

Modelling the perception of English F0 scaling in a segmental context

Jonathan Barnes¹, Alejna Brugos¹, Nanette Veilleux², and Stefanie Shattuck-Hufnagel³

¹Boston University, ²Simmons College, ³MIT

Segmental context in tone and intonation systems is known to influence target F0 contours in both production and perception. Many languages, for example, prefer to realize critical F0 events during maximally sonorous intervals, a tendency manifest variously through altered pitch movement timing ([3], [15], [14], [11]), and in distributional limitations on certain contour types ([6]). Current analytic practice, by contrast, routinely ignores segmental backdrop when considering the perceptual efficacy of putative pitch cues such as F0 turning points: Tone scaling is typically equated simply with measured turning-point height, while tonal timing is read directly off turning points, whose salience is estimated visually, rather than auditorily. Even voiceless intervals may receive no special treatment, owing to widespread belief in perceptual completion during missing F0 intervals ([13], [7]), a process modelled explicitly in many stylization algorithms (e.g., MOMEL, [8]). [1], however, cast doubt on this, demonstrating that listeners' tone-scaling judgments are better understood by ignoring F0 gaps, relying instead solely on recorded F0 values. This study goes further, arguing that the perceptual effects of a range of lower-sonority segments require a scaling model based not on single points, but instead on the integration of weighted F0 samples over a broader region ([4]).

In a 2AFC task, listeners judged the relative scaling of two L+H* accents, realized during versions of the English words *day*, *Dane*, *Dave*, and *Dade*, resynthesized with identical rhyme durations, nucleus:coda ratios (where relevant), and F0 rises (150 Hz at rhyme onset to 200 at offset). A carrier phrase, *X might fit*, bore a rise-fall-rise contour, with nuclear accent on X. Seven additional versions of each stimulus phrase were also created, these with high, plateau-shaped accents, beginning at 200 Hz, and descending by .5 semitones/step, to create a continuum of level reference standards. Sharp-peaked test stimuli were compared only with reference standards of identical segmental make-up (Figures 1a-d). Since plateau-shaped accents are known to sound higher than sharp-peaked analogues with identical maximum F0 ([5], [10]), sharp-peaked stimuli should sound somewhat lower than their highest corresponding reference levels, with 50% crossover points within the continuum providing estimates of perceived F0 for each sharp-peaked accent.

Identical perceived scaling for test stimuli regardless of segmental make-up would predict identical response patterns for all syllable types. This was not the case for 35 listeners (Fig 2), analyzed using mixed-effects logistic regression. Instead, subjects heard *day* as highest, followed by *Dane*, then *Dave*, with *Dade* slightly (though non-significantly) lower. No single point in the test stimuli predicts this result, nor is House's Spectral Stability Hypothesis ([9], [12], [14]) helpful in this instance. We account for these facts instead using an extension of the Tonal Center-of-Gravity ([2]) approach, modelling perceived scaling via a weighted average over a broad window of F0 samples, with lesser sample weights for F0 realized during lower-sonority segments, reflecting decreased perceptual salience (as per [6]). These results demonstrate both the dangers of F0 analysis abstracted away from segmental context, and the attractiveness of a global approach to tonal implementation, such as TCoG.

Figure 1 a-d: F0 contours superimposed on spectrograms for level standards (solid lines) and test items (dashed lines) for *day* (a), *Dane* (b), *Dave* (c), and *Dade* (d).

