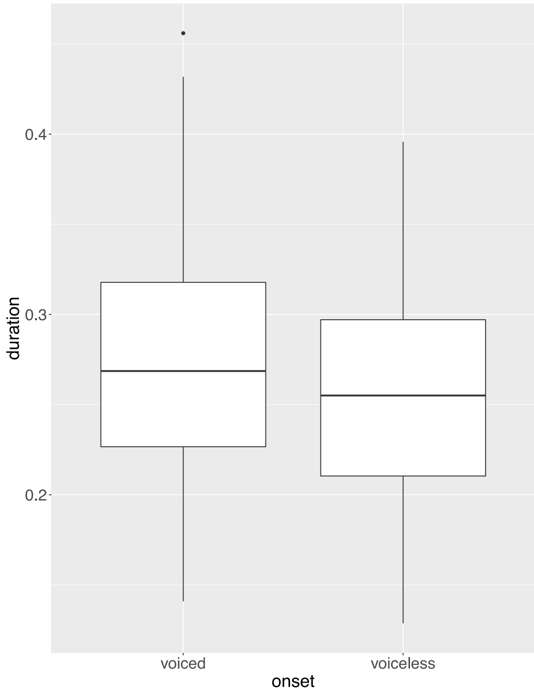


Figure 4. F0 (Hz) V1 **Figure 5. Duration (ms) V1**

In the Table 3 below B and P stand for any voiced vs. voiceless stop and R stands for the medial sonorant consonant. For the regression tests, the factor of place and its interaction with voice proved to be non-significant and was dropped from the final models reported here. The numbers in square brackets for the regressions designate the models to be found in the appendixes.

Table 3 also indicates a significant difference in F0 with the voiced stops showing a lower value—a result consistent with all of the earlier studies on this correlate. There was a trend in the direction of increased duration after the voiced stops but it fell short of reaching the 2.0 level of significance.

Table 3. Mean (st dev) in Hz for V1 in CaRaC and CaRa (=[CɔCɔ]) stems

factor	BaRaC		PaRaC		regression (t)		BaRa		PaRa		regression (t)	
F1(Hz)	948	(39)	1083	(86)	9.94	[1]	722	(79)	808	(57)	4.95	[2]
F2(Hz)	1579	(85)	1610	(82)	1.61	[3]	1057	(77)	1050	(58)	–0.29	[4]
F0(Hz)	194	(7.4)	205	(9.3)	9.89		195	(6.9)	203	(6.6)	6.04	
duration(ms)	275	(64)	247	(60)			270	(55)	265	(55)		

merged	voiced		voiceless			
F0(Hz)	194	(7.1)	204	(8.2)	8.63	[5]
duration(ms)	273	(60)	255	(58)	–1.44	[6]

Figures 6 and 7 below show the averaged F1 values for the stem-initial [a] and [ɔ] vowels as a function of the voicing of the onset consonant obtained from time-normalized measurements taken across the middle 80% window of the vowel using Formant-Pro (Xu 2020). The time course for the voiceless stops is relatively flat while the voiced stops start at a lower point and climb towards the middle of the vowel. This trajectory arguably tracks the lowered larynx gesture.

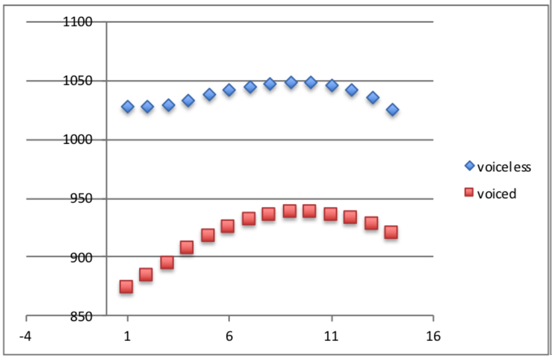


Figure 6. F1(Hz) in V₁[a] (normalized)

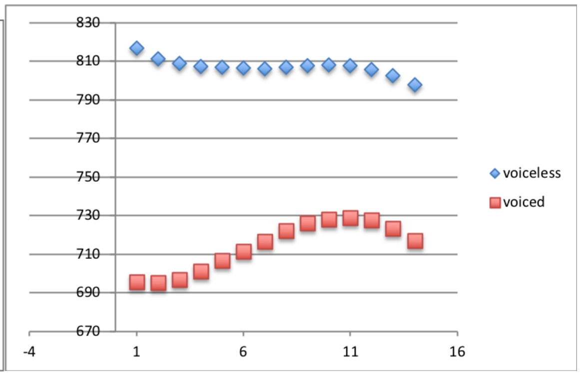


Figure 7. F1(Hz) in V₁[ɔ] (normalized)

We also investigated the spectral tilt of the vowel following the onset consonant as reflected by H1–H2. This factor measures the difference between the amplitudes of the first and second harmonics. A larger (positive) value indicates a sharper drop-off of the energy in the spectrum and is customarily taken as a reflex of breathy voice. For measurement of this factor we utilized VoiceSauce (Shue et al. 2011). 25 ms. intervals at the beginning of the vowel as indicated by the autocorrelation setting were marked in the Praat textgrids for each initial-syllable vowel. The values for the vowel across this window returned by VoiceSauce were then averaged. Figure 8 below summarizes the measurements obtained. Error bars are standard errors. Here as well, our results replicate in part earlier findings. The vowels following the voiced stops show a greater spectral tilt compared to the voiceless ones and point to breathy phonation similar to Hayward's (1993) findings and in contrast to the findings of Kulikov (2010) for Sundanese and in part by Thurgood (2004) for Javanese. However, as shown by the large variances, the data are quite noisy. Part of this can be attributed to place of articulation of the consonant. As shown in Figure 8, the H1–H2 values are much greater following the voiced velar stops. We recall from Figure 1 that this was also the pattern with VOT and appeared in Kulikov's study of Sundanese mentioned above as well. This suggests that the VOT and phonation factors are related.

Table 4. Mean (st dev) in dB for H1–H2 in V_1 of CaCaC and CɔCɔ stems (merged)

voiced	2.35 (3.82)
voiceless	–1.04 (3.63)

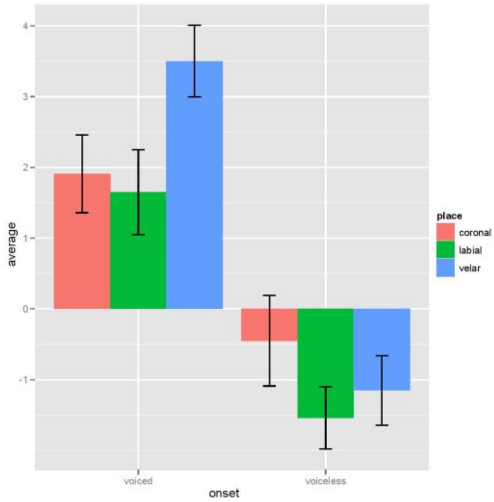


Figure 8. H1–H2 (dB) in V_1 of CaCaC and CɔCɔ stems (merged)

Mixed-effects linear regression finds the voice factor to be significant while the effect of velar place and its interaction with voicing falls at the margins of significance (Table 5). In this model, treatment coding was employed with voiced and coronal as the baselines. An overall change to voiceless significantly depressed the H1–H2 factor while change to velar did so marginally; but when combined a greater effect was achieved relative to the baseline. There was no significant difference between baseline coronal and labial.

Table 5. Mixed-effects linear regression model: H1–H2

	Estimate	Std. Error	t value
(Intercept)	1.8998	0.6431	2.954
voice:voiceless	–2.2708	0.9216	–2.464
place:labial	–0.1921	0.8500	–0.226
place:velar	1.5883	0.8109	1.959
voice:voiceless-place:labial	–1.0409	1.3481	–0.772
voice:voiceless-place:velar	1.8998	0.6431	2.954

Testing for the spectral measures of H1–A1 and H1–A2 found no significant differences for voicing; but there was a nearly significant positive effect for labial place vis a vis the coronal baseline for H1–A1 (beta = 2.1, t= 1.9) and a marginally significant effect for H1–A2 (beta = 2.5, t=2.2).

