VOICELESS INTERVALS AND PERCEPTUAL COMPLETION IN F$_0$ CONTOURS: EVIDENCE FROM SCALING PERCEPTION IN AMERICAN ENGLISH

Jonathan Barnes[a], *Alejna Brugos*[a], *Nanette Veilleux*[b] *& Stefanie Shattuck-Hufnagel*[c]

[a]Boston University, USA; [b]Simmons College, USA; [c]MIT, USA
jabarnes@bu.edu; abrugos@bu.edu; veilleux@simmons.edu; sshuf@mit.edu

ABSTRACT

Intonation models describing F0 alignment and scaling in terms of peak and valley localization can face challenges when F0 contours are interrupted (e.g., during voiceless segments). It is often assumed that some form of perceptual completion or "filling in" of such intervals occurs that resolves these issues. This study uses the perceived scaling of High pitch accents both with and without missing peaks due to F0 gaps to adjudicate between three possible accounts of how speakers treat missing F0 in intonation perception. Results provide strong evidence against both extrapolation and interpolation across the missing region, supporting instead the hypothesis that listeners simply ignore these regions. This suggests that a non-turning-point-based model, such as TCoG, should be considered as an alternative to standard target-and-interpolation models.

Keywords: F0 alignment, intonation, interpolation, extrapolation, F0 plateau

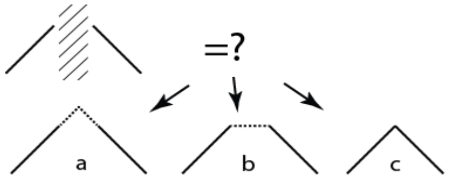
1. INTRODUCTION

It is often observed in the literature on intonation that, although F0 contours are routinely shot through with discontinuities (e.g., at voiceless intervals), our experience of the signal is one of a continuous melody carried over the length of the utterance (e.g., [10, 11, 17]). This might suggest that gaps in the F0 record are sufficiently unobtrusive, in duration, positioning, or both, to avoid disruption of perceived continuity of the F0 signal. A stronger interpretation, apparently informing mainstream assumptions in speech intonation research, is that the "missing" F0 in voiceless intervals is actually restored perceptually. Nooteboom [17], for example, uses a term from the perceptual completion literature, "filling in", to characterize this process (p. 644). Furthermore, nearly all approaches to F0 contour modelling (with notable exceptions, e.g., [1, 13]) employ some form of F0 interpolation through gaps in the F0 record. In some cases (e.g., [8, 19]), this move may be purely pragmatic. In other cases, though, perceptual issues are clearly at stake. The MOMEL algorithm [12], for example, using quadratic spline fitting to produce a continuous F0 curve, which is then reduced to "a series of target points" that can serve as an "appropriate phonetic representation" of the contour in question. Crucially, these target points will often fall within the missing F0 intervals.

If perceptual "filling in" of missing F0 is real, this is good news for target-and-interpolation approaches to tonal phonetics/phonology, such as the Autosegmental-Metrical model [16, 18]. In this model, F0 turning points (such as those provided by MOMEL, hereafter TPs) are typically seen as critical cues for phonological tone specifications. The absence of key TPs should thus cause serious problems for tonal perception. Perceptual completion, on the other hand, predicts such problems should not arise.

To see how perceptual completion might be accomplished in tone perception, take, for example, a symmetrical F0 rise and fall separated by a voiceless interval. One solution would be for listeners to extrapolate the missing F0 peak based on observed trajectories to either side of the gap. (Fig. 1a.) Dannenbring [6], however, provides evidence against this option: when presented with tone glides of this shape, separated by noise, listeners failed to extrapolate a peak between rise and fall, reporting instead a peak F0 equal to or somewhat lower than the real F0 maximum.

Figure 1: Schematic showing 3 predictions for the perceptual contribution of a no-F0 region to the scaling of a high accentual contour: a) extrapolation, b) interpolation, and c) "ignoring an absence".



Bregman [4] and Ciocca and Bregman [5] conclude on this basis that listeners do not extrapolate missing peaks or valleys, but instead integrate disjoint pitch movements by interpolating directly between them. (See Fig. 1b.) It is unclear what such interpolation would mean for the localization of F0 TPs that might otherwise have landed within the critical interval.

