PERCEPTION OF LARYNGEAL CONTRAST IN MADURESE

James Kirby[1], Misnadin[2]
[1]University of Edinburgh, [2]Universitas Trunojoyo Madura
j.kirby@ed.ac.uk, misnadin@trunojoyo.ac.id

ABSTRACT

We investigate native speaker perception of cues to voiceless plosives in the Malayo-Polynesian language Madurese. Madurese is described as having a three-way laryngeal contrast between voiced, voiceless aspirated, and voiceless unaspirated plosives. However, voiceless aspirated and unaspirated plosives are always followed by vowels of different but predictable height, and their VOT distributions overlap heavily, raising the question of whether VOT or F1 is primary perceptual cue to this contrast. The trading relation between VOT and F1 in Madurese was investigated using 2AFC identification and AXB discrimination paradigms. Results indicate that the VOT differences between voiceless plosives which exist in production are not exploited in perception, suggesting that Madurese speakers may not have distinct phonetic targets for aspirated and unaspirated plosives. The surface VOT distributions may instead be a result of differences in following vowel height.

Keywords: *Voicing, laryngeal contrast, speech perception, Austronesian, phonology*

1. INTRODUCTION

Madurese is a Western Malayo-Polynesian language spoken primarily on the island of Madura and parts of East Java, Indonesia [7, 22]. It is unusual among languages of the region for its putative three-way laryngeal contrast between prevoiced (/D/), voiceless unaspirated (/T/), and voiceless aspirated (/Tʰ/) stops. In addition, while there are 8 phonetic vowel qualities in Madurese (see [6, 17] and Figure 1), they are subject to an unusual CV co-occurrence restriction: voiceless unaspirated stops, nasals, and initial liquids are always followed by 'non-high' vowels [a ɔ ɛ ə], while voiced stops, aspirated stops, liquids and glides preceded by a high vowel are always followed by 'high' vowels [ɤ i ɨ u] (Table 1).

Phonetic vowel height is therefore predictable based on the phonation type of a preceding consonant, an analysis supported by morphophonological evidence [5, 7, 22]. For example, prefixes can trigger vowel height alternations in roots: /N/ 'AV' + [patɛ] 'die' → [matɛ] 'AV.die', with accompanying nasal place assimilation, but /N/ + [bɤbɤ] 'low' → [mabɤ] 'AV.low', with the first [ɤ] surfacing as its non-high counterpart. The phonation type of consonants also controls vowel height in suffixes: [pʰuŋkɔs] 'wrap' + /an/ → [pʰuŋkɔsan] 'wrapped thing', but [bɤlis] 'return' + /an/ → [bɤlisɤn] 'returned thing'.

While the /T/- and /Tʰ/-series plosives show distinctive phonological behaviour, it is nevertheless unclear whether they involve separate phonetic targets. In Thai, perhaps the best-studied language with a three-way laryngeal contrast, the three categories show discrete distributions in production [14], are robustly distinguished on the basis of Voice Onset Time (VOT) differences in perception [15], and show discrimination peaks at the language-specific margins of the labelling distribution for each phoneme [1]. In Madurese, the VOT distributions for the two voiceless plosive series are heavily overlapping (Figure 2), with a mean difference between /T/ and /Tʰ/ of only around 20 ms across all places of articulation (Table 2). This raises the

Figure 1: Formant plot of F1/F2 values for Madurese vowels based on 15 speakers (from [17]).

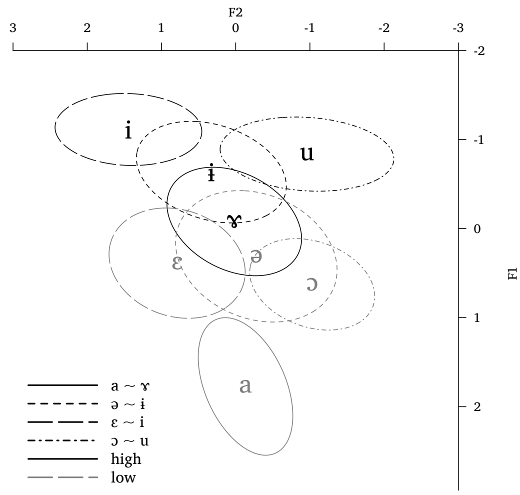


Table 1: CV co-occurrence restriction in Madurese.