We report one final measurement made with the low vowel data set A—the F1 value for the second stem vowel V_2. Our interest in this factor is prompted by two considerations. First, there is a strong tendency for the stem vowels of Javanese to be identical. Uhlenbeck’s (1950) study of Javanese stem structure found that 85% of roots are disyllabic with CVCVC the most common shape. He states that stems with identical vowels (i.e., CV_iCV_iC) are over-represented statistically. Furthermore, when one of the stem vowels is altered by a phonological process, the change may be passed on to the preceding vowel, as in the case of the rounding of word-final /a/ mentioned above. Second, we recall that Misnadin & Kirby's (2017) summary of Cohn's (1993) study of Madurese reports such an effect (3) as well. We were curious whether the F1 difference in V_1 as a function of the voicing of the onset consonant would show up in the second vowel of our CaCaC and CaCa (= [CɔCɔ]) stems. In fact, such a difference was observed in the data, as summarized in Table 6 and Figure 9 below. In the chart error bars are standard errors. Regression tests found this difference to be strongly significant. It suggests that the stem harmony extends to this more secondary reflex of vowel quality.

Table 6. V_2 mean (st dev) in Hz

factor	BaRaC	PaRaC	regression (t)	BaRa	PaRa	regression (t)
F1	927 (78)	1054 (134)	6.75 [8]	758 (102)	841 (85)	4.91 [9]

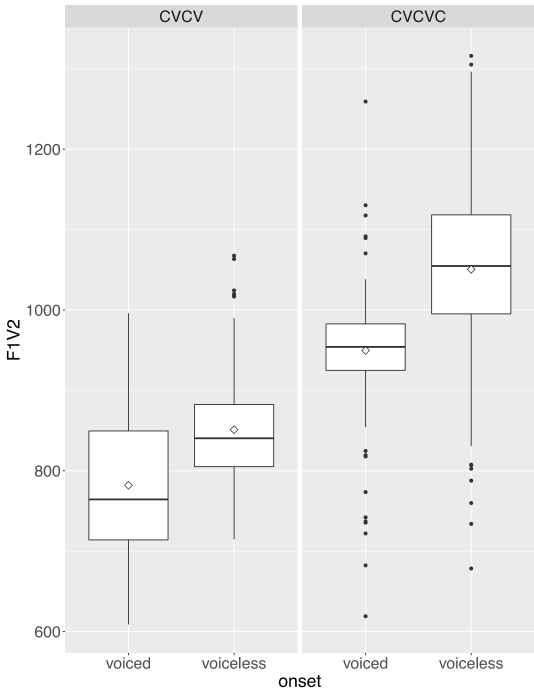


Figure 9. V_2 F1 (Hz); CɔCɔ and CaCaC stems

6. Word-initial position: high vowels and schwa

We analyzed a smaller set of data to see how the Javanese word-initial voicing contrast was reflected in a following high [i, u] and mid central vowel [ə]. This set consisted of 24 words of the structure C_1VC_2əC. C_1varied among a voiced, voiceless, and sonorant consonant for the three places of articulation (labial, coronal, velar). C_2 was held constant as a sonorant (liquid, nasal, glide) and V_2 was restricted to schwa. The words were recorded in randomized list format with five repetitions. Vowel formant measures were taken at the midpoint of the vowel by Praat scripts. For the regression tests, the data were treatment coded with sonorant and schwa as the baselines. See Appendix Set-B for the complete word list. As seen in Tables 7 and 8 and the boxplots (Figure 10) below, F1 was significantly lower after the voiced stops for each of the three vowels. For F2 (Figure 11), the voiced stops were associated with a greater mean value compared to voiceless and sonorant in the central and back vowels (as in Sundanese and for [u] in Thurgood's study of Javanese) but the difference was not significant.

Table 7. F1 mean (st dev) in Hz for V_1 midpoint

onset	i	u	ə	regression (t)
sonorant	449 (16)	495 (73)	766 (32)	
voiced	404 (18)	447 (13)	664 (33)	–3.79 [10]
voiceless	437 (24)	482 (18)	776 (18)	0.24

Table 8. F2 mean (st dev) in Hz for V1 midpoint

onset	i	u	ə	regression (t)
sonorant	2722 (95)	1027 (183)	1505 (155)	
voiced	2732 (84)	1103 (139)	1534 (103)	0.47 [11]
voiceless	2771 (87)	1024 (164)	1420 (109)	–0.43

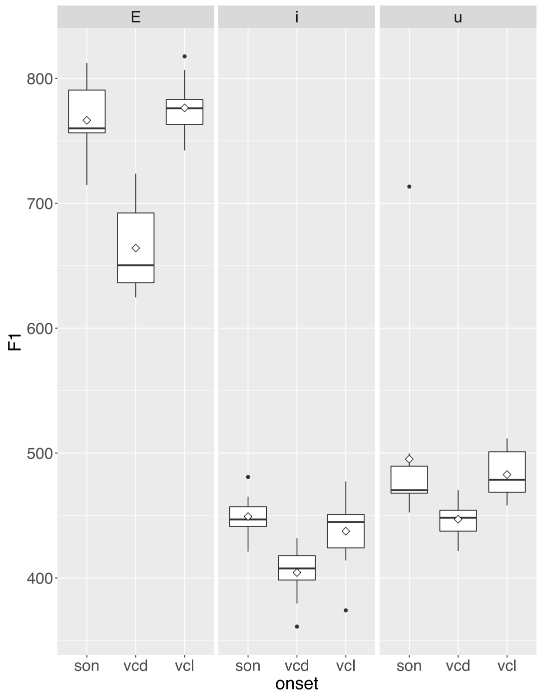


Figure 10. F1(Hz) for V_1 ; E = schwa

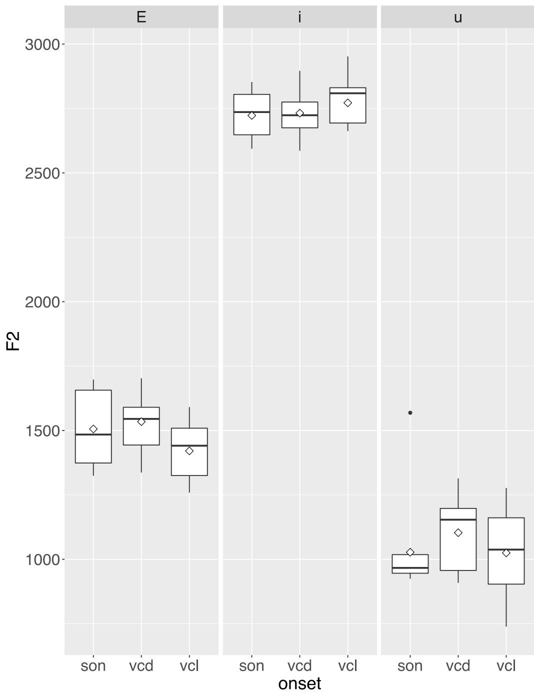


Figure 11. F2 (Hz) for V_1; E = schwa

The voiced stops also differed significantly in the expected direction with regard to F0 in the following vowel as shown in Table 9 and Figure 12 below. This measurement was taken across the entire vowel. For the regression tests the baseline was sonorant; voiced stops differed significantly from sonorants while voiceless stops did not do so.

