A perception study on cue weighting in Madurese stops

Josiane Riverin-Coutlée[1], Misnadin[2], and James Kirby[1]

[1]Institute for Phonetics and Speech Processing, Ludwig-Maximilians-Universität München (Germany),
[2]Universitas Trunojoyo Madura (Indonesia)

Madurese is a Malayo-Polynesian language spoken by 8 million people on the islands of Madura and East Java, Indonesia [1]. Early studies describe Madurese as maintaining a three-way contrast between voiced, voiceless, and aspirated stops in syllable onsets at five places of articulation [2,3], abbreviated here as /D/, /T/, and /TH/. However, this account is challenged by several findings. First, Madurese maintains a strict phonotactic restriction on CV co-occurrence: /D/ and /TH/ precede "high" vowels /i ɨ u ɤ/, whereas /T/ exclusively precedes their non-high counterparts /ɛ ə ɔ a/, such that /T/ never forms minimal pairs with /D/ or /TH/ independently of vowel quality [1]. Second, acoustic studies [1,4] have found that while /D/ is realized with robust prevoicing, there is considerable overlap in voice onset time (VOT) distributions for /T/ and /TH/ (see Fig. 1), atypical of languages where these are contrastive, e.g. Thai [5]. Third, results of a pilot study using a /p/ ~ /p^{h}/ continuum suggested that Madurese listeners attended only to F1, and not VOT, when making categorization judgments [6].

Together, these findings suggest that only differences in VOT lead times, but not lag times, function as a perceptual cue to the laryngeal contrast in Madurese. This would follow from an "associative" perspective on cue weighting, under which listeners' cue weights in perception can be predicted as a function of their acoustic distributions in production (see review in [7]). First, however, we need to establish whether there is *any* distributional configuration under which Madurese listeners will shift attention to the generally less informative VOT cue. In [6], the experimental stimuli contained F1 values which were clear exemplars of /T/ or /TH/-category words. This may have led listeners to ignore the secondary cue (VOT), despite its correlation with the primary cue (F1). Since secondary cues to voicing exert the strongest influence on voicing perception when the primary cue is ambiguous [8,9], in this study, we neutralize the informativeness of the primary cue, i.e. F1. If Madurese listeners do use (positive) VOT to distinguish /T/ and /TH/, this configuration should provide the most favorable conditions for listeners to shift their attention to the distributionally less reliable VOT cue. Moreover, if their cue weight for (negative) VOT is predicted by its acoustic distribution, we expect them to use it robustly to distinguish /D/ from /T/ and /TH/ when F1 is ambiguous.

Recordings were made of a phonetically-trained native speaker of Madurese producing CV syllables corresponding to phonotactically legal (e.g. /dɤ/, /ta/, /t^{h}ɤ/) and illegal (e.g. */da/, */tɤ/, */t^{h}a/) combinations at three places of articulation (POA: labial, alveolar, velar) followed by the vowels /ɤ/ and /a/. Using Praat [10], f0 was made identical in all syllables. Then, STRAIGHT [11] was used to create a vowel quality ambiguous between /ɤ/ and /a/ and a 14-step VOT continuum ranging from -60 ms to +70 ms in 10 ms increments at each POA. These stimuli were administered to 42 native speakers of Madurese through a 3AFC listening experiment in which they identified the syllable onset presented in Madurese standardized orthography, e.g. <d> for /d/, <t> for /t/ or <dh> for /t^{h}/. Multinomial mixed-effect regressions were fit to the data using the *mclogit* package [12] in R [13]. The model of best fit included an interaction between the fixed factors POA (3 levels) and VOT (continuous), and random intercepts per listener. Estimated marginal means and post-hoc pairwise comparisons with Bonferroni adjustment were computed with *emmeans* [14].

Model estimates are shown in Fig. 2. The probability of /D/ responses compared to /T/ or /TH/ is very high in the negative VOT range, suggesting little confusion in the presence of any prevoicing. In the positive VOT range, an increase in probability of /T/ and /TH/ responses is observed; however, /TH/ remains a less probable response than /T/ over the positive range, except at the highest VOT values with the labial POA (leftmost panel). 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 physiological basis [15].

