ACOUSTIC CORRELATES OF THE JAVANESE LIGHT VS. HEAVY DISTINCTION: A LARGE-SCALE CORPUS STUDY

S.C. Angela Xu, Colin Wilson

Johns Hopkins University
angela.xu@jhu.edu, colin.wilson@jhu.edu

ABSTRACT

According to the influential *continuum model* of phonation, only voiced segments can be specified as creaky or breathy. The present study investigated many possible phonetic correlates of the laryngeal contrast in Javanese word-initial prevocalic stop consonants, drawing upon a spoken corpus of more than 180,000 utterances. The results indicate that the laryngeal contrast is cued by voice onset time (VOT) and several acoustic-phonetic properties of the following vowel, including the first formant (F1) in addition to voice source measurements such as H1*-H2* and cepstral peak prominence (CPP). Taken together these findings indicate that Javanese stops can be both voiceless and breathy, supporting a revision of the continuum model in which voicing and other aspects of phonation are decoupled.

Keywords: laryngeal contrast, acoustic correlates, phonation type, Javanese, corpus phonetics.

1. INTRODUCTION

Javanese is a relatively understudied Malayo-Polynesian language of the Austronesian family with approximately 100 million speakers, mostly concentrated on the island of Java in Indonesia [1]. Several dialects of Javanese have been distinguished in the literature. Javanese also has several speech levels, including Krama/Basa (formal) and Ngoko (less formal). This study examines the speech of individuals recorded in Yogyakarta, Indonesia, and focuses on the Ngoko level that is typical of informal settings (e.g., [2]).

The laryngeal contrast in Javanese stop consonants has proven difficult to characterize. The IPA symbols /p t ʈ k/ and /b d ɖ ɡ/ have been used to represent the stops, suggesting a contrast in phonetic voicing. However, previous literature has referred to the two groups of stops in various ways, indicating a contrast that is not, or at least not only, a difference in the presence and timing of vocal fold vibration. Here we follow a tradition of using the terms 'light' vs. 'heavy' ([3, 2, 4, 5]), while others have described the contrast as 'intensive' vs. 'non-intensive' ([6]), 'clear' vs. 'breathy' ([7]), 'tense' vs. 'lax' ([8, 9]), 'stiff' vs. 'slack' ([10, 11, 12]), or 'voiceless' vs. 'voiced' ([13]).

The Javanese laryngeal contrast can be considered in the context of the influential *continuum model* of phonation types, first proposed in [14] and later revised in [10, 15]. Fig. 1 gives a visual schematic of the voicing and phonation terms and their relation with the states of the glottis, as described in [10]. Notice that 'voiceless' refers to lack of vocal fold vibration regardless of glottal aperture. According to this model, classifying a stop as one of the five phonation types (creaky through breathy) implies that it is voiced.
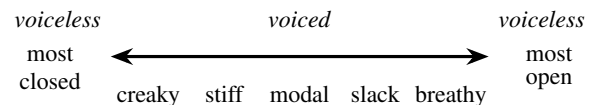

Figure 1: Voicing and phonation types as described in the continuum model.

Problematically for the continuum model, [16] found that glottal stops are frequently (partially) voiced and that the phonetic difference between the so-called 'voiceless' glottal stop and 'voiced' creaky sounds is not one of voicing. The authors hence argued that the model should be modified so that voicing is unspecified for glottal consonants. The laryngeal contrast in Javanese stops may motivate a similar decoupling of voicing and phonation.

The rest of this section reviews previous studies (from the last 20 years) on the acoustic correlates of the Javanese contrast, focusing on cross-linguistically common acoustic-phonetic correlates of voicing and phonation contrasts. For voicing, common cues include a delay of vocal fold vibration in a following sonorant or vibration during closure, as measured by positive and negative voicing onset time (VOT) of stops, as well as the duration and fundamental frequency (f0) of neighboring vowels. For phonation, [17] examined the phonetic space of vowel contrasts with data from 11 different

languages, and concluded that the most informative parameters include the formant-corrected amplitude difference between the first and second harmonics (H1*-H2*), cepstral peak prominence (CPP), the subharmonic-to-harmonic ratio (SHR), and the harmonic-to-noise ratio in the range 0-500 Hz (HNR05), with amplitude difference between the first harmonic and the harmonic closest to the first formant (H1*-A1*), strength of excitation (SoE), and root-mean-squared (RMS) energy also being informative.

