Barnes, Jonathan, Alejna Brugos, Stefanie Shattuck-Hufnagel & Nanette Veilleux. (Forthcoming.) On the nature of perceptual differences between accentual peaks and plateaux. In Oliver Niebuhr & Hartmut Pfitzinger (eds.), Prosodies: Context, Function, Communication. Berlin/New York: Mouton de Gruyter.

On the nature of perceptual differences between accentual peaks and plateaux

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

1. Background.

It is probably not much of an overstatement to say that the theoretical mainstream in intonational phonology today takes the historical debate between level-based and configuration-based approaches to intonation to be resolved in such a way that explicit reference to contour shape no longer has any place in phonological representations. Part of what has made this an attractive position is the high degree of within-category variability one encounters in the shapes of phonological entities such as pitch accents in phonetic realization. It is well-known, for example, that a given pitch accent category (e.g., English L+H*, as shown in Figure 1) can be realized either as a sharp peak (as on the left), or instead can linger for a time near the F0 maximum, creating what is often referred to as a plateau-shaped configuration (as on the right). This distinction is not known to form the basis of a phonological contrast in any intonation system, nor is it clear whether the distinction is categorical, or even whether it is conditioned by contextual factors.

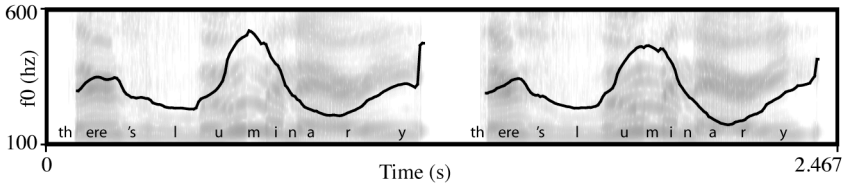

Figure 1: Two productions of L+H L-H% on* there's luminary*, one showing a sharp peak (left) and the other a plateau (right).*

Plateau-shaped accents have nonetheless proven problematic for mainstream level-based theories of intonational phonology, such as the

Autosegmental-Metrical approach (hereafter AM, Pierrehumbert 1980, Ladd 1996/2008), for two reasons. First, to the extent that tonal realization is typically conceived in level-based models as a matter of the precise alignment (in time) and scaling (in F0 space) of a tonal target corresponding to (more or less) each tone specification in the phonology, it becomes unclear how any unique point within the high-F0 region of the plateau could be reliably singled out as 'what matters' in terms of alignment and scaling; even where a single maximum can be located, it is often not obviously distinct from surrounding points in any meaningful way.

More worrying still, however, is a second consideration: the substantial and growing body of evidence suggesting that variations in contour shape, such as that between peaks and plateaux, in fact influence the perception of tone in both the timing and scaling dimensions in ways that seem incompatible with standard assumptions of level-based models (See D'Imperio 2000, Niebuhr 2007a, and Barnes et al. 2010). In the case of plateau scaling, it has been known for some time ('t Hart 1991, D'Imperio 2000, Knight 2008) that all things being equal, a pitch accent realized as an F0 plateau will sound higher to listeners than an analogous sharp peak with identical maximum F0. Most interpretations of this result seem to rely on the notion that listeners have greater exposure to, and thus more opportunity to perceive, the F0 maximum within the plateau, as compared to the briefer maximum of a sharp F0 peak, which is thereby somehow rendered less effective perceptually. The search for a single point within the plateau to identify as the target therefore begins to seem somewhat misguided, since even in the case of sharp peaks, the perception of scaling apparently relies on something more than just the scaling of a single F0 point.

In the temporal domain, the picture might initially appear simpler. Knight & Nolan (2006), for example, find evidence of greater stability in the alignment of the offset of F0 plateaux relative to segmental material than in either the onset of the plateau, or the F0 maximum during same.[1] This leads them to posit that the end of the plateau is the crucial point in terms of signalling differences in linguistic structure. D'Imperio (2000), however, in the context of a perception study involving a contrast in Neapolitan Italian between one pitch accent timed earlier with respect to

1 Knight and Nolan locate plateau onsets and offsets by identifying a region around the measured F0 maximum wherein F0 has declined to either side by less than 4%. This criterion, arrived at heuristically, is based on an estimate of jnd for pitch.

