The effect of global F0 contour shape on the perception of tonal timing contrasts in American English intonation

Jonathan Barnes[1], *Nanette Veilleux*[2], *Alejna Brugos*[1] *and Stefanie Shattuck-Hufnagel*[3]
[1] Boston University, [2] Simmons College, [3] Massachusetts Institute of Technology
jabarnes@bu.edu, veilleux@simmons.edu, abrugos@bu.edu, stef@speech.mit.edu

Abstract

Results from an ABX perception task involving the contrast between default- and late-timed pitch accents ((L+)H* and L*+H) in American English intonation demonstrate that pitch movement curvature, in addition to turning-point alignment, plays a role in determining listener categorization. A model based on Tonal Center of Gravity, effectively integrating both F0 turning-point and global contour-shape information, is shown to provide a better account of these results than can a model based on turning-points alone. Results suggest further that additional factors, such as scaling of the pitch accent in the frequency domain, may also play a role.

Index Terms: intonation contrasts, F0 alignment, F0 turning points, tonal center of gravity, global F0 contour shape

1. Introduction

It has been common, at least since Bolinger (1951), to categorize models of intonation by the nature of their primitive elements. So-called "configuration-based" models characterize intonation primarily in terms of contrasting contour shapes (e.g., rises or falls). "Level-based" models, on the other hand, analyze contours as sequences of target tone levels (primarily Highs and Lows). The Autosegmental-Metrical approach (Pierrehumbert 1980, Ladd 1996/2008) numbers among the latter: intonation contours are viewed as strings of phonological tone-level specifications implemented as highly localized, systematically positioned F0 events, between which lie often lengthy, tonally unspecified stretches of the segmental string, through which the F0 curve is simply an interpolation between the nearest targets to either side.

These tonal targets are furthermore commonly identified with observable turning points in the F0 contour (maxima, minima, inflection points, hereafter TPs). Indeed, the intuitive core of the AM target-and-interpolation approach was famously presaged by Bruce (1977: 132), who observed that "reaching a certain pitch level at a particular point in time is the important thing, not the movement (rise or fall) itself". Ladd (1996/2008: 67) describes Bruce's discovery thus: "...the F0 configurations that happen to span the accented syllables play no useful role in phonetic description of the overall contour; the invariant features of the pitch system appear to be the *turning points* in the contour rather than the transitions that connect them." The investigation of TPs as reflections of tone specifications in the phonology has uncovered substantial systematicity in the way pitch movement onsets and offsets are timed with respect to the segmental string (Arvaniti, et al. 1998, Ladd, et al. 1999, Ladd & Schepman 2003, *inter alia*).

Nonetheless, there is increasing awareness in the literature that a direct equation of phonological tones and phonetic TPs is problematic in a variety of respects (Arvaniti, et al. 2006: 685, Ladd 1996/2008:180). Among the problems confronting a purely TP-based approach to intonational phonetics and phonology are the following: (1) Crucial TPs are frequently missing from the F0 curve owing to, e.g., intervals of voicelessness or irregular phonation. At the same time, psychoacoustic evidence suggests that human listeners lack the ability to extrapolate F0 trajectories to perceptually "restore" missing TPs such as peaks or valleys (Dannenbring 1976, Ciocca & Bregman 1987, Bregman 1994). (2) Even when the F0 curve is uninterrupted, precise locations of TPs are often ambiguous to the extent that no single point can be clearly identified as the target itself. Examples include high accentual plateaux (D'Imperio 2000, Knight 2008), and extended low regions flanked by shallow rises or falls lacking clear inflection points (Barnes, et al., *in press*). (3) Perhaps more worrisomely, an increasing number of sometimes difficult-to-quantify aspects of global contour shape have been shown to influence listener categorizations of F0 events. These include: "peak" shape (i.e. sharp peak vs. plateau: t'Hart 1991, D'Imperio 2000, Knight 2008), rise or fall duration (e.g., the fall following an L*+H pitch accent: D'Imperio 2000, Niebuhr 2007), and the relative scaling of rise onsets and fall offsets for High pitch accents (D'Imperio 2000). Pitch curvature differences of various kinds have also been implicated in potential differences of meaning by, e.g., Dombrowski & Niebuhr 2005, and Welby 2003.