[11] recorded 24 words from one Javanese speaker to compare word-initial labial and velar stops /p b k g/ before the vowels /a o u/. Each word was produced twice in isolation. The fundamental frequency (f0), first formant (F1), second formant (F2), and amplitudes of H1, H2, F1, and F2 were measured in post-stop vowels. The results showed negative H1-H2 and H1-F1 values for vowels after heavy consonants (which [11] refers to as 'slack'); this would indicate modal rather than slack or breathy voice. However, H1-F2 values were mostly positive (except for /a/ after light stops) and significantly higher after light stops, which would support the categorization of heavy consonants as slack. F1 was lower after heavy stops for /a o/, and F2 was consistently higher after heavy stops for all three vowels. (See [11] for a review of earlier investigations by Fagan [4] and Hayward [18, 19].)

[8] compared the articution of Javanese light and heavy stops in fiberoptic recordings of two speakers (using data originally collected by [20]). The epiglottal width was significantly greater for heavy ('lax') stops regardless of place of articulation. To the extent that this indicates larynx lowering, it could to some extent explain the longer VOT of heavy stops and their effects on f0, F1, and voice source measures in following vowels.

[12] claimed that the Javanese contrast is one of phonation (i.e., stiff vs. slack) on the basis of acoustic measurements in stop-vowel and stop-lateral-vowel productions from one speaker. Heavy consonants induced lower f0 and formant frequencies, and higher H1-H2 and H1-F2 values, supporting their characterization as slack. These effects varied to some extent across vowel qualities, but importantly persisted through intervening lateral approximants. As [12] suggests, this is consistent with the presence of contrastive phonation on Javanese stops that is spread to the following vowel.

[21] described the stop contrast as 'tense' vs. 'lax' and investigated both the acoustic correlates of this contrast and whether it is neutralized word-finally. Measurements of VOT, f0, F1, F2, the

bandwidth of the first formant (B1), H1*-H2*, and CPP were obtained from recordings of 28 speakers who produced words in a carrier sentences. The results showed that f0, F1, and CPP were lower, while B1 was higher, in vowels following heavy ('lax') stops; underlying light and heavy stops were not distinguished at the end of the word. Word-initial heavy stops had negative VOT for some but not all speakers. These findings support the claim that heavy stops are breathier than their light counterparts, a phonation contrast that exists before sonorants but not word-finally (see also [2]).

Lastly, [13], stating the contrast in Javanese stops as one of voicing, recorded a list of words mainly obtained from [22] produced in isolation by one speaker. Measurements were taken for stop VOT, and duration, f0, F1, F2, H1-H2, H1-A1, and H1-A2 of the following vowel. The results showed that in word-initial context both light and heavy stops have positive VOT, suggesting that all of the stops are voiceless, with longer values for heavy /b d g/ than light /p t k/. Among other findings, measurements of F1 in following /i e u/ agreed with previous reports that this is a reliable correlate of the laryngeal contrast.

The present study aimed to provide further evidence about the phonetic nature of the laryngeal stop contrast in Javanese by examining several possible acoustic correlates in a large-scale speech corpus. We first describe the corpus and the measurements that were considered, then report the statistical analysis of each measure, and conclude with a discussion of how the findings bear on phonation in Javanese and cross-linguistically.

2. METHODS

2.1. Corpus description

In this paper, we report acoustic-phonetic analyses of a Javanese speech corpus originally provided by Google for the purpose of training automatic speech recognition systems (henceforth Google ASR; [23]). The corpus was collected by Google in collaboration with the Javanese Literature Department of Universitas Gadjah Mada (UGM) in Yogyakarta, Indonesia, who trained local volunteers to make scripted recordings from native speakers using custom data collection tools.

The written recording prompts were taken from open online resources and used the modern Javanese writing system based on the Latin alphabet. The corpus contains approximately 185,000 utterances from more than 1000 speakers, with an average utterance length of about ten words. Demographic

information, such as dialect of Javanese and additional language background, was not provided. The audio was recorded in 16-bit linear PCM with a sample rate of 16kHz. A pronunciation lexicon containing all words from the corpus with their phonemic transcriptions was generated using a grapheme-to-phoneme (G2P) model.

2.2. Corpus preparation

We selected the 50 speakers who had the most recorded utterances in the data set. After listening to sample utterances from each speaker, we decided to exclude 6 of them for reasons such as significant background noise. For the remaining speakers (44), audio files were aligned to their transcriptions using custom acoustic models trained with the Montreal Forced Aligner (MFA, version 2.0.0b3) ([24]).

