

Cross-linguistic variability in cue weighting of consonant voicing

In a recent study of American English, Schultz et al. (2012) examined the relative weighting of two cues to consonant voicing, onset F0 and voice onset time (VOT), in both production and perception. Using discriminant function analysis, they found a significant negative correlation between onset F0 and VOT weights in production, but not in perception. Schultz et al. interpreted their findings as consistent with a model on which speakers have the goal of efficiently generating an integrated acoustic property, as suggested by e.g. Kingston et al. (2008): covariance between weights computed across categories is what might be expected if speakers compensate for producing extreme values of one cue by producing less extreme values of another.

While their primary interest was in the correlation, or lack thereof, between cue weights in production and perception, the production data from Schultz et al. are interesting in their own right, as they raise the question of whether this type of correlation is specific to English, or a more general property of the voicing contrast cross-linguistically. That is, do speakers *generally* seek to produce an “efficient” composite voicing cue structured in terms of (possibly idiosyncratic) correlations between more basic properties? Furthermore, does such a correlation reflect the acoustic realization of the voicing contrast in a given language?

As a first pass at answering these questions, this study examines the covariance in cue weights to onset voicing in six languages: American English, French, Italian, Khmer, Thai, and Vietnamese. This is an interesting set of languages to consider as it contains languages with both two-way (English, French, Italian) and three-way (Khmer, Thai, Vietnamese) onset voicing contrasts; different canonical acoustic realizations of voicing (prevoiced vs. unaspirated and unaspirated vs. aspirated); and two tone languages (Thai and Vietnamese), in which the use of onset F0 as a cue to voicing may be attenuated (Francis et al., 2006).

Our data consist of stressed syllable onsets beginning with bilabial stops (/b p/) in English, French and Italian and coronal stops (/d t t^h/) in Khmer, Vietnamese and Thai. The following vowel was almost always /a/. Production targets were lexical items (or occasionally, phonotactically licit nonce words) in the relevant language. In the tone languages, items were recorded in a range of tonal contexts. For each language, raw F0 measures made at 10ms after voicing onset were converted to semitones relative to each talker’s mean F0 on the log scale. Both this semitone measure and VOT were then scaled to facilitate comparison of the relative linear discriminant weights.

The relative weights of the two cues in production is shown for each language in Figure 1. For English, two conditions are shown: *citation* (“Say pad again”) and *natural* (“What’s the difference between a soft pad and a knee pad?”). For Khmer, Thai, and Vietnamese, two contexts are shown: *isolation* and *carrier*, the latter with the target syllable (/daa/, /taa/ or /t^haa/) in a prosodically focused position (X) of an alternative question frame, e.g. “Did you say X or Y?”. For French and Italian, the data were similarly drawn from an alternative question frame, but here context contrasts high (X) with low (Y) focus position. For Khmer, both /t/ and /t^h/ were labeled as voiceless, while for Thai and Vietnamese, /t/ was grouped with /d/ based on other data on the patterning of onset F0 effects in these languages (see Kirby, LabPhon 2016).

Given the small numbers of speakers involved (6-14 per language) and the diversity of the materials, direct comparison is challenging, but some interesting observations can be made. First, at least in English, the degree of “cue trading” observed may be at least partially a function of task and/or focus; the present findings replicate those of Schultz et al. in the citation, $r(6)=-3.42$, $p=0.01$, but not the natural condition, $r(6)=-0.97$, $p=0.37$. Second, no correlation in weights is apparent in French or Italian, languages in which /b/ is typically characterized by extensive negative VOT (i.e., prevoicing), although caution must be exercised given the extremely small number of data points. This suggests that while the temporal extent of positive VOT may be controlled by speakers, the temporal extent of negative VOT may not be. Third, while the lexical use of F0 appears to have some impact on this pattern, it is influenced by tonal context. Although there is not enough data to permit a

by-tone analysis of the Vietnamese data, in Thai, a strong negative correlation between VOT and onset F0 is seen on isolated syllables bearing a high-falling tone, $r(10)=-5.39$, $p<0.001$, precisely the environment in which robust onset F0 effects may be observed for this language (Kirby, LabPhon 2016). In Vietnamese, a weak negative correlation is observed in the carrier phrase context across all tones, $r(12)=-2.02$, $p=0.047$, while in Thai the correlation is not significant in this context.