Table 9. F0 mean (st dev) in Hz for V_1

onset	F0 (Hz)	regression (t)
sonorant	232 (11)	
voiced	211 (12)	–4.97 [12]
voiceless	232 (12)	–0.47

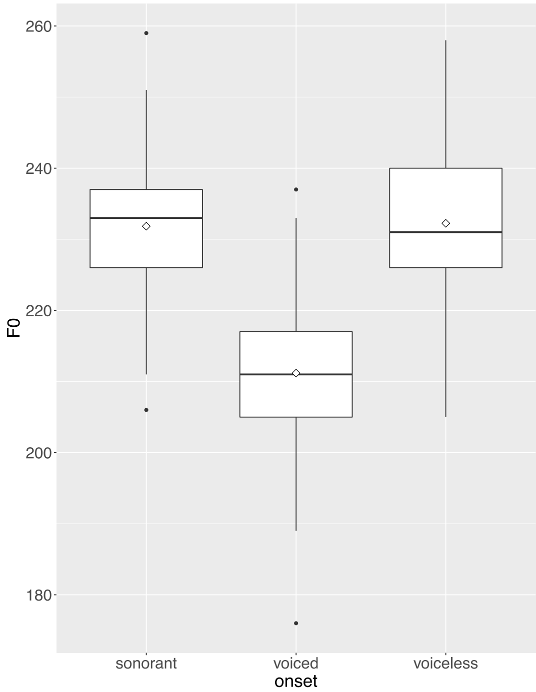


Figure 12. F0 (Hz) by onset for V_1

The high vowel and schwa data set are of the form CV_1RəC where the medial consonant is a sonorant and the second syllable vowel is schwa. We tested whether the second syllable vowel would differ for F1 as a function of the voicing of the word-initial consonant. No significant difference was found even when the test was restricted to stems whose first vowel was also a schwa (t = –0.36).

In sum, the Javanese voicing contrast for stops in word-initial position is reliably reflected in several phonetic factors: VOT, F1 and F0 of the following vowel and somewhat less reliably as a breathy vs. modal voice difference in the earliest vocalic region following the stop. No significant difference was found for F2. In addition, the phonological proclivity for identical stem vowels in Javanese compelled the F1 difference signaling the voicing contrast in the initial stop to penetrate the medial sonorant consonant and show up in the second vowel of the CaCaC and CaCa (= [CɔCɔ]) stems. But this effect was not found for the $CV_1RəC$ stems where V_1 was a high vowel or schwa.

7. Word-medial position

In order to investigate the phonetic correlates of the Javanese stop-voicing contrast in word-medial intervocalic position another set of words was constructed. This set (see Appendix set C) consisted of 27 words of the shape RaCaC. The initial consonant was restricted to a sonorant to provide a more neutral baseline compared to a voiced or voiceless stop while the medial stop varied for voicing (voiced vs. voiceless) and place (labial, coronal, velar). The words were recorded and analyzed using the same methods as for the word-initial stops. Once again, B and P are labels in our tables for the voiced and voiceless, respectively, stops; voiced was set as the baseline for the regression tests. With respect to the stop consonant, no statistically significant effects were found for the properties of closure duration or for VOT as a function of the voicing of the medial consonant (Table 10).

Table 10. Mean (st dev) in ms for medial stops in RaCaC stems

factor	RaBaC	RaPaC	regression (t)
closure duration	151 (45)	142 (27)	–1.16
VOT	17 (9)	16 (9)	–0.33

Table 11 shows the results for the various measurements of the vowel following the medial stop. Comparable to word-initial position, they indicate that the stop voicing contrast is reliably correlated with differences in the first formant and the fundamental frequency of V_2. Voiced stops were associated with significantly lower F1 and F0; they show only weak trends in the direction of greater duration and greater F2. The corresponding boxplots are also provided in Figures 13, 14, and 15.

Table 11. V_2 mean (st dev) for RaCaC stems

factor	RaBaC		RaPaC		regression (t)	
duration (ms)	120	(30)	107	(30)	–1.13	
F1 (Hz)	886	(43)	995	(45)	9.29	[12]
F2 (Hz)	1685	(106)	1649	(117)	–1.34	[13]
F0 (Hz)	226	(23)	254	(15)	5.84	[14]

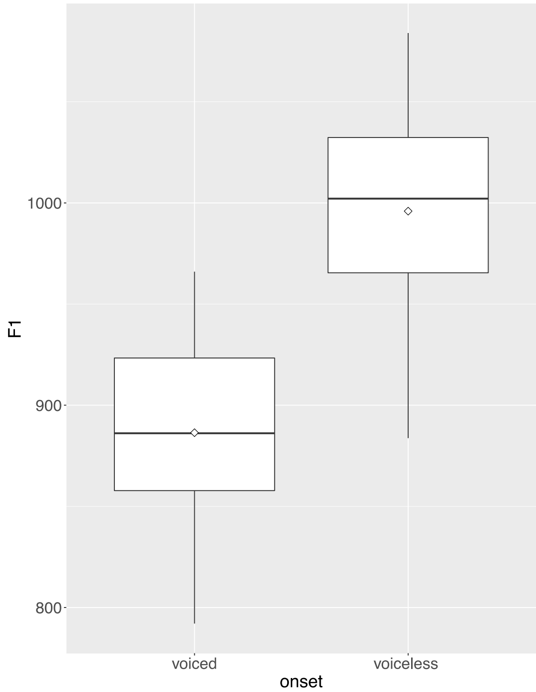


Figure 13. Medial V_2 F1 (Hz)

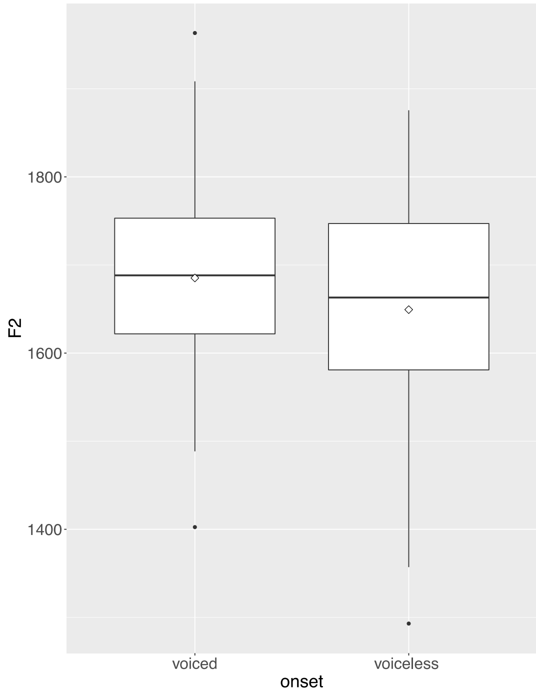


Figure 14. Medial V_2 F2 (Hz)

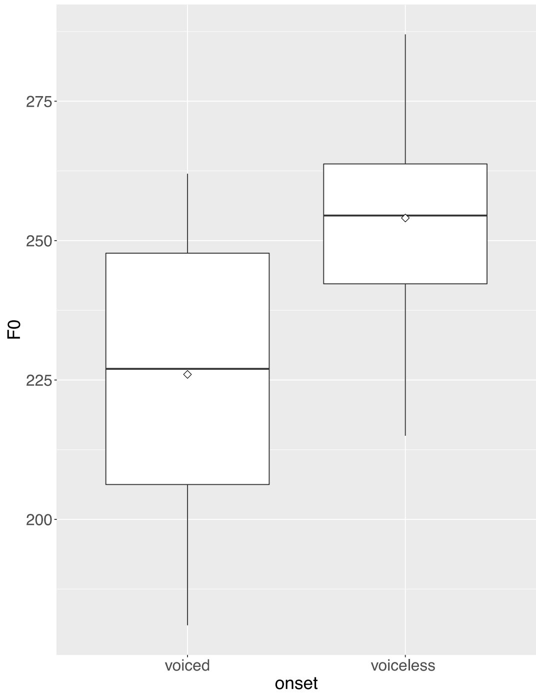


Figure 15. Medial V_2 F0 (Hz)

What about the vowel preceding the stop? The results are shown in Table 12 and the accompanying plots in Figures 16, 17, and 18. They indicate that the preceding vowel is reliably longer before the voiced stop, a difference analogous to what is found in English and many other languages. Interestingly, the closure duration of the stop itself did not reliably vary by voicing, suggesting that the durations of V_1 and the following consonant are disassociated and that V_1 duration is an independent cue for the voicing contrast in

Javanese. Another noteworthy finding seen in Table 12 is that the vowel quality reflexes of the medial voicing contrast as reflected in F1 and F2 are reliably associated with V_1. This supports the idea that there is a compulsion for the stem vowels to harmonize in Javanese. The absence of any difference in F0 indicates that the harmony is restricted to vowel quality.

