APRIL 02 2025

Acoustic cues to the perception of plosive voicing in Madurese

Josiane Riverin-Coutlée ; Misnadin ; James Kirby

J. Acoust. Soc. Am. 157, 2365–2375 (2025)
https://doi.org/10.1121/10.0036350

JASA ARTICLE

Acoustic cues to the perception of plosive voicing in Madurese

Josiane Riverin-Coutlée,[1,a] Misnadin,[2,b] and James Kirby[1,c]
[1]Institute for Phonetics and Speech Processing, Ludwig-Maximilians-Universität München, 80799 Munich, Germany
[2]Department of English, Universitas Trunojoyo Madura, Bangkalan, Madura 69162, Indonesia

ABSTRACT:
Madurese, a Malayo-Polynesian language of Indonesia, is described as having a three-way phonation contrast between voiced, voiceless, and aspirated plosives. However, acoustic evidence suggests that the voiceless vs aspirated contrast might be marginal because of small differences in voice onset time (VOT) and large differences in the following vowel height (F1). This raises the question of how these cues are weighted in the perception of the voicing contrast. This paper presents a series of experiments designed to see if Madurese listeners discriminate differences in the positive VOT range, and to what extent they use VOT and F1 to identify plosives. Although listeners were able to discriminate between VOT differences of naturally occurring magnitudes in an AXB task, use of positive VOT when distinguishing voiceless from aspirated plosives in a three alternative forced choices task was highly individually specific, even when F1 was uninformative. Conversely, negative VOT emerged as a more robust cue to the voiced category. These results suggest that the Madurese laryngeal contrast is primarily a two-way contrast signaled through differences in (pre-)voicing but not aspiration. The weak but reliable acoustic covariance between vowel height and aspiration may instead have a diachronic and/or physiological-aerodynamic basis.
 https://doi.org/10.1121/10.0036350
(Received 6 September 2024; revised 5 February 2025; accepted 13 March 2025; published online 2 April 2025)
[Editor: Benjamin V. Tucker] Pages: 2365–2375

I. INTRODUCTION

It is well known that the contrast between two phonetic categories is signaled along multiple acoustic-phonetic dimensions, or cues. For example, Lisker (1986) catalogued no less than 16 acoustic differences between /p/ and /b/ in English, including voice onset time (VOT), co-intrinsic fundamental frequency (F0) perturbations, and the onset frequency of the first formant (F1). Cues to a contrast differ both in terms of their distributional informativeness, as defined on the basis of their acoustic distributions, as well as their perceptual weight. Typically (but not always), listeners give greater perceptual weight to acoustic dimensions which are more informative (Holt and Lotto, 2006; Schertz and Clare, 2020; but cf. Kuang and Cui, 2018). As a result, robust acoustic separation along one of these dimensions in production is often sufficient to distinguish two phonetic categories in perception. Cues that unambiguously determine phonetic category membership in this fashion are often described as "primary," whereas cues characterized by smaller mean differences and more overlapping distributions with larger variance are often described as "secondary" (Schertz *et al.*, 2020; Schertz and Clare, 2020).

The relative weight of cues varies by language (and at least to some extent, individual: see e.g., Clayards, 2018; Kong and Edwards, 2016). For example, a difference in the duration of post-release aspiration (positive VOT/voice lag) is the primary cue differentiating English voiced from voiceless plosives (Lisker and Abramson, 1964), whereas this contrast is primarily signaled via presence vs absence of pre-voicing (negative VOT/voice lead) in French (Serniclaes, 1987). The relative weight of secondary cues can similarly vary. For example, F1 transition is a more important secondary cue for perceiving the voiced-voiceless stop contrast for English-speaking than Spanish-speaking listeners (Schertz *et al.*, 2020), while in Korean, listeners rely on both VOT and F0 for perceiving the three-way stop contrast, but there are regional differences in the relative weight attributed to these two cues (Kang *et al.*, 2022); and so on.

Madurese, a Malayo-Polynesian language of Indonesia, has been described as having a three-way contrast between voiced, voiceless unaspirated, and voiceless aspirated plosives. However, in this language, the acoustic difference in VOT between the aspirated and unaspirated plosives is minimal when compared to those of other languages featuring a similar contrast (Cho and Ladefoged, 1999). In addition, as reviewed in Sec. I A, Madurese has a robust process of vowel harmony controlled by the preceding consonant, which complicates the three-way analysis of voicing. Together, these properties raise the question of which feature(s) underlie(s) the perception of the laryngeal distinction in Madurese CV syllables.

In this paper, we investigate how VOT and F1 are weighted in the perception of Madurese plosives. We find that, while Madurese listeners can discriminate tokens

[a]Email: josiane.riverin@phonetik.uni-muenchen.de
[b]Email: misnadin@trunojoyo.ac.id
[c]Email: jkirby@phonetik.uni-muenchen.de

differing in VOT differences of at least 20 ms, their use of VOT in identification appears restricted to highly ambiguous regions of the F1 space. The fact that Madurese listeners largely fail to attend to positive VOT differences suggests that the Madurese laryngeal contrast is primarily a two-way contrast signaled through differences in (pre-)voicing but not aspiration.

A. Phonetic and phonological properties of Madurese plosives

Madurese (ISO 639-3 mad) is spoken by an estimated 8 million people living on the islands of Madura and East Java, Indonesia (Davies, 2010; Misnadin and Kirby, 2020b; Stevens, 1966). Traditionally, Madurese is described as possessing a three-way contrast between voiced, voiceless unaspirated, and voiceless aspirated plosives at five places of articulation (e.g., Cohn, 1993b; Misnadin and Kirby, 2020b; Stevens, 1968), a distinction that has been encoded in nearly every orthography proposed for the language since at least Kiliaan (1897).[1] We will abbreviate the voiced, voiceless, and aspirated plosives as /D/, /T/ and /TH/, respectively, to indicate the laryngeal categories independent of place of articulation.

There are, however, reasons to question the three-way analysis of Madurese plosives. First, Madurese has a strict phonotactic restriction on CV co-occurrence: although the language distinguishes at least eight phonetic vowel qualities, voiced /D/ and aspirated /TH/ are always followed by a so-called "high" vowel from the set [i ɨ ɤ u], while voiceless /T/ is always followed by a "non-high" counterpart [ɛ ə a ɔ], forming alternating high/non-high pairs [i-ɛ], [ɨ-ə], [ɤ-a], and [u-ɔ][2] (Cohn and Lockwood, 1994; Stevens, 1968) (see Fig. 1). This phonotactic restriction, which is synchronically stable and characteristic of some 95% of the Madurese lexicon (Stevens, 1968), means that /T/ never forms minimal pairs with either /D/ or /TH/ independently of vowel quality, as exemplified in (1).