1.1. Tonal timing and Tonal Center of Gravity

This paper is concerned primarily with the influence of global contour shape on the perception of tonal timing contrasts, such as those embodied in the distinction between the American English (L+)H* and L*+H pitch accents, referred to here also as *default-timed* and *late-timed* pitch accents.[1] These two pitch accents are widely seen as differing primarily in the location of the onset and offset of their accentual pitch rises. This difference is surely important, but as we will show, facts about TP alignment cannot be the end of the story.

Barnes et al. (2008) and Veilleux et al. (2009) present an alternative to solely-TP-based models of the phonetics and phonology of tonal contrast, based on the notion of Tonal Center of Gravity (TCoG). TCoG is a *gestalt* or global measure of the localization of F0 events in the temporal and frequency domains. As such, it succeeds in capturing key configurationist insights (such as the relevance of contour shape to the phonetics-phonology mapping), while nonetheless maintaining the core tenets of a level-based AM phonology. It does this by focusing not on the onsets and offsets of pitch movements, but rather on the overall distribution of the "mass" or bulk of raised F0 in both time and frequency space. It is therefore related to earlier generalizations about tonal structure based on *area under the F0 curve* (Barnes et al. *in press*, Knight 2008, Segerup & Nolan 2006), while avoiding some of their undesireable predictions. TCoG identifies the location of the "center-of-gravity" of an F0 event, which can serve as a reference location for it in perception. (See, e.g., Vishwanath & Kowler 2003, and many others, for a similar application of the notion of center-of-gravity in visual perception). TCoG in the time domain is computed as an average of discrete time values at sample locations within a region of interest. (Here, the entire region of elevated F0 corresponding to the pitch accent, the boundaries of which, unlike TPs, crucially require no particular precision in placement.) These time values are weighted by their F0 measurements, as shown in (1).

[1] The notion that, in English at least, the L*+H pitch accent is the marked member of this timing opposition is a common one (e.g., Pierrehumbert & Steele 1989, Arvaniti & Garding 2007). Our adoption of these terms is purely expository however.

We have demonstrated elsewhere (Barnes et al. 2008, Veilleux, et al. 2009) how TCoG accounts for the influence of a variety of contour shape phenomena on the perception of tonal timing contrasts; Figure 1 illustrates how pitch movement curvature, here of the rise associated with a high pitch accent, affects the temporal alignment of TCoG

$$T_{cog} = \frac{\sum_i F0_i t_i}{\sum_i F0_i} \quad (1)$$

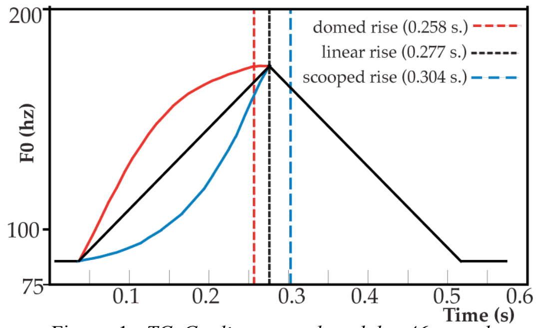

Figure 1: *TCoG alignment altered by 46 ms. due to curvature differences in the rise of a high pitch accent.*

TP-alignments for all three rise shapes depicted in Figure 1 are the same. However, since the "scooped" rise remains longer at a low F0 level, the bulk of the high F0 region falls later than for either the linear or "domed" rises. The dome, by contrast, pulls TCoG earlier. As a result, scooped rises should bias listeners toward the percept of later tonal timing patterns (English L*+H), while domed rises should have the opposite effect. In what follows, we will show that (1) pitch movement curvature, like other aspects of global contour shape, exerts an influence on listeners' perception of tonal timing contrasts, and (2) TCoG presents a more straightforward account of this phenomenon than purely TP-based models.