Pair	Non-high	High
a - ɤ	paʈɛ 'coconut milk'	pʰɤʈɛ 'profit' bɤtaʔ 'lift up'
ɔ - u	pɔʈɔ 'cake'	pʰuʈʰu 'stupid' budu 'stale fish'
ɛ - i	pɛrak 'happy'	pʰiʈak 'bird' bisa 'able'
ə - ɨ	pəsːɛ 'money'	pʰɨlːis 'furious' bɨrːɤʔ 'heavy'

question of whether Madurese listeners use these small but significant VOT differences to distinguish between voiceless stops in perception, or whether they rely primarily on differences in F1. If listeners are not sensitive to VOT differences, this might cause us to revisit the description of the Madurese laryngeal contrast as a three-category system.

Figure 2: VOT distributions by phonation type, after [16]. Dashed lines indicate mean values, pooled across all places of articulation.

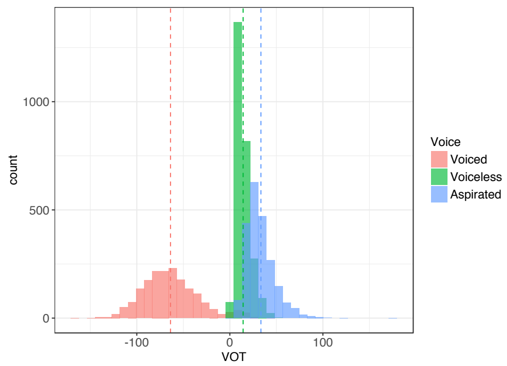


Table 2: Mean (standard deviation) of VOT durations for Madurese plosives, in ms (after [16]).

	Bilabial	Coronal	Velar	Palatal
/b d ɟ g/	-69 (27)	-57 (30)	-63 (33)	-42 (43)
/p t c k/	10 (5)	12 (5)	20 (8)	25 (8)
/pʰ tʰ cʰ kʰ/	30 (14)	29 (11)	41 (18)	52 (16)

We studied the perceptual weighting of F1 and VOT in Madurese with identification and discrimination tasks using resynthesized natural stimuli. If Madurese listeners behave like Thai listeners, we expect to find a perceptual crossover boundary at ambiguous VOT durations, and a discrimination peak at this same boundary.

2. EXPERIMENT 1: IDENTIFICATION (2AFC)

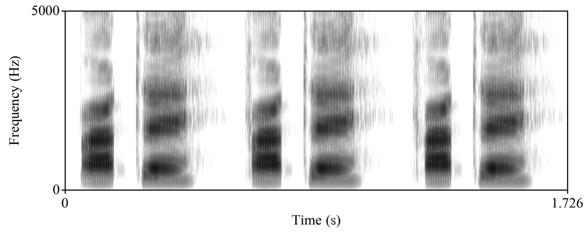
2.1. Participants

16 native speakers of Madurese (6 female) participated in the identification experiment. They were all students (ages 18-21) at Universitas Trunojoyo Madura. All participants were also fluent in Standard Indonesian, and spoke English to varying degrees. However, all were raised in Madurese-speaking households and reported being Madurese-dominant in their daily interactions.

2.1. Stimuli

The stimuli were based on recordings of the items [paʈɛ] 'coconut milk' and [pʰɤʈɛ] 'profit' produced by a male speaker. The F0 of both tokens was modified by ±10 Hz in the first syllable to give an average F0 of 170 Hz and by ±19 Hz in the second syllable to give an average of 189 Hz, using the PSOLA implementation in Praat [4]. The total duration of both items was 550 ms with closure durations of 100 ms. Some examples are shown in Figure 3.

Figure 3: Examples of three identification task stimuli based on [paʈɛ] 'coconut milk' with natural vowel [a] and VOTs of 0, 20, and 40ms.



Based on the acoustic study of [16], Madurese [p] has a mean VOT of 10 ± 5ms, and [pʰ] 30 ± 14 ms. On this basis we created a 9 step VOT continuum from 0 to 40 ms in 5 ms steps following the procedure of [23]. Onset F0 was not manipulated, as for male speakers it was found to be identical in the production of both stops [16]. VOT cutback was not altered, so vowel duration decreased as VOT increased, meaning all stimuli were of constant duration.

For each VOT duration, we then created an 8 step [a]-[ɤ] continuum starting from the naturally produced [a] vowel, following the procedure of [24], a modification of the LPC decomposition and re-synthesis procedure implemented in Praat [4]. Naturally produced tokens of [paʈɛ] (F1/F2/F3: 755/1415/2180 Hz) and [pʰɤʈɛ] (540/1550/2315 Hz) were downsampled to 10000 Hz, and the source wave was extracted using a 12 pole LPC filter. Formant contours were then computed from this inverse filtered source and used to generate 6 intermediate contours via linear interpolation. As the JND for F1 is in the range of 25-35 Hz [8], 8 steps were sufficient to cover the F1 distance between vowels. Formants in the burst and transition were not manipulated.