(1) *dâpa′* [dɤpaʔ] 'arrive'
tapa′ [tapaʔ] 'footstep'
dhâpa′ [tʰɤpaʔ] 'heel'

Three-way laryngeal contrasts among plosives are unusual among genetically related languages of the region: Indonesian, Javanese, and Sundanese all exhibit a two-way contrast (Adisasmito-Smith, 2004; Cohn, 1993b; Davies, 2010; Kulikov, 2010). Moreover, multiple acoustic studies (Cohn and Ham, 1999; Cohn and Lockwood, 1994; Misnadin, 2016; Misnadin and Kirby, 2020a) have shown that the VOT patterns of Madurese diverge from those typically observed in languages described as having aspiration-based contrasts (Cho and Ladefoged, 1999). For example, in English, VOT distributions for unaspirated and aspirated plosives (e.g., /b/ vs /p/) are well separated, with little overlap; similarly, Thai exhibits three distinct distributions for its three-way contrast between voiced, voiceless unaspirated, and voiceless aspirated plosives (Lisker and Abramson, 1964). In Madurese, on the other hand, while /T/ and /TH/ exhibit a stable statistical difference in VOT values (Misnadin and Kirby, 2020a), their distributions overlap substantially, as shown in Fig. 2. Only /D/, which is robustly and consistently prevoiced, can be reliably distinguished from the other plosive types on the basis of VOT alone. Other acoustic parameters which are known to signal the contrast between aspirated and unaspirated plosives in other languages, such as closure duration and onset F0 (e.g., Kang *et al.*, 2022; Schertz *et al.*, 2015), are very similar in Madurese /T/ and /TH/; the primary acoustic dimension that distinguishes them is the F1 (and to a lesser extent, F2) of the following vowel (Misnadin and Kirby, 2020a).

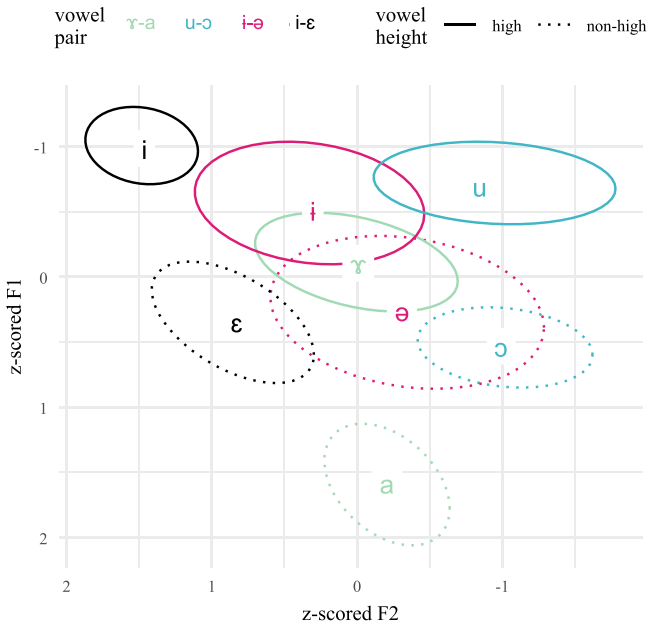

FIG. 1. Vowel space of Madurese, based on carrier phrase recordings made by Misnadin (2016). Formants are z-score normalized. Ellipses indicate one standard deviation from the mean.

Given the robust acoustic differences in vowel height, it is natural to ask whether this contrast is not better analyzed as one of phonological vowel quality with allophonically predictable differences in VOT. However, there is substantial morphophonological evidence that consonant phonation conditions the height of a following vowel, and not vice versa. To take just one example, non-high vowels are also obligatory following nasals, as in *mata* /mata/ "eye," *nyèor* /ɲɛjɔr/ "coconut," *ngolngol* /ŋɔlŋɔl/ "toothless"; when the actor voice morpheme /N/ is prefixed to a stem with a high vowel, such as *bâca* /bɤca/ "to read," the following vowel is realized as its corresponding low counterpart, as in *maca* /maca/ "AV.read." For further examples and arguments, see Cohn (1993a), Davies (2010), Misnadin (2016), and Stevens (1968). For present purposes, it suffices to point out that it is clear that the phonological contrast, such as it is, is controlled by features of the onset.

B. Aims and predictions

Our goal in this paper is to determine the relative cue weights of VOT and vowel height (F1) in the perception of Madurese plosives. In particular, we want to see if listeners

https://doi.org/10.1121/10.0036350

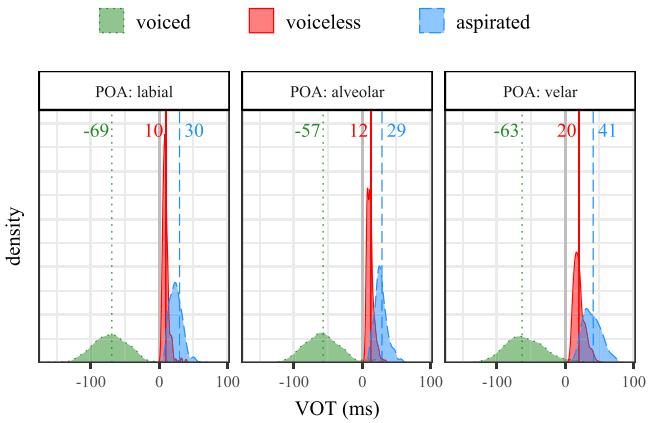

FIG. 2. VOT distributions by plosive type at three places of articulation, based on carrier phrase recordings made by Misnadin (2016). Vertical lines correspond to the mean values printed in the figure.

will make greater use of VOT when F1, the presumptive primary cue, is made uninformative. Given the near-categorical nature of the CV co-occurrence restriction in Madurese, we expect that F1 alone should be sufficient to cue the contrast between /D/ and /T/ or /T/ and /TH/, but that VOT may play a role, especially when vowel quality is ambiguous.

In order to test this, we conducted a series of perception experiments which exposed Madurese listeners to various configurations of F1 and VOT values. The first experiment (Sec. II) was a discrimination task assessing whether Madurese listeners can perceive acoustic differences in VOT of a magnitude similar to that which occurs in natural productions. The second experiment (Sec. III) used a classic two alternative forced choices (2AFC) identification paradigm with co-varying F1 and positive VOT continua to assess the relative weights of VOT and F1. The third experiment (Sec. IV) adds a third response category (prevoiced plosives) while restricting the range of F1, to provide a maximally congruous environment for listeners to focus on VOT differences.

All experiments were conducted using closed-ear headphones in quiet classrooms at the Universitas Trunojoyo Madura in Bangkalan, Madura, Indonesia.

II. EXPERIMENT 1: DISCRIMINATION (AXB) OF POSITIVE VOT

Experiment I is an AXB discrimination task investigating to what extent Madurese listeners can acoustically distinguish between positive VOT differences. We opted for a matching-to-sample rather than same-different (AX) design primarily to avoid participant fatigue, as we anticipated that the majority of the stimulus pairs in an AX task would sound similar; we hoped the AXB task would reduce response bias. An AXB (rather than ABX) design was selected to minimize the potential impact of memory load. Based on earlier literature on languages with phonemic contrasts in the positive range (e.g., English, Thai) (Lisker and Abramson, 1964), a discrimination peak is expected at the VOT boundary where this contrast occurs, which should be at more or less 20 ms at the labial place of articulation (POA) (see Fig. 2).

A. Participants

Ten native speakers of Madurese (5 female, 5 male, ages 18–21) participated in the AXB task. They were students at Universitas Trunojoyo Madura. While they were all raised in Madurese-speaking households and reported using Madurese on a daily basis, including at university, the participants were also all fluent in Standard Indonesian and spoke English to varying degrees.

B. Stimuli

The basis for the stimuli were recordings of words *pateh* "coconut milk" [paʈɛ] and *bhâteh* "profit" [pʰɤʈɛ] (one token each) by a male native speaker of Madurese. The [a]-[ɤ] vowel pair with a labial onset was chosen for Experiments 1 and 2 because exploratory acoustic analyses suggested this pair to differ minimally in terms of resonant frequencies above F1. The F0 of the initial syllables was set to 170 Hz using the PSOLA implementation in Praat 6.0.28 (Boersma and Weenink, 2017). Following the "progressive cutback and replacement" procedure detailed in Winn (2020), an 11-step VOT continuum ranging from 0 to 50 ms in 5-ms increments was created for each syllable. This method creates sound continua in which selected acoustic parameters vary linearly between two endpoints. For VOT-based continua, the aspiration phase of the most aspirated endpoint is taken to progressively replace portions of the vowel in the least aspirated endpoint over a certain number of steps. For this experiment, as VOT increased, vowel duration decreased, such that all stimuli had the same duration. Some examples are given in Fig. 3.