2. Methods

To determine the extent to which movement curvature influences the perception of tonal timing contrasts, we conducted an ABX matching-to-sample experiment. Subjects matched 49 target stimuli (synthetic F0 contours representing steps on continua for both TP-alignment and movement curvature) to one of two standard utterances containing canonical default-timed and late pitch accents.

2.1. Speech materials

The context in which our two pitch accents were presented is the rise-fall-rise contour, exemplified with both pitch accents (L+H* L-H% and L*+H L-H% in ToBI terms) in Figure 2, realized on the sentence *There's a mellower one.*[2]

All experimental presentations of this contour were based

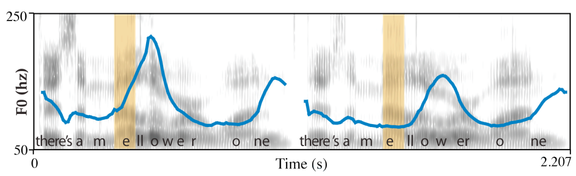

Figure 2: *Rise-fall-rise contours with default- (left) and late-timed (right) pitch accents. (Vertical bars indicate pitch-accented vowels.)*

on straight-line approximations of the rise-fall-rise pattern. These contours were superimposed on recordings of the utterance *"There's a lemonier meringue"* using the overlap-add resynthesis function in Praat (Boersma & Weenink 2009). For A and B, the standard exemplars, base recordings of default-timed and late-timed pitch accents were produced by a female ToBI-trained phonetician. Stylized pitch contours were created, with identical F0 levels for both default and late-timed standards. For the default-timed standard the rise began at the midpoint of the onset /l/ of the pitch accented syllable *lem-*, and the peak occurred at the midpoint of the ambisyllabic /m/, followed by a symmetrical fall. For the late-timed standard the rise began at the midpoint of the ambisyllabic /m/, For both standards the rise and fall were 166 ms. long.

For all 49 target stimuli, a recording of a male ToBI-trained phonetician uttering the same phrase served as the base contour for resynthesis. The recording was clipped from a larger utterance in which the relevant string was deaccented (i.e. *BOB said there's a lemonier meringue?!*), to achieve neutral duration and intensity patterns. Accentual TP-alignments and the curvature of the accent's rise were manipulated orthogonally to create seven-step continua along both dimensions (from early to late, and from domed to scooped), shown in Figures 3 and 4. As in the A and B tokens, F0 extrema were identically scaled across all 49 stimuli, with values derived from averages across 5 productions of default-timed and late rise-fall-rise contours by this male speaker.

The effect of manipulating TP-alignment on TCoG was straightforward: moving the beginning of the rise, the peak, and the end of the fall by 34 ms. per continuum step has the effect

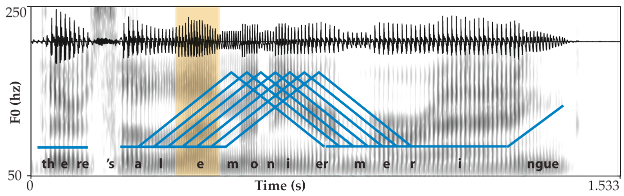

Figure 3: *A continuum of alignments, default to late.*

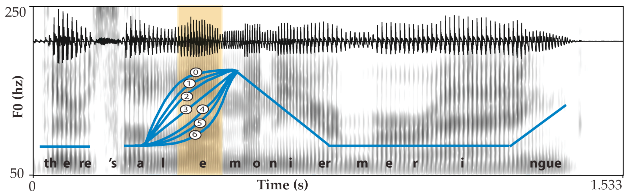

Figure 4: *A continuum of rise shapes, dome to scoop.*

of moving the TCoG of the elevated F0 region by the same 34 ms. per step.[3] The effect of the rise shape manipulation on TCoG is detailed in Table 1.

