Spectral balance and phantom pitch in intonational pitch perception

Jonathan Barnes, Alejna Brugos, Nanette Veilleux, Stefanie Shattuck-Hufnagel

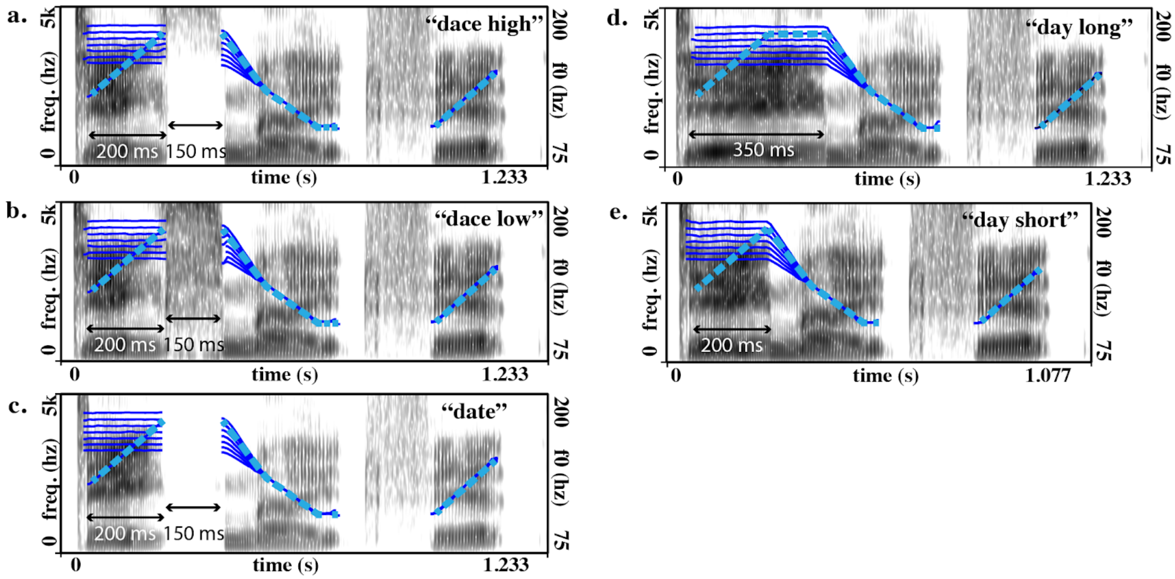
Recent tone and intonation research increasingly appreciates the contribution of non-F0 cues to tonal contrasts [e.g., 5, 6, 11]. Alongside well-known cues (duration, intensity, voice quality), spectral balance during noisy portions of voiceless segments has been seen to covary with F0 in neighboring voiced segments, and may be available to listeners in perception [9, 10]. The nature of this covariation, or "segmental intonation" [8], however, remains mysterious. We report here on a perception study testing the hypothesis, advanced thus far impressionistically, that segmental intonation allows a form of perceptual completion: Turbulent noise may "fill in" F0 discontinuities for listeners with "phantom" or interpolated pitch. Our results however fail to support this hypothesis, suggesting instead that segmental intonation provides an independent cue for tonal contrast, not integrated auditorily with pitch itself [3].

We focus here on perceived tone scaling for English intonational pitch accents. If segmental intonation indeed involves integration of sibilant noise and F0 into a single, higher-level pitch percept, then given a High accent, a coda fricative in the target syllable with higher spectral center-of-gravity should yield higher perceived scaling than an analogous lower-CoG fricative. To test this, we resynthesized two versions of the sentence *DACE might fit*, one with higher CoG during the [s], the other with lower (both plausible renditions of English /s/). 24 listeners were presented with pairs of these utterances, always segmentally identical, but differing in F0, such that one bore a natural rise-fall pitch accent, while the other was taken from a seven-step low-to-high continuum of level-f0 "standards" (see Fig. 1). Listeners then indicated which of the two utterances reached a higher pitch level on the target word.

Previous research [1] using this paradigm shows that listeners judge a long open syllable (*day*), with an F0 rise followed by a high plateau, as higher relative to its level-standard continuum than a shorter *day,* with an identical F0 rise, but no plateau. (Plateaux, all else equal, sound higher than analogous sharp peaks [4]). Additionally, a syllable durationally identical to the longer *day,* but ending in a voiceless stop (*date*), with the same rise as above, but replacing the high plateau of *day-long* with the stop closure, is perceived as identical in scaling to the shorter version of *day.* Listeners, in other words, seem not to "fill in" the pitch contour across the F0 gap (Fig. 1c, d, e). The present study replicates these findings, and asks how our two versions of *dace,* higher-balance and lower, relate to the foregoing.