For each VOT value, an 8-step vowel height continuum was then created by manipulating F1 within the range of 525–735 Hz (in 30-Hz increments) (see Flanagan, 1955), which approximate F1 values of naturally produced [ɤ] in [pʰɤʈɛ] and [a] in [paʈɛ], respectively. We used the procedure of Winn (2016a), downsampling vowels to 10 000 Hz and extracting the source wave using a 12 pole LPC filter. This was used to compute endpoint formant contours and to generate 6 intermediate contours via linear interpolation. Formants in the burst and transition were not manipulated.

For the AXB experiment, only a subset of these stimuli were selected to create experimental triads: tokens with VOT values of 0, 10, 20, 30, 40, and 50 ms; and with F1 values of 525, 615, and 735 Hz. Only the initial syllables (duration 240 ms) were used, to encourage acoustic processing and discourage any type of lexical access. Tokens with F1 of 525 and 735 Hz were judged to represent canonical [ɤ] and [a], respectively. Tokens with an F1 value of 615 Hz were chosen because they were judged impressionistically to have the most ambiguous height, and will be represented by [ɜ] hereafter. F1 values within a triad were always the same, whereas the A and B stimuli had VOT values

separated by 20 ms in both possible orders (0–20, 20-0; 10–30, 30-10; etc.). The X stimulus corresponded once to each A and B, totaling 48 different triads.

C. Procedure

The experiment was implemented as an `ExperimentMFC` in Praat (version 6.0.28). Participants listened to three repetitions of the 48 AXB triads, randomized within blocks, for a total of 144 trials. The interstimulus interval was 500 ms. They used the keyboard to signal whether X was more similar to A or B. They could not change their responses or listen to the stimuli more than once. They were given the opportunity to take breaks between blocks. Prior to the main task, a training phase featuring 8 triads allowed participants to familiarize themselves with the experimental protocol. Participants took around 20 min to complete the task.

D. Results

Responses were coded in a binary fashion as correct or incorrect. The results, averaged over all participants, are shown in Fig. 4. Individual response plots can be found in the supplementary material. A majority of correct answers (i.e., above 50%) were given across vowel heights and VOT contrasts, indicating that listeners were able to discriminate stimuli at a better-than-chance level. However, there is no obvious discrimination peak at any point of the VOT contrast continuum for either vowel height.

Mixed-effect binomial logistic regressions were fitted to the data using the lmerTest package (Kuznetsova *et al.*, 2017) in R (R Core Team, 2022). A simple model including only random intercepts for listeners was compared with more complex ones to find the model of best fit using Akaike and Bayesian information criteria. Here, this model did not include any of the fixed factors tested, i.e., vowels (3 levels: [ɤ], [ɜ], [a]) and VOT contrasts (4 levels: 0:20, 10:30, 20:40, 30:50), or their interaction; but only random slopes for VOT contrasts by listeners. This confirms statistically that no VOT contrast was better discriminated than the others (no discrimination peak), for any of the vowels, while listeners performed above chance level (estimate = 1.43; SE = 0.16; $z = 3.24$; $p < 0.001$; see supplementary material for full model summary).

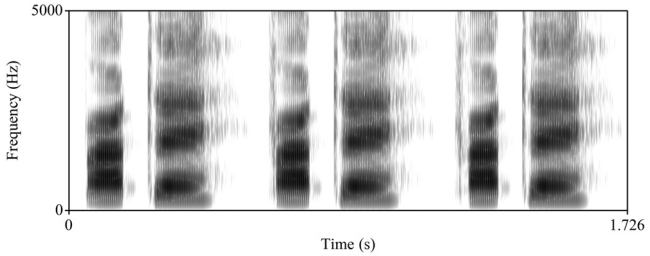

FIG. 3. Examples of three resynthesized stimuli based on [paʈɛ] "coconut milk" with natural vowel [a] and VOTs of 0, 20, and 40 ms.

E. Interim discussion

Participants in Experiment 1 were generally able to discriminate between pairs of syllables differing in 20 ms VOT, with similar performance across vowel heights. However, they did not show clear evidence of a discrimination peak, although discrimination accuracy was numerically greater for the 0:20 ms VOT pair. These results suggest that VOT differences of 20 ms, corresponding roughly to the mean difference between aspirated and unaspirated plosives in production, should be acoustically discriminable by Madurese listeners, at least for relatively short-lag VOTs.

III. EXPERIMENT 2: IDENTIFICATION (2AFC) OF POSITIVE VOT

Experiment 2 is a classic 2AFC task designed to assess the relative perceptual weight accorded to F1 and VOT in lexical identification. We expected that changes to the primary cue dimension (F1) should exert a greater influence on listeners' responses, with a crossover in category identification when the primary cue is ambiguous.

A. Participants

Sixteen Madurese listeners participated in the 2AFC task: ten who had participated in the previous AXB task, plus six additional participants (1 female, 5 male) with similar linguistic and demographic profiles.

B. Stimuli

The stimuli described in Sec. II B were used in the 2AFC experiment. Unlike in the AXB task, here the stimuli were full disyllables, based on *pateh* "coconut milk" [paʈɛ] and *bhâteh* "profit" [pʰɤʈɛ]. Each token had one of 9 VOT

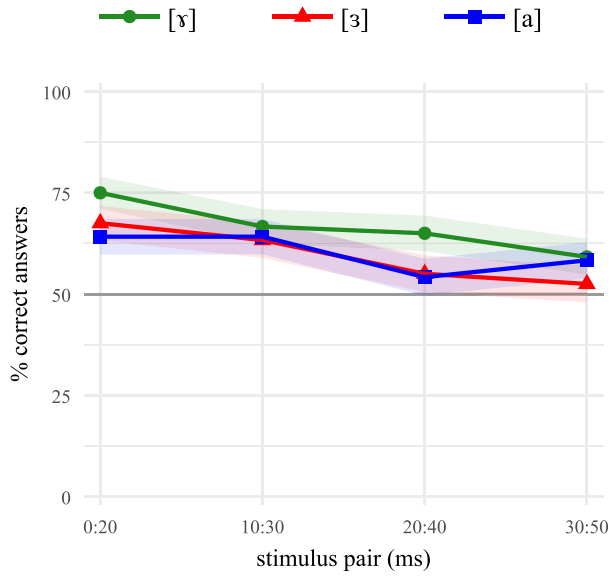

FIG. 4. Average correct responses in the A×B discrimination task as a function of stimulus pair and vowel quality. Shading represents standard error of the mean.

https://doi.org/10.1121/10.0036350

values ranging from 0 to 40 ms and one of 8 F1 values ranging from 525 to 735 Hz, for a total of 72 different stimuli. For further details about the stimuli, see Sec. II B.

C. Procedure

The experiment ran in Praat. Participants heard five repetitions of the 72 stimuli, randomized within blocks. They used the keyboard to identify words as being *pateh* or *bhâteh*, presented in the 2008 Madurese orthography (see Davies, 2010, pp. 51–60). They could listen to the stimuli only once, could not change their responses, and were encouraged to take breaks between blocks. Prior to the task, a short training phase with 10 stimuli familiarized the participants with the procedure.