To these two approaches, extrapolation and interpolation, we add a third: In a manner somewhat akin to what Dennett [7] describes as “ignoring an absence”, listeners may not actually “fill in” missing F0 at all, but rather simply skip over or ignore it for purposes of judgments about F0 events. (Fig. 1c—our reasons for representing it thus should become clear forthwith.) While this option too presents problems for TP-based accounts of tone perception, a globally-oriented model, e.g., Tonal Center of Gravity (TCoG) [2], could handle this relatively straightforwardly. Listeners might judge tonal alignment and scaling using only samples taken of “real” F0; voiceless regions would contribute no information to this process, but would not impede it either.

As far as we know, these questions have never been investigated explicitly in the domain of speech. We therefore set out to accomplish this, using a design similar to Dannenbring’s in spirit, but substituting speech sounds for tone glides.

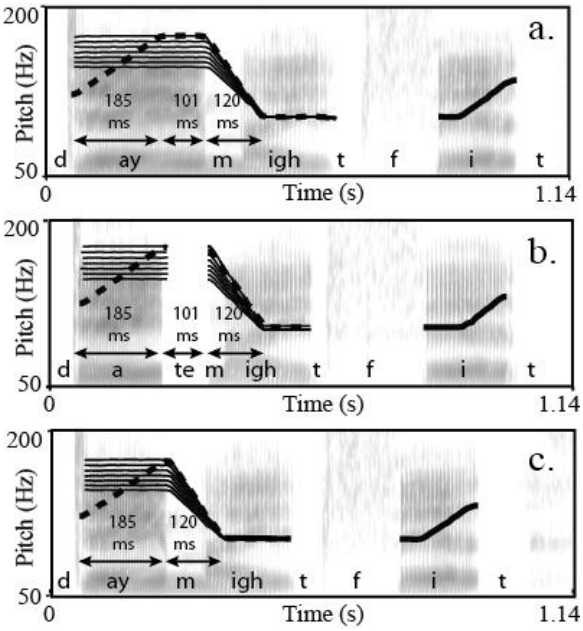
2. METHODS

Our design capitalizes on the well-known phenomenon whereby High pitch accents realized as extended “plateau” sound higher than analogous sharp-peaked accents with identical maximum F0 [9, 14, 15]. Given this, the approaches to missing F0 just described make different predictions about the perception of tone scaling in contours containing F0 gaps. Consider a stylized L+H* pitch accent realized on the first word of the phrase *‘Date’ might fit* (observed, e.g., regarding a crossword). Resynthesized as a linear rise through the accented vowel in *date*, with silence during the closure for coda [t], and then a symmetrical fall beginning in the onset [m] of *might*, this F0 contour is directly parallel to Dannenbring's, described above. Thus, if listeners do extrapolate missing peaks and valleys, the accent in this phrase should sound higher to listeners than an analogous accent on the first word of a similar phrase *‘Day’ might fit,* assuming a rhyme for *day* equal in duration to the rhyme of *date*, but where the silence of [t] is replaced by a high-F0 plateau. (I.e. *Date* would be perceived with filling in, as in Fig. 1a, with *day* corresponding to 1b. Fig. 2. depicts this directly.)

If, by contrast, listeners do not extrapolate missing peaks, but rather interpolate linearly through voiceless intervals, then our synthetic accent in *‘Date’ might fit* should sound equally high to listeners as the plateau-shaped accent on *day*. Likewise, that same *date,* with its interpolated plateau, should sound higher than a similar accent realized on a version of *day* with the rhyme shortened to be only as long as the accented vowel of *date*, and where a linear rise throughout the vowel was followed directly by a symmetrical F0 fall, creating a sharp peak at the word boundary. (i.e. the shape depicted above schematically as 1c, and concretely in Fig. 2c below.) Lastly, if listeners neither extrapolate nor interpolate, but instead just ignore the voiceless interval for purposes of scaling judgments, then the *date* stimulus should sound lower than the longer *day* stimulus, and possibly equal in pitch to the shorter, sharp-peaked *day.*

To test these predictions, we created a set of synthetic stimuli corresponding to the *date, day-long, and day-short* scenarios just described. The frame sentence (*X might fit*) was realized with a rise-fall-rise intonation contour (ToBI H* L-H%). Test items are depicted in Fig. 2.

Figure 2: F0 contours superimposed on spectrograms for standards (solid lines) and test items (dashed lines) for day-long (a), date (b) and day-short (c).



2.1. Stimulus creation

Target phrases were produced by a male native English speaker, and then resynthesized using Praat (2). Segment durations, given in Fig. 2, were based on mean values over multiple utterances from the same speaker. F0 rises were identical in duration (185 ms) and scaling (125-180 Hz) for all stimulus types. This was followed either by a 101 ms high plateau (*day-long*), 101 ms of silence (*date*), or a 120 ms fall (*day-short*).