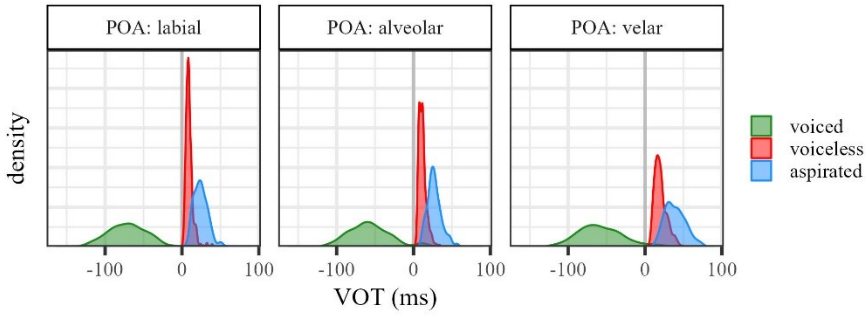

Fig. 1. VOT distributions by stop type and POA, based on data from [4].

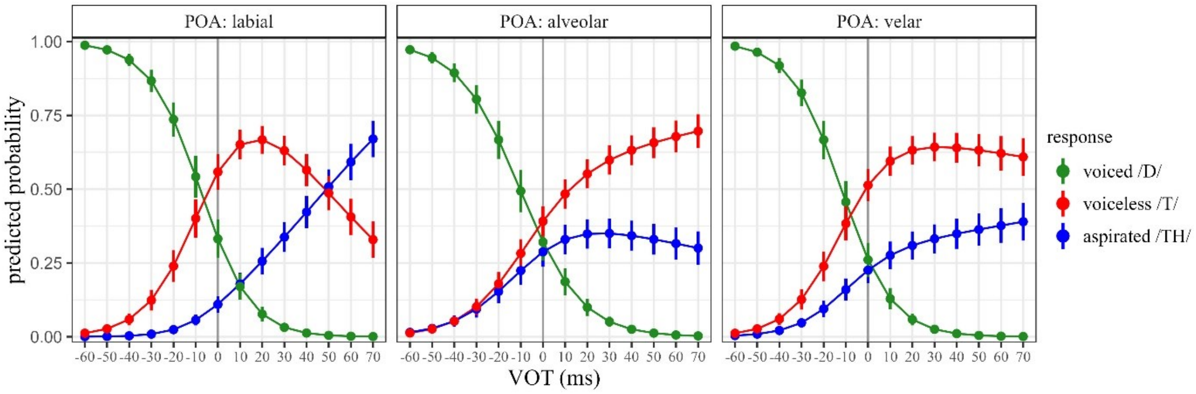

Fig. 2. Predicted probability of voiced, voiceless and aspirated responses at 14 VOT steps and 3 POAs.

References

[1] Misnadin & J. Kirby (2020a). Madurese. *Journal of the International Phonetic Association*, 50(1), 109–126.

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

[3] Cohn, A. C. (1993). Voicing and vowel height in Madurese: A preliminary report. *Oceanic Linguistics Special Publications*, (24), 107–121.

[4] Misnadin & J. Kirby (2020b). Acoustic correlates of plosive voicing in Madurese. *Journal of the Acoustical Society of America*, 147(4), 2779–2790.

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

[6] Kirby, J. & Misnadin (2019). Perception of laryngeal contrast in Madurese. In *Proceedings of the XIXth International Congress of Phonetic Sciences*, 2378–2382. Melbourne, Australia.

[7] Schertz, J. & E. J. Clare (2020). Phonetic cue weighting in perception and production. *WIREs Cognitive Science*, 11(2), e1521.

[8] Summerfield, Q. & M. Haggard (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, 62(2), 435–448.

[9] Abramson, A. & L. Lisker (1985). Relative power of cues: F0 shift vs. voice timing. In V. Fromkin (ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 25–33). San Diego: Academic Press.

[10] Boersma, P. & D. Weenink (2023). Praat: Doing phonetics by computer. http://www.praat.org/.

[11] Kawahara, H., M. Morise, T. Takahashi, R. Nisimura, T. Irino & H. Banno (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In *ICASSP 2008*, 3933–3936. Las Vegas, USA.

[12] Elff, M. (2022). mclogit: Multinomial logit models, with or without random effects or overdispersion. https://CRAN.R-project.org/package=mclogit.

[13] R Core Team (2023). R: A language and environment for statistical computing. https://www.R-project.org.

[14] Lenth, R. (2023). emmeans: Estimated marginal means, aka least-squares means. https://CRAN.R-project.org/package=emmeans.

[15] Berry, J. & M. Moyle (2011). Covariation among vowel height effects on acoustic measures. *Journal of the Acoustical Society of America*, 130(5), EL365–EL371.