Due to a scripting error, no responses were recorded for the stimulus with a VOT of 0 ms and F1 of 525 Hz. The rest of the responses were coded in a binary fashion (onset /p/ or onset /p^h/) and analyzed with mixed-effect binomial logistic regressions following the procedure described in Sec. II D. The model of best fit included F1 (continuous) as fixed effect, but not VOT (continuous) or their interaction. Random slopes for F1 by listeners were also included.

D. Results

The results of the 2AFC experiment are plotted in Fig. 5. Individual response plots can be found in the supplementary material. Identification patterns largely depend on F1 values, with lower F1 (similar to [ɤ]) triggering less *pateh* responses than higher F1 (similar to [a]). Responses for stimuli with an F1 value of 615 Hz, impressionistically the most ambiguous vowel height and the one used to represent an [ɜ] quality intermediate between [ɤ] and [a] in Experiment 1, are less decisive. There is no evidence that VOT played any role in lexical identification, with fairly constant identification rates across VOT values at any given F1 value. Accordingly, the only factor significantly affecting the participants' responses in the fitted logistic regressions was F1 (estimate = –1.33; SE = 0.22; $z = -5.94$; $p < 0.001$; see supplementary material for full model summary).

E. Interim discussion

In general, the participants' lexical decisions in the 2AFC experiment were guided by F1, confirming the role of vowel height as the primary perceptual cue. A category crossover occurred when this cue was ambiguous, i.e., at the F1 value of 615 Hz. However, no such crossover occurred at any point of the positive VOT continuum. Indeed, there is no clear indication from this experiment that listeners used VOT as a secondary cue.

Nevertheless, the finding that at the maximally ambiguous F1 value of 615 Hz, listeners were closer to chance level (50%) throughout the VOT continuum warrants further investigation for at least two reasons. First, previous studies on the perception of voicing have found that secondary cues exert the strongest influence when the primary cue is maximally ambiguous (Abramson and Lisker, 1985; Idemaru and Holt, 2011; Summerfield and Haggard, 1977). Capitalizing on this ambiguity may help assess whether VOT has any weight in Madurese listeners' perception of the plosive voicing contrast. Second, while listeners did not seem to use VOT at all for lexical identification in the 2AFC task, they were sensitive to it in the AXB task. One possibility for this is that the mere presence of non-ambiguous vowels in the 2AFC task discouraged participants from attending to VOT.

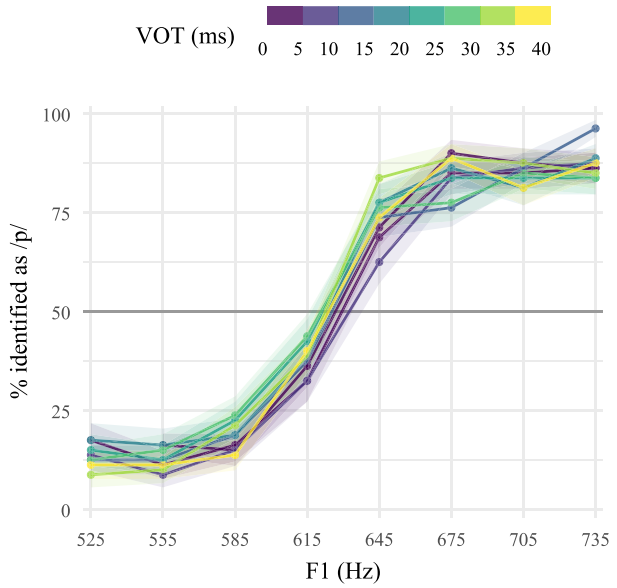

FIG. 5. Percentage of /p/ (*pateh*) responses in the 2AFC task by VOT and F1 values, averaged over participants and repetitions. Shading represents standard error of the mean.

Alternatively, or in addition, the fact that formants were not manipulated in the burst and transition may have resulted in contradictory cues, potentially compromising stimulus naturalness and lexical access. Because the Winn procedure uses a single recording of aspiration to generate all VOT steps, either by chopping the original aspiration single (if shortening is required) or extending and blending it (if lengthening is required), whatever formant structure is inherent during the particular aspiration recording used is present in all continuum steps. This is particularly salient at longer VOTs. As a result, listener responses may have been influenced by conflicting F1 cues. To avoid this confound, we conducted a third experiment in which we ensured the spectral properties of the aspiration phase matched those of the vocalic phase.

IV. EXPERIMENT 3: IDENTIFICATION (3AFC) OF NEGATIVE AND POSITIVE VOT

The third perception experiment is an identification task with three alternative forced choices (AFC), which aims to determine the weight accorded to VOT when vowel height is maximally ambiguous. In an attempt to circumvent the methodological issue related to stimulus naturalness mentioned in Sec. III E and to generalize our findings, we generated a new set of stimuli with three ambiguous vowel

qualities and three places of articulation (POA). We included stimuli with negative VOT values to obtain a more complete picture of the perceptual role of laryngeal cues in the Madurese plosive contrast. If VOT plays no role whatsoever in the perception of this contrast (because vowels following voiced and voiceless unaspirated plosives will always be different, exactly as for voiceless unaspirated and aspirated plosives), we expect to observe a similar lack of crossover between the /D/ and /T/ categories. Finally, to focus attention on the acoustic cues, listeners were presented with CV monosyllables and the task was phoneme identification instead of lexical decision (cf. Sec. III).

A. Participants

Forty-two listeners (37 female, 5 male) who did not participate in either of the previous tasks took part in the 3AFC experiment. They were students at Universitas Trunojoyo Madura, aged 19–22 years old, native and regular speakers of Madurese with high proficiency in Standard Indonesian and varying degrees of knowledge of English.

B. Stimuli

A male native speaker of Madurese who is also a trained phonetician produced 30 phonotactically legal and *illegal CV syllables combining three POAs, three vowel pairs, and three voicing types as shown in Table I.

Using Praat, the F0 contour was first made identical across syllables, starting at 140 Hz at vowel onset and declining to 100 Hz at vowel offset. Subsequent manipulations were made with STRAIGHT (Kawahara *et al.*, 2008) because it can generate highly natural-sounding resynthesized speech, but also because it allows straightforward manipulation of parameters, such as the anticipatory coarticulatory cues contained in the stops' aspiration phase, which may signal the following vowel quality. First, the legal-*illegal CV combinations shown in each row of Table I were morphed so as to create the rightmost column's ambiguous CV (e.g., [bɤ]-*[ba] > [bɜ]).[3] Second, each ambiguous CV triplet shown in the rightmost column of Table I (e.g., [bɜ]-[pɜ]-[pʰɜ]) was morphed one more time to make the vowel quality identical across voicing types. Third, a 14-step VOT continuum was created from each ambiguous CV triplet, totaling 70 different stimuli. Fourth, using the Praat script from Winn (2016b), VOTs were adjusted to precise values ranging from –60 to 70 ms in 10-ms increments, without altering VOT cutback. We focused on this VOT range for three reasons: first, to keep the experiment to a reasonable length; second, because the positive range selected covers most plausible values in the language, including (arguably carefully articulated) read speech (see Fig. 2) (Misnadin and Kirby, 2020a); and third, to include a similar number of values in both the positive and negative values, so as not to introduce a bias toward differences in the positive range.

TABLE I. Legal and *illegal CV syllables morphed into the 15 ambiguous CV syllables used as basis for the VOT continua used in the 3AFC task, per POA, vowel pair, and voicing type.