2.3. Acoustic analysis

The release of each word-initial, prevocalic stop consonant /p t k b d g/ (omitting rarer /ʈ ɖ/) and the onset of voicing in the following vowel were located with AutoVOT ([25]). The difference between these two time points provides an automatic measure of positive VOT. A separate AutoVOT model was trained for each stop from a minimum of 30 hand-labeled word tokens beginning with the target consonant followed by various vowels. In each of the forced-aligned textgrids, a new tier was added with regions marked within which AutoVOT searched for the point of stop release and the following onset of voicing. The start of each region was at the midpoint of the closure of the word-initial stop and the end was at the midpoint of the following vowel. VOT values for all tokens were extracted using the trained models.

The duration, f0, F1, and F2 of vowels following the word-initial stops were extracted using Praat ([26]) with the FastTrack plugin ([27]). Voice quality measurements, including H1*-H2*, H1*-A1*, CPP, SHR, HNR05, SoE, and RMS energy, were made for the same vowels using VoiceSauce ([28]) with 1 ms frame shifts. FastTrack and VoiceSauce were applied with their standard settings for all speakers. The output of FastTrack provides f0 and formant values for 5 equal-duration time intervals in each token; we averaged the outputs of VoiceSauce in the same intervals and analyze only the central interval here. Frequencies were converted to the mel scale, which is more closely aligned with human auditory perception than traditional frequency values (Hz).

Outliers were identified separately for each acoustic measure. For all measures except vowel formants, the raw data were grouped by speaker and stop consonant; for F1 and F2 the data were grouped by speaker and vowel phoneme (/i e a o u ə/). Within each group, data points 2 standard deviations beyond the mean were taken to be outliers and excluded from data summaries and statistical analyses.

3. RESULTS

Means and standard deviations of F1 and F2 for vowels after word-initial stops are provided in Table 2, with the other measures summarized in Table 1. We first averaged within each speaker and then computed the summary statistics across speakers. Separate linear mixed-effects models were conducted to analyze the influence of the light vs. heavy stop contrast on each measure.

The model for VOT included fixed effects of stop laryngeal specification (light vs. heavy), stop place (labial, coronal, or dorsal), and their interaction, as well as a random intercept and slope of the laryngeal effect for speaker. All fixed factors in this and subsequent models were entered with weighted effect coding. The main effect of laryngeal specification was significant and positive ($\beta_{lar} = 3.85$), indicating that heavy consonants have slightly longer VOT values than light consonants. The main effect of place was also significant, with dorsal stops having significantly longer values ($\beta_{dor} = 9.17$) and, surprisingly, coronals being significantly shorter overall ($\beta_{cor} = -4.50$). The interaction between laryngeal and place distinctions was also significant; post-hoc tests revealed a reliable effect of light vs. heavy effect on the VOT of coronal and dorsal but not labial stops (significance here and throughout is determined by $p < .001$). The same pattern was found in an analysis of log-transformed VOT.

These results are in line with previous reports that the heavy stops /d g/ have longer VOTs than their light counterparts /t k/ ([13]). The lack of a significant difference for heavy /b/ vs. light /p/ could have two explanations. First, labial stops may have frequently had weak bursts that were not detected by AutoVOT. Second, AutoVOT enforces a minimum VOT duration, here set to 5 ms, that may have artificially inflated the measured values for /p/.

The mixed-effects models for the other measures in Table 1 had the same structure. The results indicate a significant effect of light vs. heavy on f0 ($\beta_{lar} = -6.68$), H1*-H2* ($\beta_{lar} = 1.63$), and SHR ($\beta_{lar} = 0.091$) of the following vowel for stops at all three places of articulation. There were also significant effects on log duration and CPP of the

Stop	N	VOT(ms)	Duration(ms)	f0(mel)	H1*-H2*(dB)	CPP(dB)	SHR(dB)
p	4824	9.5 (2.7)	91.2 (17.1)	284.2 (58.4)	4.8 (3.0)	20.8 (2.5)	0.6(0.2)
b	4327	9.7 (2.8)	90.8 (15.9)	275.6 (54.5)	6.7 (3.2)	20.3 (2.3)	0.7 (0.1)
t	5485	11.0 (2.9)	80.0 (14.9)	297.8 (59.8)	4.6 (4.0)	21.9 (2.8)	0.5 (0.3)
d	8464	17.3 (4.6)	86.8 (17.8)	283.9 (57.8)	6.7 (3.0)	20.2 (2.4)	0.7 (0.2)
k	13653	21.4 (4.3)	87.5 (18.2)	288.4 (57.8)	4.4 (3.3)	21.3 (2.2)	0.6 (0.2)
g	1481	31.3 (7.7)	77.4 (15.0)	277.1 (54.5)	7.8 (3.1)	18.6 (2.2)	0.7 (0.1)