Table 12. V_1 mean (st dev) for RaCaC stems

factor	RaBaC		RaPaC		regression (t)	
duration (ms)	320	(56)	256	(43)	–5.96	[16]
F1 (Hz)	1040	(43)	1123	(55)	6.83	[17]
F2 (Hz)	1570	(50)	1615	(57)	3.91	[18]
F0 (Hz)	201	(13)	200	(10)	–0.57	[19]

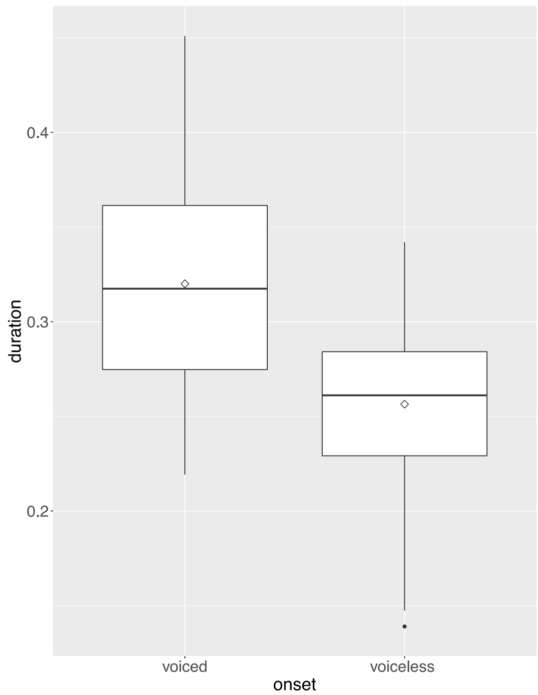


Figure 16. V_1 Duration (ms) RaCaC

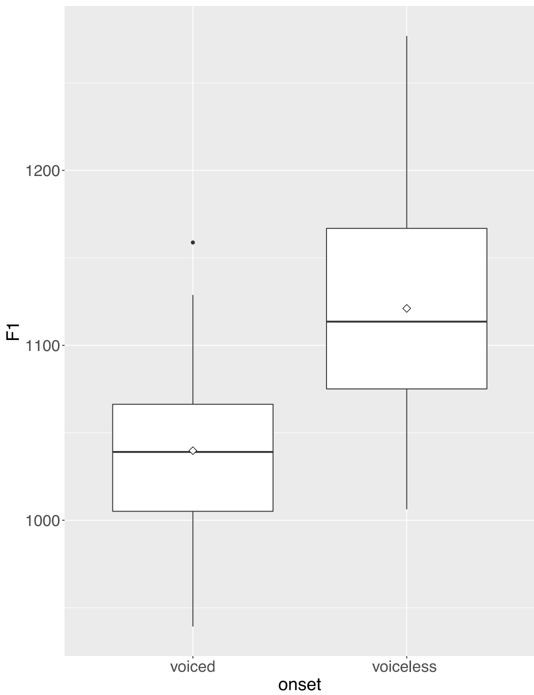


Figure 17. V_1 F1 (Hz) RaCaC

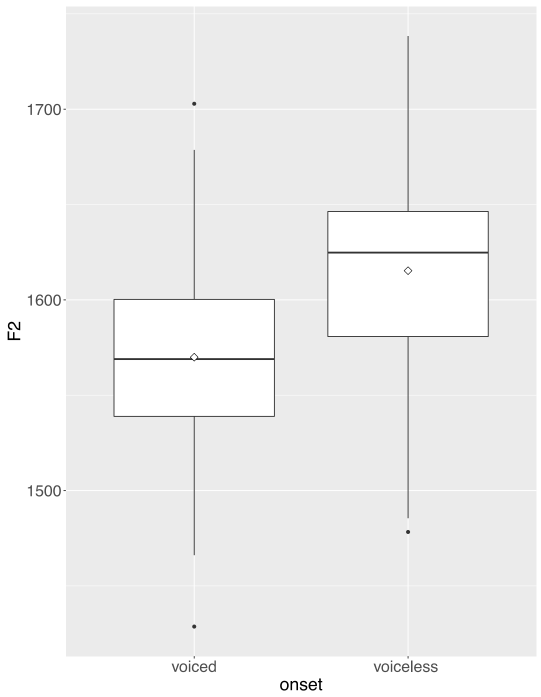


Figure 18: V_1F2(Hz) RaCaC

The spectral correlates for the medial voicing contrast are summarized in Table 13 and Figure 18 below. They indicate that there is a marginally significant effect for H1–H2 on the vowel following the medial stop where voiced stops are associated with a more breathy quality in the vowel. But for the other two measures of H1–A1 and H1–A2, there was no significant effect of voicing, just as in word-initial position.

Table 13. Spectral measures in dB mean (st dev) for V_2

factor	RaBaC		RaPaC		regression (t)	
H1–H2	5.48	(3.27)	3.69	(3.8)	–2.1	[20]
H1–A1	16.62	(6.39)	17.75	(3.7)	0.78	
H1–A2	10.92	(6.63)	14.0	(5.48)	1.77	

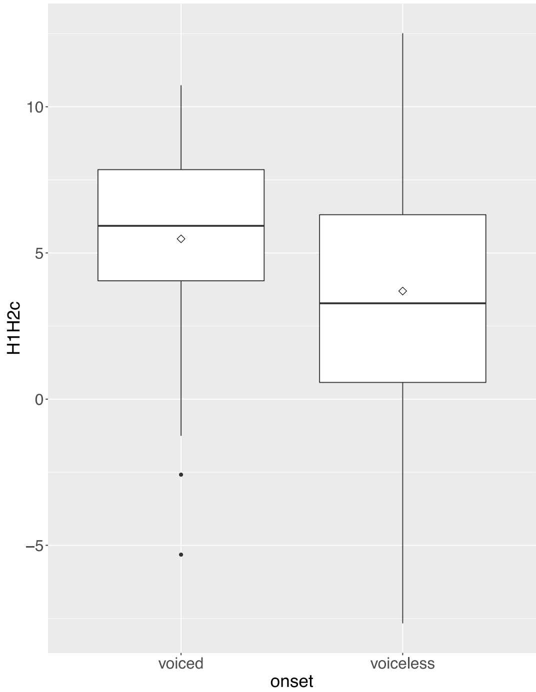


Figure 19: V_2 H1–H2 (dB)

8. Word-final position

To investigate the realization of the Javanese stop-voicing contrast in word-final position, another wordlist consisting of 33 disyllabic items of the shape CaCaC was constructed (see Appendix set D). These words varied their final stop between voiced and voiceless at the three places of articulation: labial, coronal, velar. The orthographic final /k/ was realized as a glottal stop. The words (all nouns) were also elicited in their definite form with the suffix -e to check to what extent the underlying voicing contrast is restored/preserved when the stem-final consonant is placed in intervocalic position.

Word-final stops in related languages such as Madurese are said to be unreleased (Misnaden & Kirby 2017). While this was often true for our data as well, in quite a few cases the stops were released. This was especially true for the underlying voiced stops but also occurred for some of the voiceless ones as well. Since the duration of the preceding vowel was a reliable indicator of the voicing of a following stem-medial stop in our Javanese data (Figure 16), we were interested to see whether the same cue would carry over to word-final position. Also, because the medial consonant of the CaCaC stems in this set varied between sonorant, voiced, and voiceless, we report the measurements separately as a function of this difference. Table 14 and Figure 20 show the mean durations for the stem-final vowel (V_2) as a function of the voicing of the final stop and the ternary sonorant, voiced, voiceless distinction for the medial consonant. As the data indicate, a final voiceless stop is consistently associated with a shorter V_2 except when the medial consonant is voiced, which appears to inhibit this effect. Recall that a trend in the direction of greater duration after a voiced stop was observed for initial and medial positions in Tables 3 and 11, respectively, above.