2.2. Experimental task

What interests us here is the perceived scaling of the nuclear H* pitch accents in our three stimulus types. Direct pairwise comparison of target stimuli differing in segmental composition, however, is fraught with potential confounds, necessitating innovation of an alternative method for the evaluation of relative scaling. On this new approach, each target item was compared to a continuum of reference contours (or standards), that were segmentally identical to the target item in question, but where F0 on the accented syllable held steady at one of 7 levels, the highest at 180 Hz, and descending in .5 semitone increments. After the accented syllable, F0 was identical for target items and standards. (Again, see Fig. 2.)

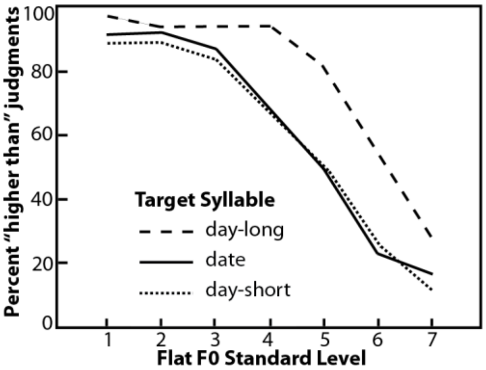
Assuming that corresponding level standards sound identical in scaling regardless of syllable type, then if one target item (e.g., *date*) sounds systematically lower or higher to listeners than another target item (e.g., *day-long*), this difference should be manifest in subjects' perception of the relative scaling of target items and their respective level standards. (e.g., *date* might sound equal in pitch to its standard level 5, while *day-long* might reach only level 4.)

The task itself was 2AFC: 39 native speakers of American English were presented with pairs of contours, either a target item and a level standard, or 2 standards of the same type, and were asked to decide which item's target word reached a higher pitch. After 6 consecutive correct responses in an initial block comparing standards separated by ≥ 3 steps, the experiment began. There, each test item was compared to its 7 standards (3 reps x 2 orders, 126 trials). 90 additional trials pairing level standards separated by 3 continuum steps or less (15 comparisons x 3 target types x 2 orders) served as a measure of participants' accuracy in discriminating pitch levels. Trials were presented in random order.

3. RESULTS AND DISCUSSION

Data from 26 participants is included in the analysis. (One participant did not continue past the initial section, and 12 did not meet criteria for inclusion based on discrimination of level standards). Fig. 3 displays results, pooled across subjects. Lines represent the percentage of trials in which a given target type (i.e. *date, day-long,* or *day-short*) was judged higher than each of its 7 level standards. Each line starts high and declines as the standard level increases, meaning that, as expected, listeners tended to judge target syllables as higher than the lowest standards, but lower than the highest standards. Comparing now target types, the percentage of "higher-than" judgments for *day-short* declines earlier than does that of *day-long*. We infer from this that *day* realized with a high plateau does indeed sound higher to listeners than *day* with a sharp peak, but identical maximum F0. But what of the missing F0 in *date*? "Higher-than" judgments for *date* resemble those for *day-short,* declining earlier than for *day-long.* This is confirmed by a logistic regression analysis, using both standard level and target-syllable type to predict responses. The full model, tested against an intercept-only model, was statistically significant, chi-square (3) = 1256.782, p < .001. Both standard level (Wald χ^2 (1) = 733.005, $p < .001$) and syllable type (Wald χ^2 (2) = 133.628, $p < .001$) yield statistically significant main effects. Furthermore, while the response pattern for *day-long* differs significantly from that of *date* (Wald χ^2 (1) = 112.362, $p < .001$), those of *date* and *day-short* do not differ significantly (Wald χ^2 (1) = .804, $p = .37$). We infer, therefore, that *date* sounded lower to our subjects than *day-long*, but was perceived as equal in scaling to the sharp-peaked *day-short*.

Figure 3: Percent "Higher-than" judgments for three syllable types as a function of the level standard against which they were compared.



Recall now our initial predictions: If listeners “fill in” F0 gaps by extrapolating missing peaks, then *date* should sound higher than *day-long.* It did not. Likewise, if listeners instead fill in using linear interpolation, then *date* should sound equal in scaling to *day-long,* and higher than *day-short*. Again, it did not. Instead, *date,* with its missing F0 “plateau”, sounded the same as *day-short,* with a sharp peak such as would result from removing the voiceless interval in *date* from the signal entirely. For F0 scaling perception, then, it appears as though listeners do not “fill in” missing F0 at all. Instead, they seem to disregard the voiceless interval entirely, perceiving scaling just as if it had never been present to begin with.