Rise shape	Domes			Line	Scoops		
	0	1	2	3	4	5	6
TCoG offset from linear rise	-27	-18	-12	0	19	27	41

Table 1. Effect of shape on TCoG. Offsets are ms. from the TCoG of the linear-rise pitch accent (i.e. rise shape 3, above).

2.2. Subject Selection and Training Procedure

31 subjects were recruited, of whom 19 reached the criteria for inclusion (see also Section 3 below). Subjects were primarily undergraduate students, all native speakers of American English, naïve as to the purpose of the experiment. For inclusion in the study, subjects were required to successfully complete a short screening procedure that also provided training in the ABX task. Only two subjects failed.

[2] Much has been written about meaning differences between these two accents (e.g., Pierrehumbert & Steele 1989, Pierrehumbert & Hirschberg 1990). For our purposes, it is crucial only that they contrast phonologically in this context.

[3] The directness of this influence has led us to hypothesize that the demonstrated systematicity of TP-alignment patterns in speech production may represent a particularly efficient articulatory strategy for speakers to achieve the differences in TCoG timing that we argue are crucial to listeners in perception.

2.3. Experimental procedure

Subjects carried out an ABX task in which default-timed and late-timed standards were balanced for order of presentation. Each of the 49 target stimuli—7 contour shapes x 7 TP-alignments—was presented in 2 trial types: default-timed-standard-first and late-timed-standard-first. Each of the 98 ABX types was presented twice, in random order, for 196 total presentations. Subjects were seated in a sound-attenuated room facing a computer monitor and wearing headphones; they pressed designated keyboard keys to indicate whether the target stimulus X sounded more like the A or B standard.

3. Results and Discussion

The literature has shown that manipulations of TP-alignment can significantly affect judgments of tonal timing (Kohler 1987, D'Imperio 2000, Niebuhr 2007, *inter alia*). Since the goal of this study was specifically to assess the extent to which manipulations of contour shape might *modulate* effects of TP-alignment, we included only subjects clearly demonstrating the influence of TPs. To this end, only subjects showing a highly significant effect of TP-alignment on categorization ($p < .001$ in a one-way ANOVA) were accepted, excluding 10 subjects who failed to reach this criterion.

Results presented here are therefore based on 3612 total judgments from 19 subjects. Figure 5 depicts percent late-timed or L*+H judgments for all subjects as a function of TP-alignment. Separate lines represent steps in our contour shape continuum, from most domed to most scooped, forming a series of seven sigmoid response curves. The regions of steepest slope, corresponding to the boundaries between pitch accent categories for each shape, shift progressively later in the alignment continuum as we move through the shape continuum. This suggests a strong influence of both TP-alignment and contour shape on listeners' responses, a conclusion reinforced by a repeated-measures ANOVA using percent late-timed responses as a dependent variable and TP-Alignment and Curve Shape as within-subjects factors. This analysis yields main effects both of Alignment, $F(3.15, 56.692) = 147.678$, $p < .001$ and Shape, $F(6, 108) = 91.296$, $p < .001$, and a significant interaction between the two, $F(10.101, 181.819) = 4.619$, $p < .001$.

In terms of TP-alignment, then, earlier continuum steps bias listeners toward default-timed or L+H* identifications, while later steps produce a bias toward late-timed or L*+H responses. In the shape dimension, more domed shapes yield proportionally more L+H* judgments, while more scooped shapes bias listeners in the other direction. (The interaction emerges because differences between responses for each contour shape are greater in the middle of the alignment continuum than they are at either end.) Planned comparisons between individual contour shapes yield significant differences (at $p < .05$) between listener response patterns to all contour shapes, save steps 1 and 2, the two less extreme domes. These results accord quite well with the patterns visible in Figure 5.