POA	Pair	Voicing	Legal	*Illegal	Ambiguous
Labial	[ɤ]-[a]	Voiced	[bɤ]	*[ba]	[bɜ]
		Voiceless	[pa]	*[pɤ]	[pɜ]
		Aspirated	[pʰɤ]	*[pʰa]	[pʰɜ]
Alveolar	[ɤ]-[a]	Voiced	[dɤ]	*[da]	[dɜ]
		Voiceless	[ta]	*[tɤ]	[tɜ]
		Aspirated	[tʰɤ]	*[tʰa]	[tʰɜ]
Velar	[ɤ]-[a]	Voiced	[gɤ]	*[ga]	[gɜ]
		Voiceless	[ka]	*[kɤ]	[kɜ]
		Aspirated	[kʰɤ]	*[kʰa]	[kʰɜ]
Labial	[i]-[ɛ]	Voiced	[bi]	*[bɛ]	[be]
		Voiceless	[pɛ]	*[pi]	[pe]
		Aspirated	[pʰi]	*[pʰɛ]	[pʰe]
Labial	[u]-[ɔ]	Voiced	[bu]	*[bɔ]	[bo]
		Voiceless	[pɔ]	*[pu]	[po]
		Aspirated	[pʰu]	*[pʰɔ]	[pʰo]

C. Pretest

Stimuli were validated in a pretest with two groups of listeners whose use of VOT in perception is well documented. Twenty native English-speaking and 20 native French-speaking listeners were recruited via Prolific (www.prolific.co) to complete a 2AFC experiment in which they identified stimulus onsets using the keyboard (e.g., ⟨b⟩ vs ⟨p⟩). Data from one English-speaking participant were removed because they reported experiencing technical issues during the task, as well as from one English-speaking and two French-speaking participants who responded at chance. The results from the remaining 36 listeners, displayed in Fig. 6, showed the expected language-related difference in category crossover, where a majority of French-speaking participants started identifying stimulus onset as voiceless at the VOT value of 0 ms, and English-speaking participants at 10 ms (see supplementary material for plots per CV combination and individual listener). The results of the pretest confirm that the VOT differences in the stimuli are indeed perceptible to listeners of languages who are known to use VOT as a primary cue to voicing.

D. Procedure

The 3AFC experiment was run using PsyToolkit (Stoet, 2010, 2017). Five blocks of stimuli were presented, one per 14-step VOT continuum, with four randomized repetitions of each step within blocks. Participants clicked on symbols shown on the screen that represented syllable onset in standardized Madurese orthography, i.e., ⟨b⟩, ⟨p⟩ and ⟨bh⟩; ⟨d⟩, ⟨t⟩ and ⟨dh⟩; or ⟨g⟩, ⟨k⟩ and ⟨gh⟩, depending on the POA of the stimuli. The participants listened to each stimulus only once, could not change their responses, and could take breaks between blocks. The main task was preceded by a short training phase in which listeners were asked to identify seven phonotactically legal, non-ambiguous CV syllables, none of which was reused during the main task.

https://doi.org/10.1121/10.0036350

Responses were coded as "voiced," "voiceless," or "aspirated," and were analyzed with mixed-effect multinomial logistic regressions via the *mclogit* library (Elff, 2022). The model of best fit included an interaction between fixed effects VOT (continuous) and CV (five levels: labial-[ɜ], alveolar-[ɜ], velar-[ɜ], labial-[e], labial-[o]), and random intercepts by listeners. The introduction of a fifth degree orthogonal polynomial term to model the continuous VOT factor was also found—numerically and visually—to produce a better fit of the empirical data. *Post hoc* pairwise comparisons were computed with *emmeans* (Lenth, 2022).

E. Results

Figure 7 shows the model-predicted probabilities of listeners identifying stimulus onsets as voiced, voiceless, or aspirated depending on VOT value and CV combination, two factors that were found to interact in the model (see supplementary material for details of the empirical responses and model summary).

For stimuli with negative VOTs, the probability predicted by the statistical model is overwhelmingly voiced, reflecting the strong tendency toward voiced responses. A crossover between voiced and voiceless responses is observed at the VOT value of approximately 0 ms for all CV combinations. In the positive range, all combinations exhibit an early peak of voiceless responses and a progressive ramping-up of aspirated responses, but some differences across CV combinations also emerge at the group level. For alveolar-[ɜ] and labial-[e], the probability of voiceless responses decreases over the positive range but remains higher than the probability of the aspirated responses until the penultimate step (60 ms). A similar pattern is observed for velar-[ɜ], but the probabilities of voiceless and aspirated responses converge slightly earlier, with a crossover at the last step. For labial-[o], voiceless and aspirated responses crossover at 30 ms, and for labial-[ɜ], around 40–50 ms.

Individual response plots per CV combination can be found in the supplementary material. These show substantial inter-listener variability for voiceless and aspirated responses. While several listeners' response curves resemble those observed at the group level in Fig. 7, others exhibit crossovers at different points of the positive VOT continuum, or other patterns altogether. Response patterns in the voiced (VOT < 0 ms) region are much more consistent across listeners and CV combinations, but a few participants still exhibit some degree of variability in the negative VOT range (e.g., listener 30).

F. Interim discussion

The presence of any prevoicing triggered a strong probability of stimulus onset to be identified as voiced, and participants practically never gave voiced responses if no prevoicing was present (i.e., if the stimulus did not have a negative VOT). A category crossover occurred at the VOT value of 0 ms, consistent with production data (see Fig. 2). This suggests the presence of a boundary at this point of the VOT continuum, and that prevoicing alone is probably sufficient to cue the voiced vs voiceless contrast even when F1 is ambiguous.

In the positive VOT range, both voiceless and aspirated responses were more probable than voiced ones. Unaspirated responses dominated the lowest positive VOT values, while an increase in aspirated responses was observed as VOT increased. However, while a clear crossover between aspirated and unaspirated responses was not generally observed at the group level, the duration of positive VOT does appear to have played some role in the participants' lexical decisions when vowel height was maximally ambiguous.

For two of the CV combinations involving a labial POA, group-level category crossovers occurred in the positive range (30 and 40–50 ms), but not at the same point of the continuum as in production data (20 ms, see Fig. 2). Compared to other CV combinations, labial-[o] stood out in having lower probability of voiceless responses in the early positive continuum and a particularly early category crossover at 30 ms. Responses to the labial-[ɜ] stimuli also showed some evidence of a crossover, albeit later, around 40–50 ms. We discuss possible reasons for these findings below.

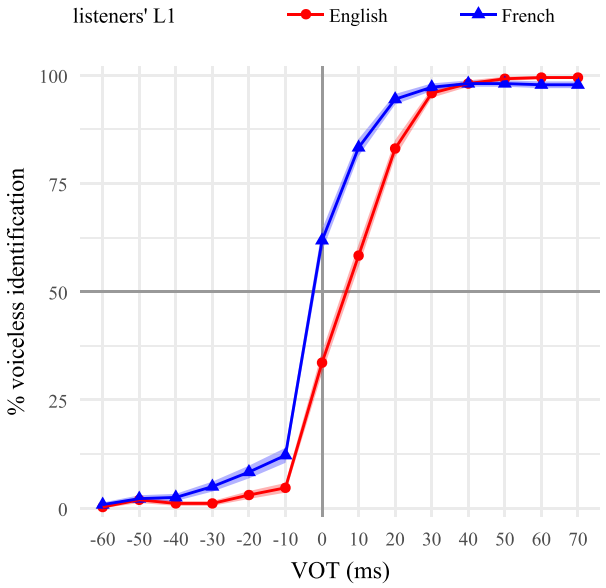

FIG. 6. Percentage of voiceless identification in the pretest as a function of VOT and listeners' L1. Shading represents standard error of the mean.