Table 1: Acoustic measures of word-initial stops and following vowels.

following vowel, but these held for coronal and dorsal stops only; the same pattern was found for analyses of H1*-A1*, HNR05, SoE, and RMS energy (not shown in the table). The signs of the effects on the voice quality measures are in line with the expectation from previous literature that heavy stops are more breathy than light stops in Javanese. For example, H1*-H2* is larger (i.e., there is greater low-frequency spectral tilt) following heavy stops.

Stops	F1 (mel)	F2 (mel)
	high V	*front V*
/p t k/	539.0 (51.4)	1654.0 (127.1)
/b d g/	496.9 (45.5)	1675.4 (140.8)
	mid V	*central V*
/p t k/	706.2 (62.9)	1366.4 (76.2)
/b d g/	583.6 (51.3)	1413.9 (80.2)
	low V	*back V*
/p t k/	863.7 (78.6)	1038.4 (57.5)
/b d g/	757.0 (69.5)	1205.5 (80.1)

Table 2: F1 and F2 of vowels following word-initial stops.

The separate models for vowel formants included fixed effects of stop laryngeal specification, vowel height for F1 and vowel backness for F2 (as in Table 2), and their interaction. Each model had a random intercept and slope of the laryngeal effect for speaker. F1 was significantly lower after heavy stops ($\beta_{lar} = -60.3$), consistent with larynx lowering ([8]) or other pharyngeal expansions. The laryngeal and height factors interacted significantly, indicating that low and mid vowels show a larger light vs. heavy difference than high vowels, but post-hoc tests established that the laryngeal effect was significant at all three vowel heights.

F2 was significantly higher after heavy stops ($\beta_{lar} = 47.3$), perhaps suggesting some degree of vowel fronting (see also [11]). The laryngeal and vowel backness factors interacted significantly, with non-front vowels showing larger effects of light

vs. heavy than front vowels, but the effect was nevertheless significant and in the same direction for all levels of backness.

4. CONCLUSION

This study investigated the phonetic nature of the laryngeal contrast in Javanese by measuring many possible acoustic-phonetic correlates in more than 36,000 tokens of word-initial stops and following vowels. The results broadly support previous findings that stops of both series have positive VOT in this context, that VOT values are larger for heavy than for light stops (with the possible exception of labials), that voice source measures such as H1*-H2* indicate breathiness on vowels following heavy stops, and that vowels have lower F1 and higher F2 in the same context (see [11, 8, 12, 21, 13]).

According to the continuum model ([14, 10, 15]), only voiced segments can be breathy. Our findings, like those of many previous acoustic studies of the laryngeal stop contrast in Javanese, appear to contradict this claim. Both the light stops /p t k/ and the heavy stops /b d g/ have positive VOT values, consistent with the hypothesis that they are all phonetically voiceless. At the same time, there were also significant differences in voice source measurements such as H1*-H2* and SHR within following vowels, and these are consistent with the hypothesis that the heavy series is breathy.

Along the lines of [16], we suggest that phonation specifications such as slack or breathy are not strictly confined to voiced segments. Indeed, voiceless aspirated stops require the glottis to be abducted, and if this gesture persists after stop release it could naturally induce a breathier voice quality on the following vowel. Because measurements other than VOT were taken in the central interval of each vowel, extension of [+spread glottis] or a related feature from word-initial heavy stops to following vowels is a plausible phonological process in the sound system of Javanese.

ACKNOWLEDGMENTS

5. REFERENCES

[1] D. M. Eberhard, G. F. Simons, and C. D. Fennig, Eds., *Ethnologue: Languages of the World*, 25th ed. Dallas, Texas: SIL International, 2022, online version: http://www.ethnologue.com.

[2] K. M. Dudas, "The phonology and morphology of Javanese," Ph.D. dissertation, University of Illinois at Urbana-Champaign, Urbana, Illinois, 1976.

[3] E. C. Horne, *Beginning Javanese*. New Haven: Yale University Press, 1961.

[4] J. L. Fagan, "Javanese intervocalic stop phonemes: The light/heavy distinction," in *Studies in Austronesian Linguistics*, R. McGinn, Ed. Athens, Ohio: Ohio University Press, 1988, pp. 173–202.