2.3. Procedure

Instructions and response choices were presented in Madurese orthography (*pateh* for [paʈɛ] and *bhâteh* for [pʰɤʈɛ]). Participants were first presented with 10 stimuli, drawn from the endpoints and ambiguous regions of the continuum, to familiarize them with the experiment. They then responded to 5 repetitions of the 72 stimuli, randomized within blocks, by pressing a keyboard button corresponding to one of the two choices (*pateh* or *bhâteh*). Participants were encouraged to take a short break after each block.

Due to errors in the data collection process, results from 2 participants (1 female) could not be analyzed. The responses from 2 additional participants (1 female) were discarded prior to analysis: one who was at chance for all stimuli, and one who only ever responded *bhâteh*.

2.4. Results

Aggregate results from the remaining 12 respondents are shown in Figure 4. With the possible exception of step 4, where mean F1 was in the ambiguous range (645 Hz), there is no evidence of a trading relation with VOT: identification is based solely on vowel height, and tokens with ambiguous values are judged more or less at chance. Accordingly, in a generalized logistic mixed model (with random intercepts for subject), the estimate for F1 is significant (β = -0.63, p < 0.001) but not the estimates for VOT or the F1:VOT interaction ($\beta \sim 0$ for both predictors).

Figure 4: Identification curves as function of VOT by F1 step value, averaged over participants and repetitions.

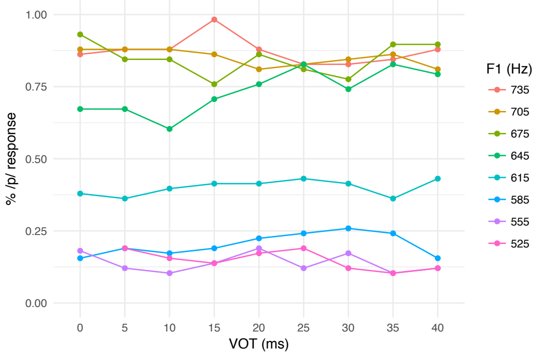


Figure 5: Identification curves as function of VOT by F1 step value by participant, averaged over repetitions.

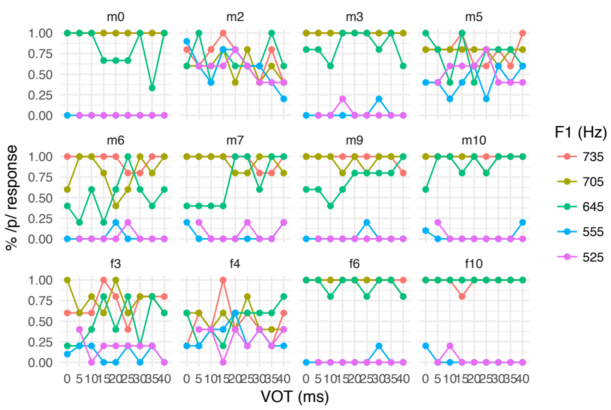


Figure 5 plots the individual response patterns. When F1 was maximally ambiguous (645 Hz), some listeners show evidence for a faint trading relation, but in the unexpected direction. Other listeners appear to be responding categorically based on vowel height.

3. EXPERIMENT 2: DISCRIMINATION (AXB)

Experiment 1 indicated that listeners did not, in general, rely on VOT to make lexical decisions when F1 was ambiguous. One possible explanation for this result is that the VOT steps used were below the threshold of discriminability. If Madurese listeners do use VOT to distinguish between /T/ and /T^{h}/, they should show a discrimination peak at the [p-p^{h}] category boundary (so at ~10-20ms on the VOT continuum).

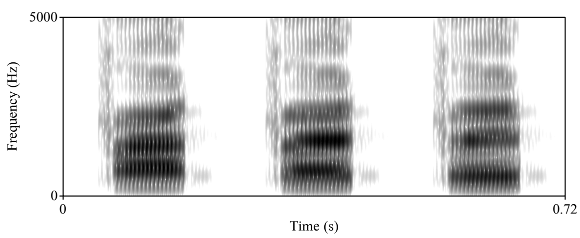
3.1. Participants

10 of the same listeners (5 female) who completed 2AFC task also completed the discrimination task.