Table 14. V_2 duration ms (mean and st dev) for CaCaC stems

final C / medial C	sonorant		voiced		voiceless	
voiced	127	(20)	140	(20)	128	(18)
voiceless	112	(17)	141	(15)	97	(21)

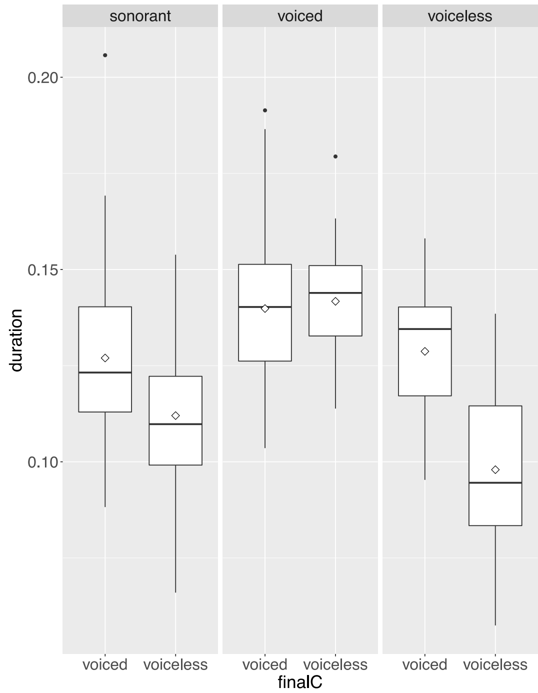


Figure 20: V_2 duration (ms) CaCaC stem

Regression tests found both effects to be significant: Table 15. In this model the baseline was set to a final voiced stop in the context of a medial sonorant. Changing the final consonant to voiceless was associated with significantly shorter V_2 while changing the medial consonant to voiced had a significant lengthening effect on this vowel.

Table 15. Mixed-effects linear regression model: V2 duration (ms)

	Estimate	Std. Error	t value
(Intercept)	0.125459	0.003590	34.945
finalC:voiceless	–0.012555	0.003720	–3.375
medial:voiced	0.018427	0.004183	4.405
medial:voiceless	–0.007670	0.005182	–1.480

Below (Table 16) are the F0 measures for V_2 as a function of voicing in the final stop in the three medial contexts. Although the lowest scores are found with a medial voiced stop, the differences were too small to reach significance (t=1.17; [20]). This held true even when the effect of the final consonant was dropped from the model (t=1.24).

Table 16. F0 of V_2 in Hz (mean and st dev) for CaCaC stems

final C / medial C	sonorant	voiced	voiceless
voiced	216 (27)	210 (32)	227 (31)
voiceless	219 (27)	213 (22)	224 (31)

As for the F1 value of the second stem vowel in this data set, both the medial consonant as well as the final consonant play a significant role. This point is evident from Table 17 and the associated boxplots in Figure 21. When the final consonant is changed from voiced to voiceless, the F1 value increases in all three medial contexts. Similarly, change from a medial voiced stop to voiceless or sonorant also increases F1. Regression modeling in Table 18 suggests that the voicing of the medial stop has a greater effect in both magnitude and reliability.

Table 17. F1 of V_2 in Hz (mean and st dev) for CaCaC stems

final C / medial C	sonorant		voiced		voiceless	
voiced	969	(78)	882	(42)	955	(47)
voiceless	1053	(93)	897	(38)	1058	(44)

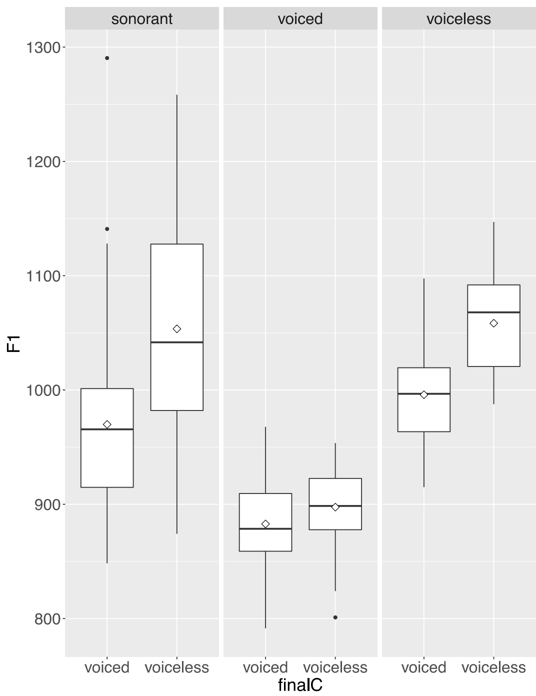


Figure 21: V_2 F1(Hz) RaCaC stem

Table 18. Mixed-effects linear regression model: F1(Hz) of V_2

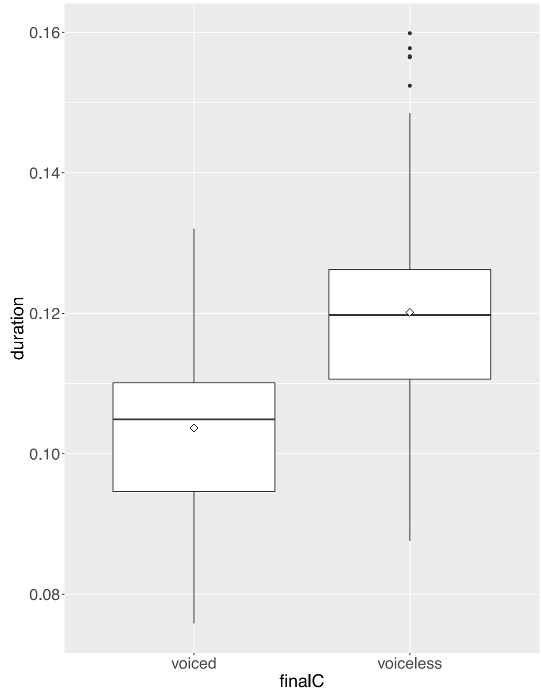
	Estimate	Std. Error	t value
(Intercept)	982.38	15.99	61.438
medial:voiced	–113.79	18.45	–6.169
medialC:voiceless	13.72	22.72	0.604
finalC:voiceless	62.13	16.33	3.804

In sum, for the disyllabic CaCaC stems the duration of the second vowel is reliably associated with the underlying voicing of the final consonant; but this property can be masked when the medial consonant is a voiced stop, which has a lengthening effect on the following vowel. Voiced stops in either the medial or final position tended to decrease F0 in the second stem vowel; but neither effect reached significance. Finally, lowering of the first formant was found in the context of a following word-final voiced stop as well as a preceding medial one. The latter effect was greater in both magnitude and statistical significance. Thus, compared to word-initial and word-medial positions, the number as well as the magnitude and statistical reliability of cues to the Javanese stop voicing contrast are reduced in word-final position.

Table 19 indicates the effect of the voicing of the stem-final consonant on the vowel of the definite suffix -e for the various phonetic dimensions of interest. The second last row also shows the duration of the stem-final stop before the definite suffix. The only factor that reaches significance is the duration of the stem-final stop, which is longer when it is voiceless. This is plotted in Figure 21 below. Figure 22 indicates the duration of the second stem vowel as a function of the voicing of the stem final consonant crossed with the voicing category of the medial consonant in the forms with the definite suffix. The data reveal that V_2 duration is decreased before a stem final voiceless stop but that this effect is minimized in the presence of a medial voiced stop—the same duration pattern seen in the unsuffixed forms in Figure 19 above.