To summarize these results, then, it appears that early TP-alignment and domed rise shapes work synergistically to cue earlier timing patterns, while late TP-alignment and scoopier rise shapes have the same synergistic effect toward the percept of later tonal timing patterns. This is precisely the pattern that the TCoG hypothesis predicts we should find.

Turning now to the analysis of these facts, TP-alignment alone fails to account for the variation in listeners' responses arising from contour shape. Figure 6 recasts the data in Figure 5 as a scatterplot, in which percent L*+H judgments are plotted as a function of TP-alignment. Compare this now to a depiction of percent L*+H judgments as a function of the location of TCoG (Figure 7). Datapoints are more tightly clustered around a single notably sigmoid trend when depicted in terms of

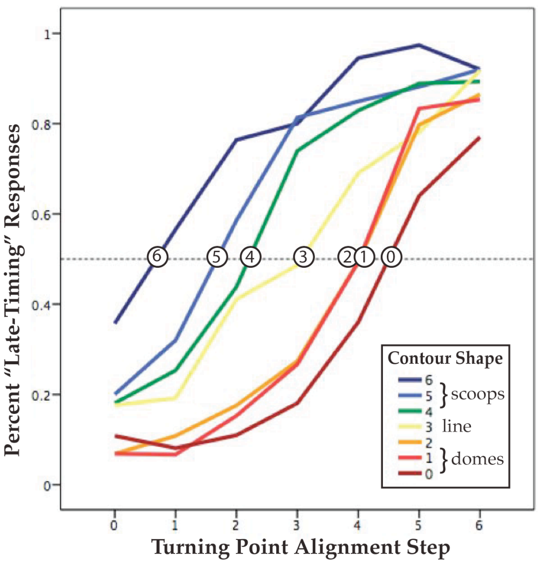

Figure 5: *Percent "late-timing" responses as a function of TP-alignment step, for all seven contour shapes.*

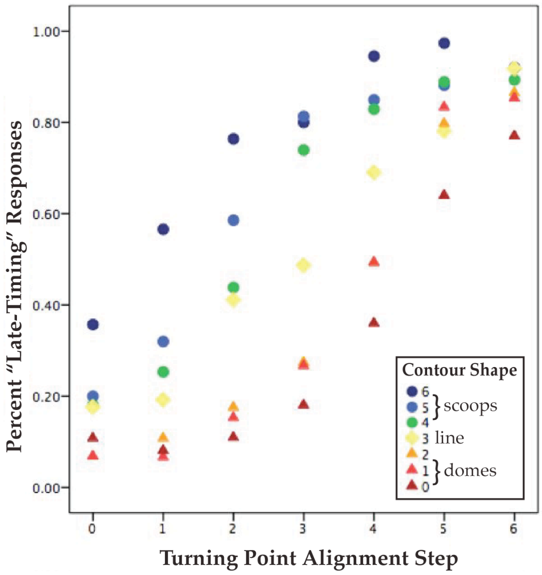

Figure 6. *TP-alignment as a predictor of listener categorizations.*

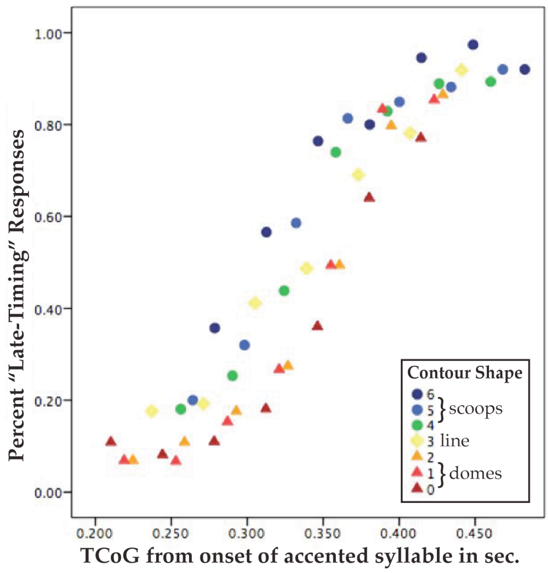

Figure 7. *TCoG-alignment as a predictor of listener categorizations.*

TCoG. This is as predicted: because TCoG distills the effects of both TP-alignment and contour shape, a model based on TCoG should provide a better account of our data.