V. GENERAL DISCUSSION

The results of the AXB discrimination task (Sec. II) suggest that, at least for relatively short-lag VOTs, Madurese listeners were able to perceive VOT differences of 20 ms with (slightly) better than chance accuracy. However, we found no evidence of a discrimination peak normally associated with a category boundary. In a 2AFC lexical decision task where both VOT and F1 were orthogonally covaried (Sec. III), we observed little to no influence

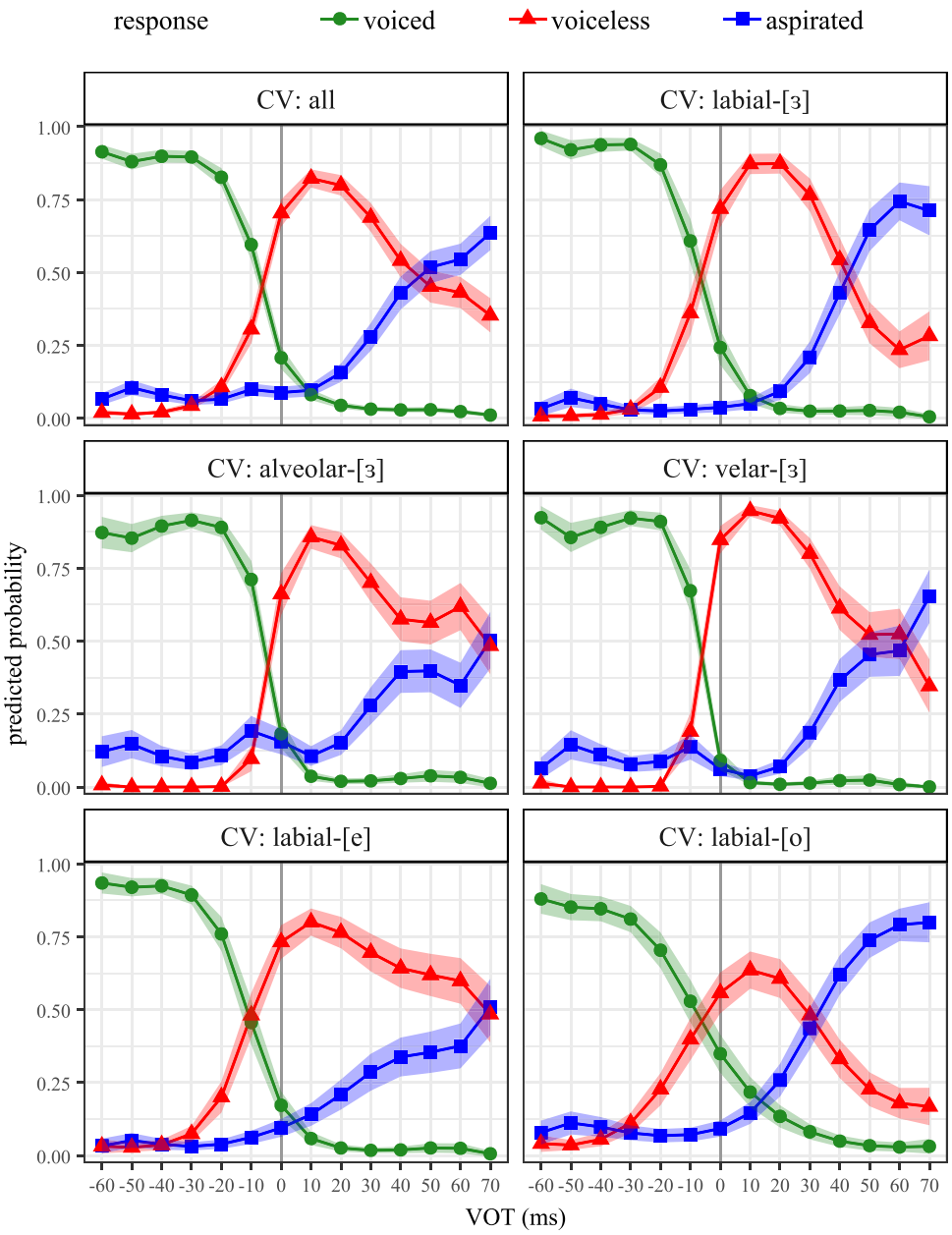

FIG. 7. Model-predicted probability of voiced, voiceless, and aspirated identification in the 3AFC task as a function of VOT for different CV combinations. Shading represents the 95% confidence interval. Top left panel shows results pooled over CV combinations.

of VOT on category judgments. However, the results of a more extensive follow-up 3AFC experiment (Sec. IV), which controlled for several potential experimental confounds, suggest that Madurese listeners can use VOT as a perceptual cue to plosive voicing at least when the primary cue, F1, is highly ambiguous.

That being said, while we were eventually able to construct an experimental scenario in which Madurese listeners were effectively forced to use positive VOT differences to aid them in reaching a categorization decision, this context was highly unnatural. Even when the primary cue was made maximally uninformative, the kind of two-crossover 3AFC function observed for languages such as Thai (Lisker and Abramson, 1970) failed to emerge at the group level: identification rates for the voiceless aspirated plosive were never greater than 75%. For some CV combinations, especially labial-[o] and labial-[ɜ], and to a lesser-extent velar-[ɜ], there is something like an aspirated-unaspirated crossover point, albeit much later in the continuum than would be expected based on the production data. For the case of labial-[o], we suspect this could be an artifact of the morphing procedure having yielded a vowel too close to /u/ to be truly ambiguous, thus skewing responses toward the phonotactically legal option [p^hu]. Some (but not all) individual response patterns shown in the supplementary material tend to support this hypothesis, for example those of listeners 10 and 19, who virtually never used the voiceless response for labial-[o] stimuli while mostly conforming to group patterns for other CV combinations.

Comparison of vowel qualities in the labial-[ɜ] stimuli (where we observe a group-level voiceless/aspirated crossover) with the alveolar-[ɜ] (where we do not) reveals small

https://doi.org/10.1121/10.0036350

differences in steady-state F1 and F2 (bilabial-[ɜ] F1 630 Hz, F2 1315 Hz; alveolar-[ɜ] F1 615 Hz, F2 1375 Hz), so here again, it seems plausible that the responses could be skewed by vowel quality. While this could potentially be controlled for by first determining the by-participant boundary between each vowel pair à la Brunner and Żygis (2011), we note that differences between the vowel qualities of our labial-[ɜ] and alveolar-[ɜ] stimuli are at the discriminability threshold under optimal listening conditions (Hawks, 1994; Kewley-Port and Watson, 1994). Given that the difference limens are rather higher under more ordinary listening conditions (Kewley-Port and Zheng, 1999), we suspect that the contexts in which differences in positive VOT are likely to play a perceptual role are extremely limited.

Our findings are consistent with much previous work demonstrating the high degree of language- and individual-specificity in cue weighting. Especially in the 3AFC task, some listeners seemingly relied more on VOT than others, with two-crossover functions similar to those of Thai listeners observed in their response plots (see, e.g., listeners 27 and 34 in the supplementary material). While it is beyond the scope of this study to account for such individual differences in cue weighting, which have been discussed extensively elsewhere (e.g., Clayards, 2018; Idemaru *et al.*, 2012; Kim *et al.*, 2020; Kong and Edwards, 2016; Schertz *et al.*, 2015; Yu, 2022), it is worth mentioning that even for those listeners who did exhibit two crossovers, the point at which voiceless and aspirated responses crossed was not necessarily consistent across continua (cf. Kong and Edwards, 2016) or with production data, suggesting once more an inconsistent role of VOT. Moreover, inter-individual variability in the 3AFC task was far greater for our Madurese participants than for the French- and English-speaking listeners who rated the same stimuli in the pretest and who are known to rely on VOT as primary cue (with the caveat that the French and English pretests were 2AFC).