Table 19. Suffixal -e in CaCaC-e

factor	voiced		voiceless		regression (t)	
F1	548	(77)	550	(49)	0.05	
F2	2421	(127)	2446	(183)	0.81	
F0	215	(27)	223	(30)	1.23	
e-duration	140	(26)	137	(23)	–0.70	
stop duration	104	(11)	121	(14)	6.15	[21]
V_2 duration	125	(19)	107	(21)	–2.98	



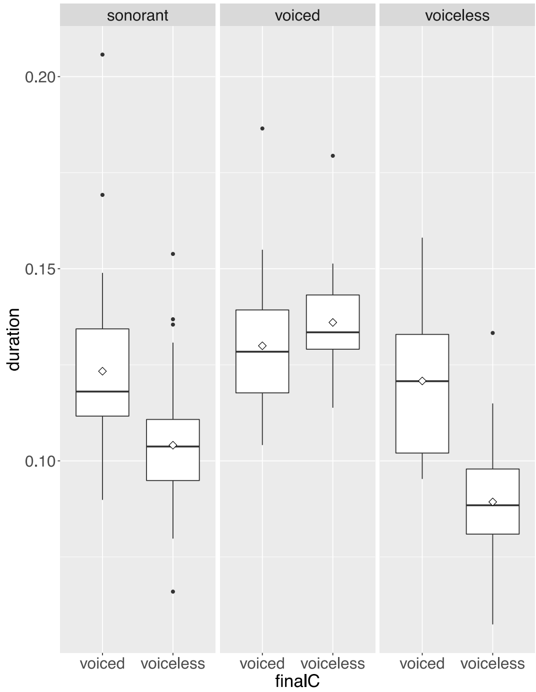


Figure 22. C_3 duration (ms) CaCaC-e **Figure 23. V_2 duration (ms) CaCaC-e**

Regression testing finds that the voicing of both the word-final as well as the word-medial consonant significantly affects the duration of the second stem vowel. In this model (Table 20) a final voiced stop and medial sonorant were the baseline categories. Changing the final consonant to voiceless significantly decreases the duration of the preceding vowel. And changing the medial consonant from baseline sonorant to voiced significantly increases the duration of the following vowel. This finding suggests that the stop external, vocalic reflexes of the voicing contrast in Javanese stops may be restricted to the stem since they were not found for the -e definite suffix (Table 19). Testing with a larger variety of suffixes is needed to determine how systematic this phenomenon is.

Table 20. Mixed-effects linear regression model: duration of V_2 in CaCaC-e

	Estimate	Std. Error	t value
(Intercept)	0.119809	0.004682	25.590
finalC:voiceless	–0.013873	0.004646	–2.986
medial:voiced	0.016130	0.005247	3.074
medial:voiceless	–0.009573	0.006466	–1.480

Table 21 below summarizes the various factors correlated with the Javanese stop voicing contrast investigated in our study as a function of the location of the stop in initial, medial, and word final positions. Yes indicates that a significant difference was found; no indicates that no significant difference was found. The results are consistent with the intervocalic > word-initial > word-final hierarchy found for the neutralization hierarchy in the typology of laryngeal features proposed in Steriade (2009).

Table 21. Distribution of cues to the Javanese stop voicing contrast

	Word-initial	Word-medial		Word-final	
		V_1	V_2	V_2	-e
Stop duration	no	no			yes
VOT	yes	no		no	
F0	yes	no	yes	no	no
F1	yes	yes	yes	yes	no
F2	no	yes	no		no
Duration	no	yes	no	yes	no
H1–H2	yes		yes		
H1–A1	no		no		
H1–A2	no		no		

With respect to this intervocalic > word-initial > word-final hierarchy, we were curious whether a difference would emerge in the magnitude of the major phonetic correlates of the voicing contrast. Given our data, we can compare V_1 in the CaRaC stems of set A with V_2 in the medial RaCaC stems of set C for F1 and F0 as a function of the voicing of the preceding stop. We can also compare V_1 in medial CaCaC stems of set B with V_2 in final CaCaC stems of set D for duration as a function of the voicing of the following stop. To make these comparisons, we normalized the V_1 and V_2 vowels with z-scores for their F1, F0, and duration values. We then ran regression tests with stop voicing and stop position as predictors. In these tests, the baselines were set to initial position and voiced (B) for the parameters of F1 and F0. And for the parameter of preceding vowel duration, the baselines were final and voiced (B). As seen in Tables 22–25, in none of the three comparisons was stop position significant while stop voicing continued to be so. While the absence of an effect for position might be overshadowed by stem harmony for F1, there is no harmony for the prosodic features of duration or F0 in Javanese. This suggests that there is no trading relation among the major correlates for the Javanese stop voicing contrast.

Table 22. Mixed-effects regression test of F1 (normalized) as a function of stop voicing and stop locus

	Estimate	Std. Error	t value
(Intercept)	–0.68728	0.09734	–7.061
locus:medial	–0.05943	0.11273	–0.527
onset:P	1.48536	0.11442	12.981

Table 23. Mixed-effects regression test of F0 (normalized) as a function of stop voicing and stop locus

	Estimate	Std. Error	t value
(Intercept)	–0.4889	0.1285	–3.805
locus:medial	–0.0951	0.1265	–0.752
onset:P	1.1969	0.1275	9.384

Table 24. Mixed-effects regression test of duration (normalized) as a function of stop voicing and stop locus

	Estimate	Std. Error	t value
(Intercept)	0.59063	0.19548	3.021
locus:medial	–0.09439	0.18423	–0.512
onset:P	–0.95838	0.18092	–5.297

9. Summary discussion and conclusions

The goal of this study was to replicate and extend previous research on the phonetic correlates to the stop voicing contrast in Javanese. Like its sister languages Madurese and Sundanese, Javanese lacks the internal cues of closure voicing and VOT and realizes the contrast in virtue of its effects on the quality, pitch, and phonation of the following vowel. This reliance on external cues makes these languages worthy of special attention. Of particular interest is the factor of vowel height, which has been phonologized into a split of the phoneme inventory in Madurese where it also serves as the basis for a stem harmony process. We endeavored to extend the investigation by exploring the realization of the Javanese voicing contrast in word-medial and word-final contexts. Finally, our study sampled a greater range of data than earlier investigations and employed consistent tests of statistical significance.

We examined three positions in the word: initial, medial, and final. For initial position, our findings largely coincide with earlier results. There was a large and reliable difference in vowel height as a function of stop voicing; this difference was observed not only for the low vowels but also for the central and high vowels. In the latter respect our results align with the findings of Kulikov (2010) and Perwitasari et al. (2017) for Sundanese and differ from Thurgood's (2004) only with respect to [u], which she reported to be lower in F1*F2 space after voiced stops. But unlike prior research we only found a minimal difference in F2. And as in many other languages, F0 was reliably lower after voiced stops. We also investigated the spectral properties of H1–H2, H1–A1, H1–A2 to determine their role in supporting the voicing contrast. For the lower region of the spectrum measured by H1–H2, our subject evidenced a breathy phonation profile for the voiced stops similar to Hayward's (1993) findings for Javanese and unlike Kulikov's (2010) results for Sundanese. Also different from Sundanese, we did not find any significant differences for H1–A1 and H1–A2. The same general pattern of results carried over to word-medial position: the stop-internal correlates of duration, closure voicing, and VOT did not distinguish the voicing contrast. Rather the burden fell primarily on the following vowel, which differed in vowel height and F0 in ways comparable to word-initial position. But an additional factor supporting the voicing contrast word-medially was the duration of preceding vowel, which was reliably longer before voiced as opposed to voiceless stops. Here our results differ from Fagan (1988) who did not find duration of the preceding vowel to be a significant factor distinguishing the heavy (voiced) vs. light (voiceless) contrast. For word-final position there is no following vowel to carry the cues and so a merger of the stop voicing contrast is expected. The only factor available to signal the contrast is the preceding vowel and in fact its duration patterned with the treatment of word-medial stops showing longer vowels before an underlying voiced stop compared to a voiceless one. However, this effect was overshadowed by the voicing of the medial consonant, which had a greater effect on the duration of V_2. Finally, when the stem was followed by the definite suffix -e, the voicing contrast in the stem-final stop emerged as a difference in closure duration with underlying voiceless stops reliably longer. However, we did not find that factors of vowel height or pitch showed up on this suffixal vowel, suggesting that these properties may be restricted to the stem. In sum, the overall correlates to the stop voicing contrast in Javanese exemplify the cross-linguistically common profile of cue distribution noted in Steriade's (2009) well-known study: intervocalic > word-initial > word-final.