3.2. Stimuli

The stimuli were the initial syllables of the identification task stimuli (duration 240 ms). Given the goal of the task, we took care to avoid lexical items. Three vowel qualities were used: a natural [a] (expected to accompany short-lag VOT), a natural [ɤ] (expected to accompany longer-lag VOT) and the resynthesized vowel with ambiguous F1 of 615 Hz (unexpected). An AXB design was utilized to minimize the potential impact of memory load. Vowel qualities within an AXB triad were always the same; they differed only in VOT, with the A and B stimuli separated by 20 ms in both possible orders (so 0-20, 20-0; 10-30, 30-10; etc.), for a total of 48 unique triads. Some examples are given in Figure 6.

Figure 6: Examples of three discrimination task stimuli with VOTs of 20 ms and three vowel qualities, moving from natural [a] to natural [ɤ].



3.3. Procedure

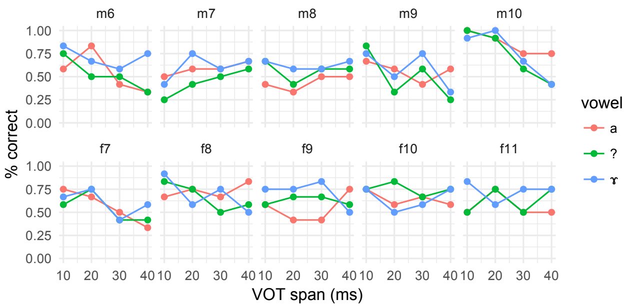
Participants heard three repetitions of the 48 AXB triads, randomized within block. The interstimulus interval was 500 ms. Responses were given via laptop keyboard. Participants were allowed and encouraged to take a break after each block.

Due to errors in the data collection process, results from 2 participants (1 female) were discarded prior to analysis.

3.4. Results

Mean accuracy across all subjects and trials was 62%. Individual results, separated by vowel quality, are given in Figure 7. The x-axis gives the VOT value spanned by the stimulus pair (e.g. the pairs 0-20/20-0 are indicated by 10; 10-30/30-10 by 20; etc.) Although a few participants (m10, f7) show some indication of better discrimination for the 0-20 and 10-30 pairs, a clear discrimination peak is not generally visible. In a linear mixed model with terms for SPAN, VOWEL and their interaction (including subject-specific slopes for SPAN), no terms were significant, although the subject-specific slope for SPAN contributes significantly (χ^2=14.12, p<0.001).

Figure 7: Discrimination as a function of VOT pair by participant, averaged over repetitions.



4. DISCUSSION

In general, Madurese listeners do not appear to attend to differences in VOT when making lexical decisions between stimuli differing both in covarying vowel height and VOT. The AXB results further suggest that Madurese listeners, although able to discriminate between different VOT durations, may not have a category boundary along the VOT continuum.

Before interpreting these findings further, we wish to point out several potential methodological issues. As seen in Figure 5, VOT and F1 may in fact trade (albeit in an unexpected direction) for at least some listeners in a very narrow F1 range. This effect might be enhanced by employing a more acoustically intermediate quality. It is also possible that the inclusion of unambiguous F1 values discourages listeners from attending to VOT at all, so it may be necessary to repeat with an identification continuum that includes *only* the ambiguous F1 range. With respect to the discrimination task, the duration of stimuli (240 ms, with 500 ms ISI) may have played a role; extending the vowel length may give different results, as may using stimuli with different onset consonants, some of which have longer inherent VOTs than the bilabial plosive used here.

Assuming the results are robust, however, it seems likely that a difference in vowel height, rather than VOT, is the primary acoustic-perceptual cue distinguishing words minimally differing in /Tʰ/ vs. /T/-series plosives. What, then, accounts for the small but significant differences in voicing lag time? Cross-linguistically, a correlation between vowel height and VOT is not uncommon [12, 18, 20]. 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 [19]. As suggested by [3], the mechanical relationship between vowel articulation and intrinsic F0 proposed by [9] might be extended to explain these effects: if contraction of the genioglossus and extrinsic laryngeal muscles increases vocal fold tension (and thereby phonation threshold pressure), this could in turn delay voicing onset, leading to longer VOTs before higher vowels.

Alternatively, speakers might be actively increasing the lag before high vowels to make the onset of the following vowel breathy, thereby increasing spectral tilt and enhancing the low frequency concentration of energy brought about by high vowels' low F1 [11]. The fact that many aspirated onsets in Madurese were probably voiced historically [10, 21] may also go some way towards explaining the minimal differences in voicing lag.