4. CONCLUSION

We have shown that our experience of illusory continuity in intonation contours is not achieved by “filling in” missing portions of the F0 contour. Instead, it resembles the kind of “ignoring an absence” posited by Dennett [7] (rightly or wrongly) in connection with perceptual completion in the visual and other domains. This treatment of the missing region must furthermore be highly task-specific: To the extent that the duration of the voiceless interval is used as a cue for various other aspects of the signal (e.g., phonemic identity, syllabification, prosodic constituency etc.), gaps must be ignored for purposes of F0 judgments alone.

If correct, this conclusion raises interesting questions for target-and-interpolation models of tonal implementation: if interruptions in the F0 contour routinely obscure what would otherwise be the locations of the F0 TPs that hypothetically cue tonal targets, and missing TPs are not restored perceptually, then it is unclear how a TP-based model of tonal perception should proceed. If critical perceptual cues to tonal identity are either missing altogether, or radically misplaced, why is the analytic ambiguity researchers face when presented with gap-filled F0 contours not matched by a corresponding difficulty on the part of listeners? To us, this suggests the virtues of a non-TP based approach to tonal implementation, such as Prosogram (5), or TCoG (1).

5. ACKNOWLEDGMENTS

We gratefully acknowledge the support of NSF grants 1023853, 1023954, and 1023596.

6. REFERENCES

[1] d’Alessandro, C., Mertens, P. 1995. Automatic pitch contour stylization using a model of tonal perception. *Comput. Speech Lang*. 9, 257-288.

[2] Barnes, J, Veilleux, N., Brugos, A., Shattuck-Hufnagel, S. 2010. The effect of global F0 contour shape on the perception of tonal timing contrasts in American English intonation. *Sp. Pros.* Chicago, IL, 100445, 1-4.

[3] Boersma, P., Weenink, D. Praat: Doing phonetics by computer. *http://www.praat.org*

[4] Bregman, A. 1994. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: The MIT Press.

[5] Ciocca, V., Bregman, A. 1987. Perceived continuity of gliding and steady-state tones through interrupting noise. *Perception and Psychophysics* 42(5), 476-484.

[6] Dannenbring, G. 1976. Perceived auditory continuity with alternately rising and falling frequency transitions. *Canadian J. of Psych.* 30(2), 99-114.

[7] Dennett, D. 1992. “Filling in” versus finding out: a ubiquitous confusion in cognitive science. In Pick Jr, H.L., van den Broek, P., Knill, D.C. (eds.), *Cognition: Conceptual and Methodological Issues*. Washington, D. C.: American Psych. Assoc.

[8] Fujisaki, H., Hirose, K. 1984. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *JASJ (E)* 5(4), 233-241.

[9] 't Hart, J. 1991 F0 stylization in speech: Straight lines versus parabolas. *JASA* 90(6), 3368-3370.

[10] Hermes, D. 1998. Auditory and visual similarity of pitch contours. *J. of Speech, Lang., and Hearing Res.* 41(1), 63-72.

[11] Hermes, D.J., 2006. Stylization of pitch contours. In Sudhoff, S., Lenertová, D. Meyer, R., Pappert, S., Augurzky, P., Mleinek, I., Richter, N., Schließer, J. (eds.), *Methods in Empirical Prosody Research.* Berlin-New York: de Gruyter, 29-62.

[12] Hirst, D., Espesser, R. 1993. Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l‘Institut de Phonétique d’Aix.* Univ. de Provence, 15, 71-85.

[13] House, D. 1990. *Tonal Perception in Speech*. Lund, Sweden: Lund University Press.

[14] d'Imperio. M. 2000. *The Role of Perception in Defining Tonal Targets and their Alignment*. Ph.D. Thesis, The Ohio State University.

[15] Knight, R.-A. 2007. The shape of nuclear falls and their effect on the perception of pitch and prominence: Peaks vs. Plateaux. *Lang. and Speech* 51(3), 223-244.

[16] Ladd, D.R. 1996/2008. *Intonational Phonology* (2nd ed.). Cambridge University Press.

[17] Nooteboom, S. 1997. The prosody of speech: Melody and rhythm. In Hardcastle, W., Laver, J. (eds.), *The Handbook of Phonetic Science*. Oxford: Blackwell, 640-673.

[18] Pierrehumbert, J. 1980. *The Phonetics and Phonology of English Intonation*. PhD Thesis, MIT.

[19] Taylor, P. 1998. The Tilt intonation model. In Mannell, R., Robert-Ribes, J. (eds.), *Proc. ICSLP 98*, 4, 1383-1386.