Typologically, the effects of stop voicing on the F0 of a following vowel and of duration on a preceding vowel are not at all unusual. What is more striking is the effect on vowel quality and to some extent on phonation as well. If Cohn (1993) is correct that the lower F1 of Madurese is to be attributed to larynx lowering then one is reminded of implosives: they are canonically voiced and the larynx lowering is a strategy to sustain vocal fold vibration in the face of an oral occlusion by increasing the volume of the back cavity to minimize the transglottal pressure differential. But mysteriously, vocal fold vibration is precisely what is missing in Javanese voiced stops. One might conjecture that implosive voicing occurred at an earlier stage of the language, which was then transformed into a phonation difference. Another possible implication of the loss of a direct connection between voicing and vowel height as reflected in F1 may be the phonologization of this factor in Madurese. The fact that the F1 difference associated with voicing has been extended both progressively and regressively by stem harmony in our Javanese data supports this interpretation as well. If true, it would imply that the stem harmony process operates at an abstract level of phonological feature structure comparable to the root node that dominates the features of tongue body height, rounding, as well as the tense vs. lax distinction in mid vowels (Dudas 1976). Of course, the prosodic feature of F0 is not part of the vowel copy phenomenon since it reflects the intonation contour of the word, which was largely rising due to the list format for the elicitation of our data.

Tasks for future research include testing with a larger number of subjects as well as further articulatory study to document and more directly investigate the largely hypothetical larynx-lowering gesture that has been postulated to underlie the F1 correlate to the voiced stops as well as the range of phonation effects that at this stage of our knowledge appear rather variable and inconsistent. The perceptual correlates of the voicing contrast should also be investigated by cross-splicing of vowels as well as through manipulation of their F0 and F1 values.

References

Adisasmito-Smith, Niken. 2004. *Phonetic and phonological influences of Javanese on Indonesian*. Ph.D. thesis. Ithaca, N.Y.: Cornell University.

Bates, D. & M. Maechler. 2010. Linear mixed-effects models using S4 classes. Version 0.999375-37.

Boersma, Paul & David Weenink. 1992–2017. Praat: doing phonetics by computer. Version 6.0.33. http://www.praat.org.

Brunelle, Marc. 2010. The role of larynx height in the Javanese tense~lax stop contrast. In Mercado, Ralph, Eric Potsdam & Lisa Travis (eds.), *Austronesian and Theoretical Linguistics*. 7–24. Amsterdam: John Benjamins.

Brunelle, Marc & James Kirby. 2016. Tone and phonation in Southeast Asian languages. *Language and Linguistics Compass* 10. 191–207.

Catford, John. 1964. Phonation types. In David Abercrombie (ed.), *In honour of Daniel Jones*. 26–37. London: Longmans.

Catford, John. 1977. *Fundamental problems in phonetics*. Bloomington: Indiana University Press.

Coetzee, Andries, Patrice Speeter Beddor, Kerby Shedden, Will Styler & Daan Wissing. 2018. Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics* 66. 185–216.

Cohn, Abigail. 1993. Consonant-vowel interaction in Madurese: the feature [Lowered Larynx]. *Chicago Linguistics Society* 29. 105–119.

Cohn, Abigail & Katherine Lockwood. 1994. A phonetic description of Madurese and its phonological implications. *Working papers of the Cornell phonetics laboratory* 9. 67–92.

Dudas, Karen. 1976. *The phonology and morphology of modern Javanese*. Ph.D. thesis. Urbana, Il: University of Illinois.

Fagan, Joel L. 1988. Javanese intervocalic stop phonemes: the light/heavy distinction. In Richard McGinn (ed.), *Studies in Austronesian linguistics* 76. 173–202.

Golston, Christopher & Wolfgang Kehrein. 2004. A prosodic theory of laryngeal contrasts. *Phonology* 21. 325–357.

Hayward, Katrina. 1993. /p/ vs. /b/ in Javanese: some preliminary data. *SOAS Working Papers in Linguistics & Phonetics* 3. 1–33.

Hayward, Katrina. 1995. /p/ vs. /b/ in Javanese: The role of the vocal folds. *SOAS Working Papers in Linguistics & Phonetics* 5. 1–11.

Horne, Elinor Clark. 1974. *Javanese-English dictionary*. New Haven: Yale University Press.

Kulikov, Vladimir. 2010. Voicing and vowel raising in Sundanese. Presented at the 17th Annual Meeting of the Austronesian Formal Linguistics Association, Stony Brook, NY.

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Malden, MA: Blackwell.

Matthews, Marc. 2017. An acoustic investigation of Javanese stop consonant clusters. In *The Proceedings of the 21st Meeting of the Austronesian Formal Linguistics Association*. 201–218. Canberra: Australian National University.

Misnadin & James Kirby. 2017. *Journal of the International Phonetic Association* 47. 1–18.

Perwitasari, Arum, Marian Klamer & Niels O. Schiller. 2017. Quality of Javanese and Sundanese vowels. *Journal of the Southeast Asian Linguistics Society* 10. 1–9.

Shue, Yen, Patricia Keating, Chad Vicenik, and Kristine Yu. 2011. VoiceSauce: a program for voice analysis. *Proceedings of the International Congress of Phonetic Sciences* XVII. 1846–1849. [http://www.phonetics.ucla.edu/voicesauce/]

Steriade, Donca. 2009. The P-map and its consequences for constraint organization. In Kristin Hansen & Sharon Inkelas (eds.), *The nature of the word*. 151–179. Cambridge, MA: MIT Press.

Thurgood, Ela. 2004. Phonation types in Javanese. *Oceanic Linguistics* 43. 277–295.

Trigo, Loren. 1991. On pharynx-larynx interactions. *Phonology* 8. 113–136.

Uhlenbeck, E. M. 1950. The structure of the Javanese morpheme. *Lingua* 2. 239–270.

Xu, Yi. 2020. Formant Pro. A Praat script for large-scale systematic analysis of continuous formant movements (Version 1.4.2).

Appendices

[A] test items

gloss	word	gloss	word
Set A: CaRaC, CaRa			
internal	dhalam	fetus	gana
army troops	bala	to offer peddle	tawa
land	tanah	tense	gawat
to race	balap	nutmeg	pala
fault	galap	together with	dalah
coral	karang	nickname	parab
fish species	bawal	Eve	kawa
a pity, shame	talah	commotion	gara
garlic	bawang	precious	dama
uncle	paman	food	pangan
variant of upama	pama	arrow	panah
to lose	kalah	pigeon	dara
salted	kamal	something carried	gawa
charitable gift	dana	think	barang
honeycomb	tala	hand to dry	tarang
leprosy	barah	over there	kana
good luck	bara	room	kamar
the end	tamat	danger, misfortune	gama
climbing vine	kara	season	kala
interval	tara	to fall	dhawah
love, passion	kama	the kings	para
ineffectual	tawar	to get struck	tama
to fear that	gamar	wire	kawat

handle	garan	situation	bawa
strong	ganal	to accuse	daran
demand	bana	land	dharat
hot	panas	road	dalan
to know (O.Jav)	pana	long	dawa