[5] K. M. Hayward and K. M. Muljono, "The dental/alveolar contrast in Javanese," *Bulletin of the School of Oriental and African Studies*, vol. 54, pp. 126–144, 1991.

[6] E. M. Uhlenbeck, "The structure of the Javanese morpheme," *Lingua*, vol. 2, pp. 239–270, 1949.

[7] J. C. Catford, *Fundamental Problems in Phonetics*. Bloomington: Indiana University Press, 1977.

[8] M. Brunelle, "The role of larynx height in the Javanese tense~lax stop contrast," in *Austronesian Contributions to Linguistic Theory: Selected Proceedings of AFLA*, R. Mercado, E. Potsdam, and L. Travis, Eds. John Benjamins, 2010, pp. 7–24.

[9] R. H. Sumukti, "Javanese morphology and morphophonemics," Ph.D. dissertation, Cornell University, Ithaca, New York, 1971.

[10] P. Ladefoged and I. Maddieson, *The Sounds of the World's Languages*. Oxford: Blackwell, 1996.

[11] E. Thurgood, "Phonation types in Javanese," *Oceanic Linguistics*, vol. 43, pp. 277–295, 2004.

[12] M. Matthews, "An acoustic investigation of Javanese stop consonant clusters," in *The Proceedings of the 21st Meeting of the Austronesian Formal Linguistics Association*, C. S. Amber Camp, Yuko Otsuka and N. Tanaka, Eds., 2015, pp. 201–217.

[13] M. J. Kenstowicz, "Phonetic correlates of the Javanese voicing contrast in stop consonants," *NUSA: Linguistic studies of languages in and around Indonesia*, pp. 1–37, 2021.

[14] P. Ladefoged, *Preliminaries to Linguistic Phonetics*. Chicago: University of Chicago, 1971.

[15] M. Gordon and P. Ladefoged, "Phonation types: a cross-linguistic overview," *Journal of Phonetics*, vol. 29, pp. 383–406, 2001.

[16] M. Garellek, Y. Chai, Y. Huang, and M. Van Doren, "Voicing of glottal consonants and non-modal vowels," *Journal of the International Phonetic Association*, vol. 94, pp. 1–28, 2021.

[17] P. Keating, J. Kuang, M. Garellek, C. M. Esposito, and S. Khan, "A cross-language acoustic space for vocalic phonation distinctions," *Language*, To appear.

[18] K. M. Hayward, "/p/ vs. /b/ in Javanese: Some preliminary data," in *SOAS Working Papers in Linguistics and Phonetics 3*, pp. 1–99.

[19] ——, "/p/ vs. /b/ in Javanese: The role of the vocal folds," in *SOAS Working Papers in Linguistics and Phonetics 5*, pp. 1–11.

[20] K. M. Hayward, D. Grafield-Davies, B. J. Howard, J. Latif, and R. Allen, "Javanese stop consonants: the role of the vocal folds. School of Oriental and African Studies [video]." 1994.

[21] S. Seyfarth, J. Vander Klok, and M. Garellek, "Acoustics of the tense-lax stop contrast in Semarang Javanese," *The Journal of the Acoustical Society of America*, vol. 142, no. 4_Supplement, pp. 2581–2581, 10 2017.

[22] E. C. Horne, *Javanese-English Dictionary*. Yale University Press, 1974.

[23] O. Kjartansson, S. Sarin, K. Pipatsrisawat, M. Jansche, and L. Ha, "Crowd-sourced speech corpora for Javanese, Sundanese, Sinhala, Nepali, and Bangladeshi Bengali," in *Proc. The 6th Intl. Workshop on Spoken Language Technologies for Under-Resourced Languages*, Gurugram, India, Aug. 2018, pp. 52–55.

[24] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi," in *Proc. Interspeech 2017*, 2017, pp. 498–502.

[25] J. Keshet, M. Sonderegger, and T. Knowles, "AutoVOT: A tool for automatic measurement of voice onset time using discriminative structured prediction," 2014, [Computer program].

[26] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," 2022, [Computer program].

[27] S. Barreda, "Fast Track: Fast (nearly) automatic formant-tracking using Praat," *Linguistics Vanguard*, vol. 7, 2021.

[28] Y.-L. Shue, P. Keating, C. Vicenik, and K. Yu, "VoiceSauce: A program for voice analysis," in *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, 2011, pp. 1846–1849.