While TCoG may be a stronger predictor of subjects' responses than TP-alignment alone, however, it still seems not to account for all of the effect of contour shape on accent categorization. Note that in Figure 7, particularly near the category boundary between the L*+H and L+H* pitch accents, dome-shaped rises still elicit fewer L*+H judgments than do scooped rises with similar TCoG alignments, a pattern confirmed by logistic regression analysis.

First, comparison of a binary logistic regression model using TP-alignment to predict listener responses (chi-square (1) = 1067.957, $p < .001$, Nagelkerke $r^2 = .342$) with a similar model using TCoG-alignment as a predictor (chi-square (1) = 1434.208, $p < .001$, Nagelkerke $r^2 = .437$) demonstrates the superiority of TCoG over TP-alignment.[4] That temporal TCoG does not account for all of the influence of contour shape is demonstrated by the results of a stepwise forward conditional logistic regression model, this time using both TCoG and contour shape as predictors. This model, as expected, selects TCoG as a predictor in its first step (chi-square (1) = 1434.208, $p < .001$), but then adds contour shape as a predictor in its second step (chi-square (7) = 1529.497, $p < .001$).

The relatively small, but significant, difference in chi-square values for the model at each step (95.289, $p < .001$) shows that contour shape still contributes some information to the model, over and above TCoG-alignment. What this additional information might be is still not entirely clear to us, but one possibility relates to an aspect of TCoG that has not yet entered into this discussion, namely the scaling of TCoG in the frequency domain. We are currently pursuing this line of thinking in an additional set of experiments.

4. Conclusions

In this paper we present evidence for the influence of pitch movement curvature on the perception of tonal timing contrasts in the intonational phonology of American English. We have argued elsewhere that Tonal Center of Gravity, a global measure modelling the localization of F0 events, can account for the demonstrated strength of both F0 TP-alignment and global contour shape as cues to intonational contrasts, while referring directly to neither. TCoG succeeds in modelling our experimental results involving pitch movement curvature in a way that solely-TP-based models cannot. We also hypothesize that a more complete account of the effect of global contour shape on the perception of tonal timing will need to take into account the localization of TCoG not just in time, but in the frequency domain as well.

5. Acknowledgments

This work was supported by NSF grant numbers 0842912, 0842782, and 0843181.

6. References

Arvaniti, A. & G. Garding. 2007. Dialectal variation in the rising accents of American English. In J. Cole & J. H. Hualde (eds.), *Papers in Laboratory Phonology 9*. Berlin, NY: Mouton de Gruyter: 547-576.

Arvaniti A., Ladd, D. R. & I. Mennen. 1998. Stability of Tonal Alignment: the case of Greek prenuclear accents. *Journal of Phonetics* **26**: 3-25.

Arvaniti, A., Ladd, D. R. & I. Mennen. 2006. Phonetic effects of focus and 'tonal crowding' in intonation: evidence from Greek polar questions. *Speech Communication **48**(6): 667-696.*

Barnes, J., Brugos, A., Shattuck-Hufnagel, S. & N. Veilleux. 2008. Alternatives to F0 turning points in American English intonation. *Journal of the Acoustical Society of America* **124**: 2497. Poster presented at ASA 156, Miami, FL., November 2008.

Barnes, J., Veilleux, N., Brugos, A. & S. Shattuck-Hufnagel. In press. Turning points, tonal targets, and the English L- phrase accent. *Language and Cognitive Processes*, forthcoming.