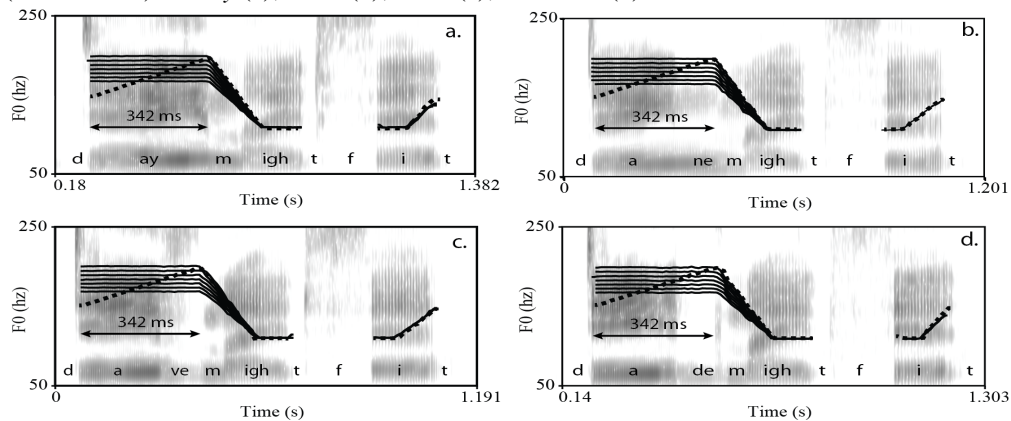
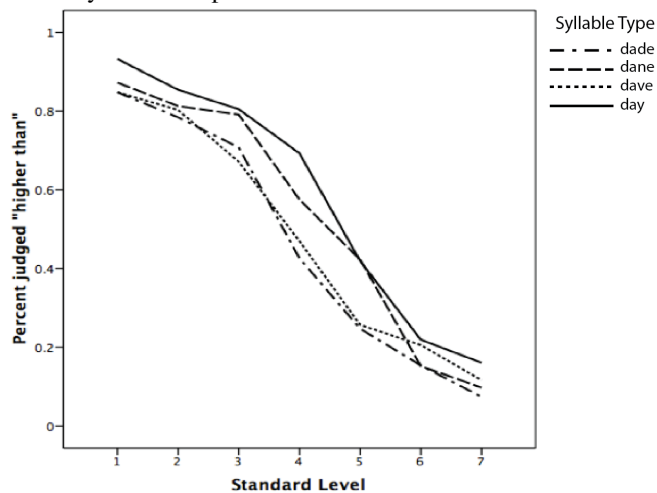


Figure 2: Percent "Higher-than" judgments for 4 syllable types as a function of the level standard against which they were compared.



References

- [1] Barnes, Brugos, Shattuck-Hufnagel, & Veilleux. (forthcoming) "Voiceless intervals..." ICPHS 2011.
- [2] Barnes, Veilleux, Brugos, & Shattuck-Hufnagel. (2010) "The effect of global..." Sp. Pros., 2010.
- [3] Caspers & van Heuven. (1993) "Effects of time pressure ..." *Phonetica* 50: 161-171.
- [4] d'Alessandro, Rosset, & Rossi. (1998) "The pitch of... glissandos." *JASA* 104, 4: 2339-2348.
- [5] D'Imperio (2000) "The Role of Perception in Defining Tonal Targets and their Alignment." Thesis.
- [6] Gordon (2001) "A typology of contour tone restrictions." *Studies in Language* 25: 405-444.
- [7] Hermes. (2006) "Stylization of Pitch Contours." In *Methods in Emp. Pros. Research*, de Gruyter.
- [8] Hirst & Espesser. (1993) "Automatic modelling of fundamental..." *Trav. de l'Institut de Phon. d'Aix* 15.
- [9] House. (1990) *Tonal perception in speech*. Lund University Press.
- [10] Knight. (2008) "The Shape of Nuclear Falls...: Peaks vs. Plateaux" *Lang. & Speech* 51, 3: 223-244.
- [11] Ladd, Mennen & Schepman. (2000) "Phonological conditioning..." *JASA* 107: 2685-2696.
- [12] Mertens. (2004) "The Prosogram: Semi-Automatic Transcription..." *Proc. of Sp. Pros. 2004*.
- [13] Nootboom. (1997) "The prosody of speech: melody and rhythm." In *Hdbk of Phon. Sci.*, Blackwell.
- [14] Prieto. (2009) "Tonal alignment patterns in Catalan nuclear falls." *Lingua* 119, no. 6: 865-880.
- [15] van Santen & Hirschberg (1994) "Segmental effects on timing and height ..." *ICSLP 94*.