Results (Fig. 2) show that *day-long,* with its high F0 plateau, is perceived as significantly higher than all targets lacking an overt plateau. This is confirmed ($p < .001$) in a mixed-level logistic regression. Neither *dace* stimulus, therefore, appears to fill in its voiceless interval with anything comparable in terms of perceived pitch to the voiced F0 of *day-long*'s plateau, arguing against perceptual completion as a mechanism for segmental intonation. Fig. 2 also hints, however, at a tendency for higher-balance *dace* to scale slightly higher than other plateau-less target items. This proves marginally significant ($p = .046$) compared with *dace-low*, significant against *date* ($p = .014$), but not significant against *day-short*. If *dace-high* separates at all, in other words, it does so unreliably, and to an extent much too small to equate with the scaling differences between tone contours that segmental intonation is purported to replace. Instead, to the extent that segmental intonation is consistently produced and perceptually robust, we suspect it arises from a production connection between raised fricative CoG and articulatory effort [12], linking it perceptually, therefore, with prominence [7], and perhaps thus modulating perceived pitch, however slightly, through an inferential channel akin to [2]'s Effort Code. Fricative spectra therefore provide a cue to tonal contrast that is auditorily unrelated to F0, much as [3]'s C/V duration ratio and Low Frequency Property serve as cues for the feature [voice] in various languages, without themselves integrating to form a single, higher-level percept.

Figure 1: *Spectrograms with overlaid f0 contours for target items, with identical carrier phrase ("X might fit"): (a) "dace" with higher freq. [s]; (b) "dace" with lower freq. [s]; (c) "date", with silent interval; (d) "day" with long vowel; (e) "day" with short vowel. Target contours (dashed lines) show an identical f0 rise (125 hz to 180 hz) across a 200 ms vowel (a, b, c and e), or the same rise followed by a 150-ms plateau at 180 hz over a 350 ms vowel (d). Seven level standards (solid lines) range from level 7 (highest .5 st above max. f0 of target) to lowest in .5-st steps. (Level 6 is equal to the max f0 of targets.)*



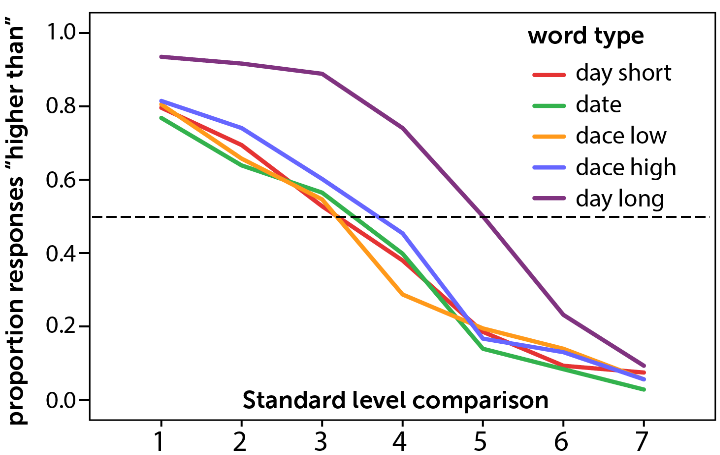


Figure 2. Proportion of responses in which target syllables were judged higher than selected word-specific level-standard, as a function of standard level (1-7), and word type.

References: **[1]** Barnes, J., Brugos, A., Veilleux, N. & Shattuck-Hufnagel, S. 2011. Voiceless Intervals and Perceptual Completion in F0 Contours. *ICPhS 17.* **[2]** Gussenhoven, C. 2004. *The phonology of tone and intonation.* CUP. **[3]** Kingston, J. & Diehl, R. 1994. Phonetic Knowledge. *Language* 70(3). **[4]** Knight, R. 2007. The shape of nuclear falls and their effect on the perception of pitch and prominence: Peaks vs. Plateaux. *Lang. & Speech* **51**(3). **[5]** Kochanski, G., Grabe, E., Coleman, C. & Rosner, B. 2005. Loudness predicts prominence: Fundamental frequency lends little. *JASA 118*(2). **[6]** Kuang, J. 2017. Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *JASA 142*. **[7]** Mixdorff, H., Hönemann, A., Niebuhr, O. & Draxler, C. 2014. Perceived prominence reflected by imitations of words with and without F0 continuity. *Speech Prosody 2014.* **[8]** Niebuhr, O. 2009. Intonation segments and segmental intonations. *Interspeech 2009*. **[9]** Niebuhr, O. 2012. At the edge of intonation – The interplay of utterance-final F0 movements and voiceless fricative sounds. *Phonetica* 69. **[10]** Niebuhr, O. 2017. On the perception of "segmental intonation": F0 context effects on sibilant identification in German. *EURASIP J. Audio, Speech, & Music Proc. 19*. **[11]** Patel, A., Xu, Y. & Wang, B. 2010. The role of F0 variation in the intelligibility of Mandarin sentences, *Speech Prosody 2010*. **[12]** Shadle, C. & Mair, S. 1996. Quantifying spectral characteristics of fricatives. *ICSLP'96.*