Boersma, P. & D. Weenink, 2009. Praat: doing phonetics by computer. (Version 5.1). Retrieved from http://www.praat.org/.

Bolinger, D. L. 1951. Intonation: Levels vs. Configurations. *Word* **7**: 199-210.

Bregman, A. S. 1994. *Auditory Scene Analysis: The Perceptual Analysis of Sound.* Cambridge, MA: MIT Press, 1990. Bradford Books: paperback edition, 1994.

Bruce, G. 1977. *Swedish Word Accents in a Sentence Perspective.* Lund: Gleerup.

Ciocca, V. & A. S. Bregman. 1987. Perceived continuity of gliding and steady-state tones through interrupting noise. *Perception & Psychophysics* **42**: 476-484.

Dannenbring, G.L. 1976. Perceived auditory continuity with alternately rising and falling frequency transitions. *Canadian Journal of Psychology* **30**: 99-114.

D'Imperio, M. 2000. *The Role of Perception in Defining Tonal Targets and their Alignment*. PhD thesis, Ohio State University.

Dombrowski, E. & O. Niebuhr. 2005. Acoustic patterns and communicative functions of phrase-final rises in German: activating and restricting contours. *Phonetica* **62**(2-4), 176-195.

Knight, R-A. 2008. The shape of nuclear falls and their effect on the perception of pitch and prominence: peaks vs. plateaux. *Language and Speech* **51** (3), 223-244.

Kohler, K. 1987. Categorical pitch perception. *Proceedings of the XIth International Congress of Phonetic Sciences*, Tallinn: 331-333.

Ladd D. R. & A. Schepman. 2003. "Sagging transitions" between high pitch accents in English: experimental evidence. *Journal of Phonetics* **31**: 81-112.

Ladd D. R., Faulkner, D., Faulkner, H. & A. Schepman. 1999. Constant "segmental anchoring" of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America* **106**: 1543-1554.

Ladd, D. R. 1996/2008. *Intonational Phonology*. New York: Cambridge University Press.

Niebuhr, O. 2007. The signalling of German rising-falling intonation categories—the interplay of synchronization, shape, and height. *Phonetica* **64**: 174-193.

Pierrehumbert, J.B. 1980. *The Phonetics and Phonology of English Intonation*. PhD thesis, MIT.

Pierrehumbert, J. & J. Hirschberg. 1990. The meaning of intonation in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (eds.), *Intentions in Communication*, Cambridge, MA: MIT Press: 271-311.

Pierrehumbert, J. & S. Steele. 1989. Categories of tonal alignment in English, *Phonetica* **47**: 181-196.

Segerup, M. & F. Nolan. 2006. Gothenburg Swedish word accents: a case of cue trading? In G. Bruce & M. Horne (eds.), *Nordic Prosody: Proceedings of the IXth Conference.* Frankfurt am Main: Peter Lang: 225-233.

't Hart, J. 1991. F0 stylization of speech: straight lines versus parabolas. *Journal of the Acoustical Society of America* **90**: 3368-3370.

Vishwanath, D. & E. Kowler. 2003. Localization of shapes: eye movements and perception compared. *Vision Research* **43** (15): 1637-1653.

Veilleux, N., Barnes, J., Shattuck-Hufnagel, S. & A. Brugos. 2009. Perceptual robustness of the tonal center of gravity for contour classification. Presented at the Workshop on Prosody and Meaning, Barcelona, September 2009.

Welby, P. 2003. *The Slaying of Lady Mondegreen, being a Study of French Tonal Association and Alignment and their Role in Speech Segmentation.* PhD thesis, Ohio State University.

---

[4] The predictably high degree of collinearity between TP-alignment and TCoG-alignment ($r = .945$, $p < .001$) makes it inadvisable to enter both as predictors into any single regression model. This is not a problem for TCoG and contour shape, however, owing to their substantially lesser collinearity ($r = .327$, $p < .001$).