Taken together, these results suggest that the degree of talker-specific correlation between onset F0 and timing-related cues to voicing appears strongest in languages where voicing is cued at least in part by long-lag VOT. In this presentation, I will discuss the possible functional and biomechanical explanations for this patterning, as why – if less F0 perturbation is naturally explained by longer time until voicing onset – we do not find VOT/onset F0 covariance on a within-category, token-by-token basis.

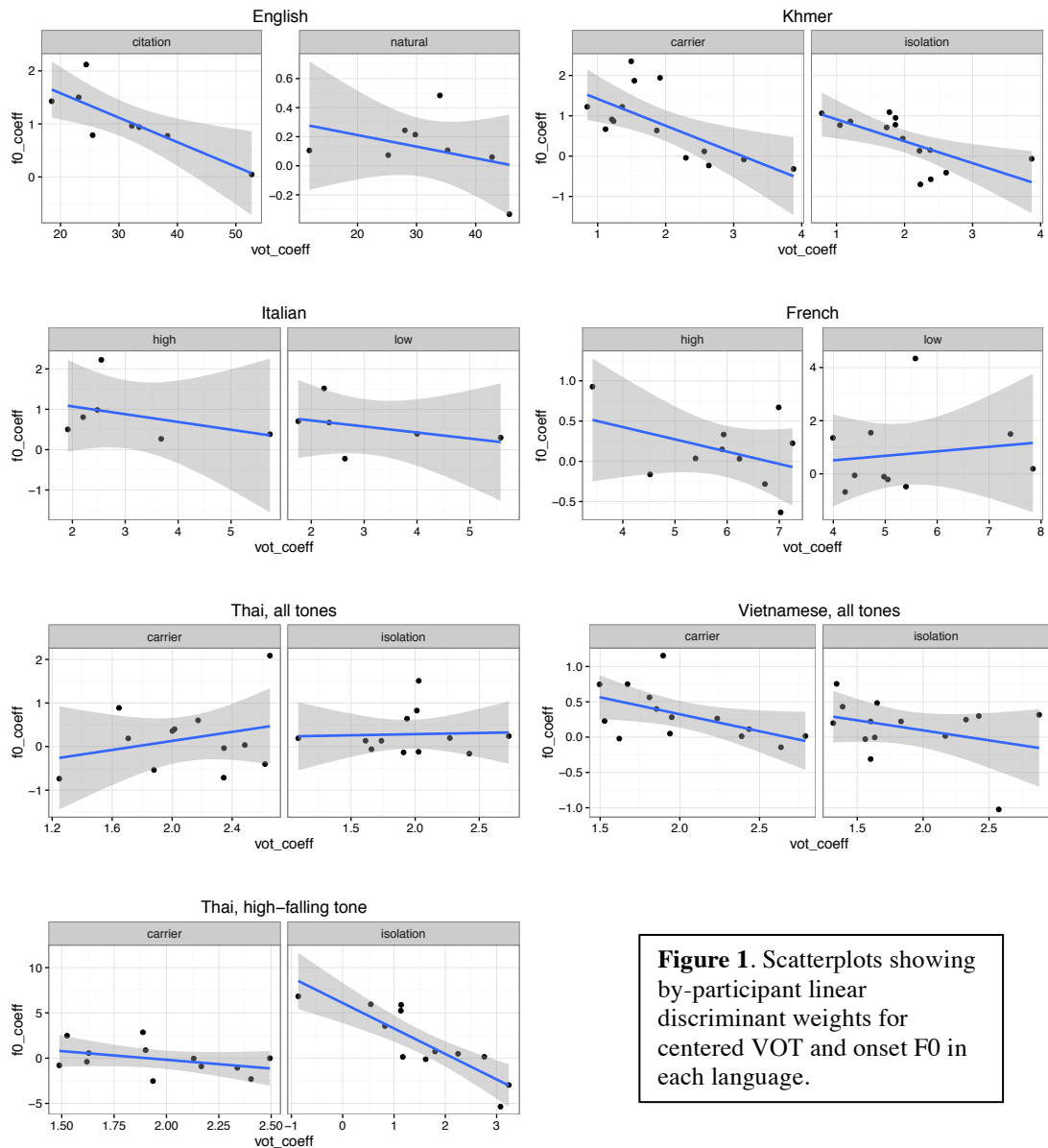


Figure 1. Scatterplots showing by-participant linear discriminant weights for centered VOT and onset F0 in each language.