As noted in Sec. I, the primary perceptual cue to a contrast can often be predicted based on its informativeness, but the same is not always true of secondary cues. In some languages where the laryngeal contrast is undergoing sound change, such as Southern Yi (Kuang and Cui, 2018), Afrikaans (Coetzee *et al.*, 2018), or Chru (Brunelle *et al.*, 2020), it has been observed that listeners display sensitivity to cues in perception that is greater than expected based on their production patterns. There is no evidence that the Madurese laryngeal contrast is anything but highly stable, but it is perhaps still somewhat surprising that VOT plays such a marginal perceptual role. While the distributional differences in positive VOT are modest in Madurese, especially when compared to those in F1, so too are the differences in F0 when compared to VOT in English, yet changes in onset F0 can substantially alter categorization behavior in that language (Abramson and Lisker, 1985; Haggard *et al.*, 1981; Whalen *et al.*, 1993).

If vowel height is the primary acoustic-perceptual cue distinguishing /T/ onsets from /TH/ onsets in Madurese, what accounts for the persistence of the small but significant differences in VOT (Fig. 2)? One possibility is that this is due to the greater aerodynamic resistance offered by high, close vowels, leading to a delay in the transglottal pressure drop necessary to sustain voicing (Ohala, 1981). As suggested by Berry and Moyle (2011), vocal fold tension—and subsequently phonation onset pressure—could be increased due to contraction of the genioglossus and extrinsic laryngeal muscles during the production of high vowels (Honda, 1983), leading to a slight delay in voicing onset and consequently longer VOTs.

It is also worth noting that, while unaspirated /T/ onsets continue Proto-Malayo-Polynesian voiceless obstruents, the aspirated /TH/ onsets in Madurese correspond to historically voiced onsets, while the present-day voiced /D/ onsets are derived from glide hardening and borrowings (Kiliaan, 1897; Stevens, 1966). In other Austronesian languages, such as Jarai and Raglai, fully or partially devoiced voiced (or low-register) stops are often realized with a short voice lag, which can be nearly indistinguishable from that of the voiceless unaspirated/high-register series (Brunelle *et al.*, 2020; Brunelle *et al.*, 2022; Brunelle *et al.*, 2024). If the "aspirated" plosives of Madurese developed in a similar fashion, this may also go some way toward explaining the minimal differences in voicing lag times.

VI. CONCLUSION

In this paper, we studied the role of VOT and F1 as perceptual cues to the onset plosive voicing contrast in Madurese. The primary acoustic cue signaling the difference between (phonologically) unaspirated + low and aspirated + high CV sequences in Madurese is vowel height (F1), not VOT. Even when F1 is rendered uninformative, the perceptual weight afforded VOT is fairly weak, confirming its status as a secondary cue. The fact that listeners largely fail to use VOT to distinguish these categories even when F1 is uninformative suggests that the Madurese laryngeal contrast is primarily a two-way contrast signaled through differences in (pre-)voicing, but not aspiration. The weak but reliable acoustic covariance between vowel height and aspiration may instead have a basis in physiological constraints and/or contrast enhancement.

SUPPLEMENTARY MATERIAL

See the supplementary material for statistical model summaries, by-speaker response plots, and, for the third experiment, empirical response plots.

ACKNOWLEDGMENTS

Results in Sec. II and Sec. III were previously discussed in a preliminary form in Kirby and Misnadin (2019). This work was supported by grants from the European Research

Council under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 758605) and the UK Arts and Humanities Research Council (Grant No. AH/P014879/1). We thank the editorial team and the reviewers for their feedback.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflict to disclose.

Ethics Approval

Ethics approval was granted by the PPLS Research Ethics Committee of the University of Edinburgh (200–1617/5). Informed consent was obtained from all participants.

DATA AVAILABILITY

The data that support the findings of this study are available from the authors upon reasonable request.

[1]See Davies (2010, pp. 51–60) for a comprehensive overview of Madurese orthographic traditions.

[2]Note that the use of "high" and "non-high" must be interpreted relative to the rest of the Madurese vowel system and, particularly with respect to the vowel [ɤ], does not necessarily imply a phonetic quality of absolute closeness in the IPA sense.

[3]We created a range of ambiguous vowel qualities, which were then submitted for identification to the speaker who produced the original stimuli. Inconsistent identifications over three repetitions of a vowel were considered evidence of an ambiguous quality. The middle step among the ambiguous qualities was then selected for further processing.

Abramson, A. S., and Lisker, L. (**1985**). "Relative power of cues: F0 shift vs. voice timing," in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, edited by V. Fromkin (Academic Press, San Diego, CA), pp. 25–33.

Adisasmito-Smith, N. (**2004**). "Phonetic and phonological influences of Javanese on Indonesian," Ph.D. thesis, Cornell University, Ithaca, NY.

Berry, J., and Moyle, M. (**2011**). "Covariation among vowel height effects on acoustic measures," J. Acoust. Soc. Am. **130**(5), EL365–EL371.

Boersma, P., and Weenink, D. (**2017**). "Praat: Doing phonetics by computer (version 6.0.28) [computer program]," https://www.praat.org.

Brunelle, M., Brown, J., and Hà, P. T. T. (**2022**). "Northern Raglai voicing and its relation to Southern Raglai register: Evidence for early stages of registrogenesis," Phonetica **79**(2), 151–188.

Brunelle, M., Leb, K., Tạ, T. T., and Đinh, G. L. (**2024**). "Voicing or register in Jarai dialects? Implications for the reconstruction of Proto-Chamic and for registrogenesis," J. Int. Phon. Assoc. **54**, 635–676.

Brunelle, M., Tạ, T. T., Kirby, J., and Đinh, G. L. (**2020**). "Transphonologization of voicing in Chru: Studies in production and perception," Lab. Phonol. **11**(11), 15.

Brunner, J., and Żygis, M. (**2011**). "Why do glottal stops and low vowels like each other?" in *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, pp. 376–379.

Cho, T., and Ladefoged, P. (**1999**). "Variation and universals in VOT: Evidence from 17 endangered languages," J. Phon. **27**(2), 207–229.

Clayards, M. (**2018**). "Differences in cue weights for speech perception are correlated for individuals within and across contrasts," J. Acoust. Soc. Am. **144**(3), EL172–EL177.

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., and Wissing, D. (**2018**). "Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis," J. Phon. **66**, 185–216.

Cohn, A. C. (**1993a**). "Consonant-vowel interactions in Madurese: The feature lowered larynx," in *Papers from the 29th Regional Meeting of the Chicago Linguistic Society*, edited by K. Beals (Chicago Linguistic Society, Chicago, IL), pp. 105–119.

Cohn, A. C. (**1993b**). "Voicing and vowel height in Madurese: A preliminary report," in *Oceanic Linguistics Special Publications, No. 24: Tonality in Austronesian Languages*, edited by J. A. Edmondson and K. J. Gregerson (University of Hawaii Press, Honolulu, HI), pp. 107–121.

Cohn, A. C., and Ham, W. H. (**1999**). "Temporal properties of Madurese consonants: A preliminary report," in *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, edited by E. Zeitoun and P. J.-K. Li, Academica Sinica, Taipei, Vol. 1, pp. 227–249.

Cohn, A. C., and Lockwood, K. (**1994**). "A phonetic description of Madurese and its phonological implications," Work. Papers Cornell Phon. Lab. **9**, 67–92.