Set B: CiCəC, CuCəC, CəCəC

cooked rice	liwêt	graceful, smooth	luwês
bottom	kurêb	overpowering	dulêg
to show	dêlêng	soft	bêrêm
speech	gunêm	be willing	gêlêm
dim, vague	rêmêng	frequent	kêrêp
abundant	biyêt	to mash	gilês
deep sleep	pulês	fed up	tumêg
stay by self	dilêp	tangled	ruwêd
oyster	tirêm	have a cold	pilêg
wet	têlês	solid	bulêt
mango	pêlêm	to narrow the eyes	kiyêr
to subside	lêrêb	melodious	wilêt

Set C: RaCaC

know by heart	apal	behavior	lagak
eat up	labas	sacrificial animal	wadal
bad	nakal	parrot	atat
harsh	ladak	barrier	athak
usual	racak	dye sediment	latak

container	wadhah	plain	wajar
pickles	acar	leaky	rajag
complaint	ratap	meeting	rapat
breath	napas	reptile skeleton	ragas
discount	rabat	groin	lakang
laugh	lakak	visible	wadhak
trace	lacak	saddle	lapak
growing fast	lagang	intercourse	wajang
do without trying	ngabas		

Set D: CaCaC

temporary roof	tratag	century	abad
nickname	parab	cobweb	lamat
wide opening	tjangap	slivers	silad
nephew	anak	a need	adjat
breath from mouth	abab	Arab	arab
saddle	lapak	slave	arad
tripe	babat	nerve	sarap
bamboo board	tabag	cover	sasab
verses in the Koran	ajat	die sediment	latak
a race	balap	hawk	alap
palm, sole	tlapak	plants set among others	adjag
raw vegetables	lalab	fault	galap
pretense	awad	household equipment	abrag
custom	adat	history	babad
fish net	ajab	convulsions	sarab
forest	alas	alcholic beverage	arak

land	dharat

Summary of R-models (Linear mixed-effects models fit by REML ['lmerMod'])

Word-initial position

[1] F1 on vowel [a]

Formula: F1 ~ onset + (1 | word) + (1 | trial) + (1 + onset | word)

Number of obs: 157, groups: word, 35; trial, 5

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	948.77	11.74	80.798
onset:voiceless	134.22	13.50	9.939

[2] F1 on vowel [ɔ]

Formula: F1 ~ onset + (1 | word) + +(1 | trial)

Number of obs: 123, groups: word, 25; trial, 5

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	727.31	11.85	61.362
onset:voiceless	77.61	15.67	4.952

[3] F2 on vowel [a]

Formula: F2 ~ onset + place + (1 | word) + (1 | trial) + (1 + onset | word)

Number of obs: 157, groups: word, 35; trial, 5

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	1583.456	46.636	33.954
onset:voiceless	41.925	26.106	1.606
place:labial	-2.147	57.112	-0.038
place:retroflex	-8.489	88.171	-0.096
place:velar	-75.791	55.565	-1.364

[4] F2 on vowel [o]

Formula: F2 ~ onset + place + (1 | word) + (1 | trial) + (1 + onset | word)

Number of obs: 123, groups: word, 25; trial, 5

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	1085.542	20.681	52.490
onset:voiceless	-5.543	18.895	-0.293

place:labial	-54.789	23.761	-2.306
place:velar	-29.922	23.018	-1.300

[5] F0 initial

Formula: F0 ~ onset + (1 | word) + (1 | trial) + (1 + onset | word)

Number of obs: 280, groups: word, 58; trial, 5

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	194.807	2.113	92.18
onset:voiceless	9.984	1.157	8.63

[6] V1 duration initial

Formula: duration ~ onset + (1 | word) + (1 | trial) + (1 + onset | word)

Number of obs: 280, groups: word, 58; trial, 5

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	0.27346	0.01091	25.059
onset:voiceless	-0.01658	0.01145	-1.448

[7] V2F1 [a]

Formula: F1V2 ~ onset + (1 | word) + (1 + onset | word)

Number of obs: 156, groups: word, 35

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	927.74	11.53	80.476
onset:voiceless	126.08	18.65	6.759

[8] V2F1 [ɔ]

Formula: F1V2 ~ onset + (1 | word) + (1 + onset | word)

Number of obs: 123, groups: word, 25

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	758.24	12.15	62.391
onset:voiceless	83.41	16.98	4.912

[9] F1 for schwa and high vowels

Formula: F1 ~ onset + segment + (1 | word) + (1 + onset | word)

Number of obs: 121, groups: word, 27

Fixed effects:	Estimate	Std. Error	t value

(Intercept)	741.928	19.102	38.840
onset:vcd	-70.999	18.730	-3.791
onset:vcl	-5.275	21.530	-0.245
segment:i	-270.387	8.852	-30.545
segment:u	-226.769	8.822	-25.704

[10] F2 for schwa and high vowels

Formula: F2 ~ onset + segment + (1 | word)

Number of obs: 81, groups: word, 19

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	1494.34	65.48	22.822
onset:vcd	37.14	75.27	0.493
onset:vcl	-54.72	72.64	-0.753
segment:u	-425.44	58.20	-7.310

[11] F0 for schwa and high vowels

Formula: F0 ~ onset + (1 | word)

Number of obs: 121, groups: word, 29

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	232.732	3.276	71.044
onset:voiced	-21.555	4.351	-4.954
onset:voiceless	-1.775	4.232	-0.419

Word-medial position

[12] V2 F1

Formula: F1 ~ onset + (1 | word) + (1 | trial) + (1 + type | word)

Number of obs: 136, groups: word, 31; trial, 5

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	886.781	8.069	109.893
onset:voiceless	112.674	12.111	9.304

[13] V2F2

Formula: F2 ~ type + (1 | word) + (1 | trial)

Number of obs: 136, groups: word, 31; trial, 5

Fixed effects:	Estimate	Std. Error	t value

(Intercept)	1692.38	28.31	59.782
onset:voiceless	-49.66	36.98	-1.343

[14] F0V2

Formula: F0 ~ onset + (1 | word) + (1 | trial) + (1 + type | word)

Number of obs: 132, groups: word, 31; trial, 5

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	225.543	4.724	47.746
onset:voiceless	28.398	4.861	5.843

[15] V1 duration

Formula: duration ~ onset + (1 | word) + (1 | trial)

Number of obs: 136, groups: word, 31; trial, 5

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	0.321011	0.009457	33.943
onset:voiceless	-0.065759	0.011042	-5.955

[16] V1F1

Formula: F1 ~ onset + (1 | word) + (1 | trial)

Number of obs: 136, groups: word, 31; trial, 5

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	1040.90	11.22	92.776
onset:voiceless	82.57	12.08	6.833

[17] V1F2

Formula: F2 ~ onset + (1 | word) + (1 | trial)

Number of obs: 136, groups: word, 31; trial, 5

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	1571.16	13.65	115.077
onset:voiceless	45.38	11.58	3.919

[18] V1F0

Formula: F0 ~ onset + (1 | word) + (1 | trial) + (1 + type | word)

Number of obs: 136, groups: word, 31; trial, 5

Fixed effects:	Estimate	Std. Error	t value

(Intercept)	201.729	3.449	58.481
onset:voiceless	-1.401	2.445	-0.573

[19] H1–H2 medial

Formula: H1H2c ~ onset + (1 | word) + (1 | trial)

Number of obs: 128, groups: word, 33; trial, 5

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	5.4561	0.7208	7.569
onset:voiceless	-1.7648	0.8521	-2.071

Word-final position

[20] F0 final

Formula: F0 ~ medialC + finalC + (1 | word) + (1 | trial)

Number of obs: 264, groups: word, 66; trial, 4

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	217.4708	6.1620	35.292
medialC:voiced	-8.5567	7.3148	-1.170
medial:voiceless	6.7466	9.0113	0.749
finalC:voiceless	0.3459	6.4762	0.053

[21] Duration of C2 in suffixed form

Formula: duration ~ finalC + (1 | word) + (1 | trial)

Number of obs: 164, groups: word, 33; trial, 5

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	0.103664	0.002188	47.381
finalC:voiceless	0.016525	0.002683	6.158