While it is clear that the voiced and aspirated plosives pattern together phonologically, their phonetic realizations are rather different. [16] failed to find any consistent phonetic properties shared by voiced and aspirated plosives that might underpin a phonetically grounded feature such as [ATR] or [lowered larynx]. Given the present findings suggesting that listeners do not use VOT to distinguish between two types of plosive in perception, the Madurese laryngeal contrast might actually be more accurately described as a 'two-way' system, like that of Malay or Sundanese [13], distinguishing voiced from voiceless plosives. In other words, despite their divergent phonological behavior, Madurese speakers may not have distinct phonetic production targets for aspirated and unaspirated plosives.

5. ACKNOWLEDGMENTS

This work was supported in part by the Arts and Humanities Research Council [AH/P014879/1]. Thanks to Alexander Martin for asking a question which resulted in us discovering a small but critical error in our analysis code, and to audiences at the Edinburgh P-Workshop and Glasgow University Phonetics Laboratory.

6. REFERENCES

[1] Abramson, A. S., Lisker, Leigh. 1970. Discriminability along the voicing continuum: cross-language tests. *Proceedings of the 6th International Congress of Phonetic Sciences*, Prague, 569–573.

[2] Anderson, S. 1991. Vowel-consonant interaction in Madurese. In: Ann, J., Yoshimura, K. (Eds.), *Proceedings of Arizona Phonology Conference 4* (Vol. 4). Tuscon, AZ: Department of Linguistics, University of Arizona, 1–15.

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

[4] Boersma, P., Weenink, D. 2015. *Praat: doing phonetics by computer*.

[5] Cohn, A. C. 1993. Consonant-vowel interactions in Madurese: the feature lowered larynx. In: Beals, K. (Ed.), *Papers from the 29th Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 105–119.

[6] Cohn, A. C., Lockwood, K. 1994. A phonetic description of Madurese and its phonological implications. *Working Papers of the Cornell Phonetics Laboratory* 9, 67–92.

[7] Davies, W. D. 2010. *A grammar of Madurese*. Berlin: De Gruyter.

[8] Flanagan, J. L. 1955. A difference limen for vowel formant frequency. *The Journal of the Acoustical Society of America* 27(3), 613–617.

[9] Honda, K. 1983. *Relationship between pitch control and vowel articulation* (No. SR-73). New Haven, CT.

[10] Kiliaan, H. N. 1897. *Madoereesche spraakkunst. Eerste stuk: inleiding en klankleer*. Batavia: Landsdrukkerij.

[11] Kingston, J., Diehl, R. L. 1994. Phonetic knowledge. *Language* 70(3), 419–454.

[12] Klatt, D. H. 1975. Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech, Language, and Hearing Research* 18(4), 686–706.

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

[14] Lisker, L., Abramson, A. S. 1964. A cross-language study of voicing in initial stops: acoustical measurements. *Word* 20(3), 384–422.

[15] Lisker, L., Abramson, A. S. 1970. The voicing dimensions: some experiments in comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences*, Prague, 563–567.

[16] Misnadin. 2016. *The phonetics and phonology of the three-way laryngeal contrast in Madurese* (PhD dissertation). The University of Edinburgh.

[17] Misnadin, Kirby, J. 2018. Madurese. *Journal of the International Phonetic Association* 1–18.

[18] Nearey, T. M., Rochet, B. L. 1994. Effects of place of articulation and vowel context on VOT production and perception for French and English stops. *Journal of the International Phonetic Association* 24(01), 1–18.

[19] Ohala, J. J. 1981. Articulatory constraints on the cognitive representation of speech. In: Myers, T., Laver, J., Anderson, J. (Eds.), *The cognitive representation of speech*. Amsterdam: North Holland, 111–122.

[20] Ohala, M., Ohala, J. J. 1992. Phonetic universals and Hindi segment duration. *International Conference on Spoken Language Processing*, Banff, 831–834.

[21] Stevens, A. M. 1966. The Madurese reflexes of Proto-Malayopolynesian. *Journal of the American Oriental Society* 86(2), 147–156.

[22] Stevens, A. M. 1968. *Madurese phonology and morphology*. New Haven: American Oriental Society.

[23] Winn, M. 2016. *Make_VOT_Continuum*. Retrieved from http://www.mattwinn.com/praat/Make_VOT_Continuum_v12.txt

[24] Winn, M. 2016. *Make_Formant_Continuum*. Retrieved from http://www.mattwinn.com/praat/Make_Formant_Continuum_v37.txthttp://www.mattwinn.com/praat/Make_Formant_Continuum_v37.txt