Davies, W. D. (**2010**). *A Grammar of Madurese* (De Gruyter, Berlin, Germany).

Elff, M. (**2022**). "mclogit: Multinomial logit models, with or without random effects or overdispersion," available at https://CRAN.R-project.org/package=mclogit.

Flanagan, J. L. (**1955**). "A difference limen for vowel formant frequency," J. Acoust. Soc. Am. **27**(3), 613–617.

Haggard, M., Summerfield, Q., and Roberts, M. (**1981**). "Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading F_0 cues in the voiced-voiceless distinction," J. Phon. **9**, 49–62.

Hawks, J. W. (**1994**). "Difference limens for formant patterns of vowel sounds," J. Acoust. Soc. Am. **95**(2), 1074–1084.

Holt, L. L., and Lotto, A. J. (**2006**). "Cue weighting in auditory categorization: Implications for first and second language acquisition," J. Acoust. Soc. Am. **119**(5), 3059–3071.

Honda, K. (**1983**). "Relationship between pitch control and vowel articulation," Technical Report SR-73, Haskins Laboratories, New Haven, CT.

Idemaru, K., and Holt, L. L. (**2011**). "Word recognition reflects dimension-based statistical learning," J. Exp. Psychol. Hum. Percept. Perform. **37**(6), 1939–1956.

Idemaru, K., Holt, L. L., and Seltman, H. (**2012**). "Individual differences in cue weights are stable across time: The case of Japanese stop lengths," J. Acoust. Soc. Am. **132**(6), 3950–3964.

Kang, Y., Schertz, J., and Han, S. (**2022**). "The phonology and phonetics of Korean stop laryngeal contrasts," in *The Cambridge University Press Handbook of Korean Linguistics*, edited by S. Cho and J. Whitman (Cambridge University Press, London, UK), pp. 215–247.

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., and Banno, H. (**2008**). "Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F_0, and aperiodicity estimation," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, NV, pp. 3933–3936.

Kewley-Port, D., and Watson, C. S. (**1994**). "Formant–frequency discrimination for isolated English vowels," J. Acoust. Soc. Am. **95**(1), 485–496.

Kewley-Port, D., and Zheng, Y. (**1999**). "Vowel formant discrimination: Towards more ordinary listening conditions," J. Acoust. Soc. Am. **106**(5), 2945–2958.

Kiliaan, H. N. (**1897**). *Madoereesche Spraakkunst. Eerste Stuk: Inleiding en Klankleer* (Landsdrukkerij, Batavia).

Kim, D., Clayards, M., and Kong, E. J. (**2020**). "Individual differences in perceptual adaptation to unfamiliar phonetic categories," J. Phon. **81**, 100984.

Kirby, J., and Misnadin (**2019**). "Perception of laryngeal contrast in Madurese," in *Proceedings of the 19th International Congress of Phonetic Sciences*, edited by S. Calhoun, P. Escudero, M. Tabain, and P. Warren (Australasian Speech Science and Technology Association Inc., Canberra, Australia), pp. 2378–2382.

Kong, E. J., and Edwards, J. (**2016**). "Individual differences in categorical perception of speech: Cue weighting and executive function," J. Phon. **59**, 40–57.

Kuang, J., and Cui, A. (**2018**). "Relative cue weighting in production and perception of an ongoing sound change in Southern Yi," J. Phon. **71**, 194–214.

Kulikov, V. (**2010**). "Voicing and vowel raising in Sundanese," Stony Brook, New York, available at http://myweb.uiowa.edu/vkulikov/voicing_vowel_ raising_sundanese.pdf.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (**2017**). "lmerTest package: Tests in linear mixed effects models," J. Stat. Softw. **82**, 1–26.
Lenth, R. V. (**2022**). "emmeans: Estimated marginal means, aka least-squares means," available at https://CRAN.R-project.org/package=emmeans.
Lisker, L. (**1986**). " 'Voicing' in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees," Lang. Speech **29**(1), 3–11.
Lisker, L., and Abramson, A. S. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**(3), 384–422.
Lisker, L., and Abramson, A. S. (**1970**). "The voicing dimensions: Some experiments in comparative phonetics," in *Proceedings of the 6th International Congress of Phonetic Sciences* (Academia Publishing House of the Czechoslovak Academy of Sciences, Prague Czech Republic), pp. 563–567.
Misnadin. (**2016**). "The phonetics and phonology of the three-way laryngeal contrast in Madurese," Ph.D. thesis, University of Edinburgh, Edinburgh, UK.
Misnadin, and Kirby, J. (**2020a**). "Acoustic correlates of plosive voicing in Madurese," J. Acoust. Soc. Am. **147**(4), 2779–2790.
Misnadin, and Kirby, J. (**2020b**). "Madurese," J. Int. Phon. Assoc. **50**(1), 109–126.
Ohala, J. J. (**1981**). "Articulatory constraints on the cognitive representation of speech," in *The Cognitive Representation of Speech*, edited by T. Myers, J. Laver, and J. Anderson (North Holland, Amsterdam, Netherlands), pp. 111–122.
R Core Team (**2022**). "R: A language and environment for statistical computing," http://www.R-project.org/.
Schertz, J., Carbonell, K., and Lotto, A. (**2020**). "Language specificity in phonetic cue weighting: Monolingual and bilingual perception of the stop voicing contrast in English and Spanish," Phonetica **77**(3), 186–208.
Schertz, J., Cho, T., Lotto, A., and Warner, N. (**2015**). "Individual differences in phonetic cue use in production and perception of a non-native sound contrast," J. Phon. **52**, 183–204.
Schertz, J., and Clare, E. J. (**2020**). "Phonetic cue weighting in perception and production," WIREs Cogn. Sci. **11**(2), e1521.
Serniclaes, W. (**1987**). "Etude expérimentale de la perception du trait de voisement des occlusives du Français" ("Experimental study of the perception of the occlusive voicing trait of the French")," Ph.D. thesis, Université Libre de Bruxelles, Institut de Phonétique, Brussels, Belgium.
Stevens, A. M. (**1966**). "The Madurese reflexes of Proto-Malayopolynesian," J. Am. Oriental Soc. **86**(2), 147–156.
Stevens, A. M. (**1968**). *Madurese Phonology and Morphology* (American Oriental Society, New Haven, CT).
Stoet, G. (**2010**). "PsyToolkit: A software package for programming psychological experiments using Linux," Behav. Res. Methods **42**(4), 1096–1104.
Stoet, G. (**2017**). "PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments," Teach. Psychol. **44**(1), 24–31.
Summerfield, Q., and Haggard, M. (**1977**). "On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants," J. Acoust. Soc. Am. **62**(2), 435–448.
Whalen, D. H., Lisker, L., Abramson, A. S., and Mody, M. (**1993**). "F0 gives voicing information even with unambiguous voice onset times," J. Acoust. Soc. Am. **93**(4), 2152–2159.
Winn, M. (**2016a**). "Make_formant_continuum," http://www.mattwinn.com/praat/Make_Formant_Continuum_v37.txt, version 37.
Winn, M. (**2016b**). "Make_VOT_continuum," http://www.mattwinn.com/praat/Make_VOT_Continuum_v12.txt, version 12.
Winn, M. (**2020**). "Manipulation of voice onset time in speech stimuli: A tutorial and flexible Praat script," J. Acoust. Soc. Am. **147**(2), 852–866.
Yu, A. C. L. (**2022**). "Perceptual cue weighting is influenced by the listener's gender and subjective evaluations of the speaker: The case of English stop voicing," Front. Psychol. **13**, 840291.