2 Conveniently, in this dialect, the choice between the pitch accents in question in the context of a particular overall contour underlies the distinction between a statement and a question.

3 D'Imperio 2000 in fact suggests something of this sort to handle the scaling question, and

the accented syllable (analyzed as L+H* in AM terms) and one timed later (L*+H), achieves a finding that is harder to interpret in terms of tonal targets identified with turning points in the F0 curve (hereafter TPs). In this study, which makes use of a paradigm pioneered in intonation research by Kohler (1987), listeners were asked to categorize a set of utterances with synthetic F0 contours representing a continuum of different alignments with respect to the accented syllable, beginning at something like canonical L+H*, and ending at something like canonical L*+H.[2] At each step in this continuum, a variety of differently shaped pitch accents were presented to listeners. Among these were symmetrical sharp peaks, and 45 ms. F0 plateaux. D'Imperio's findings were twofold: first, she demonstrated that, all things being equal, plateau-shaped pitch accents elicited a greater proportion of later-timed, or L*+H judgments, than did sharp peaks reaching their F0 maxima at the onset of these plateaux. Put another way, accents shaped like the plateau pictured in Figure 2a below tended to 'sound later' to listeners than did the sharp peak in the same illustration. At the same time, however, plateau-shaped pitch accents also elicited fewer L*+H judgments than did analogous sharp peaks timed to coincide with plateau offsets (D'Imperio 2000: 169, though as we will see, the magnitude of this difference was smaller). Figure 2b depicts this latter situation. In terms of the perception of tonal alignment, then, if we wish to posit a single point within F0 plateaux that functions as a sort of 'peak analog', the Neapolitan data suggests that this point is to be found neither right at the beginning, nor right at the end, of the F0 plateau, but instead, somewhere in between. The consequences of this finding for TP-based models, to the extent that this region contains no turning points, should be clear enough.

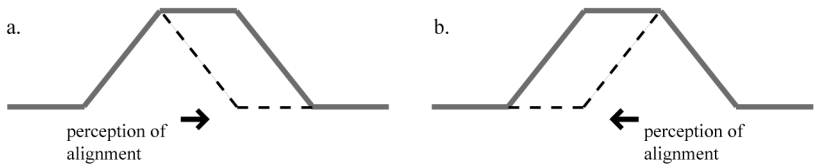

Figure 2: Schematic showing how two plateau alignments (with respect to the same sharp peak) affect listener perception of tonal timing.

2 Conveniently, in this dialect, the choice between the pitch accents in question in the context of a particular overall contour underlies the distinction between a statement and a question.

2. Tonal Center of Gravity.

Barnes et al. (2010) propose an alternative model of the alignment and scaling of F0 events that is intended to provide a non-TP based remedy to the problems of tonal implementation sketched above (among others). This approach accounts for the substantial findings concerning systematic alignment of TPs in the recent literature (Arvaniti, et al. 1998, Ladd, et al. 1999, Ladd & Schepman 2003, *inter alia*), while at the same time providing an explanation for the range of contour shape effects reported in the literature. This model, based on the notion of Tonal Center of Gravity (TCoG), is a global theory of tonal implementation. That is, the alignment and scaling of the F0 targets pertaining to intonational events are construed not in terms of the location of any specific point or points within the F0 contour, but rather in terms of the overall disposition of the bulk or "mass" of the (upward or downwardly) displaced F0 region associated with the tone specification in question. TCoG derives what can be considered a perceptual reference location for an F0 event (in both the temporal and frequency dimensions), defined as that event's "center-of-gravity". In the time domain, TCoG is computed as an average of discrete time values at sample locations within a region of interest, with each sample weighted by its measured F0 value, as shown in (1). (For High tone specifications such as those contained in the pitch accents in question, the 'region of interest' or integration window for the calculation of TCoG extends over the entire region of elevated F0 associated with the pitch accent. As Veilleux et al. (2009) and Barnes et al. (in prep.) show, however, the precise locations of the boundaries of this region are relatively unimportant.)

1. $$T_{cog} = \frac{\sum_i f0_i \times t_i}{\sum_i f0_i}$$

We have demonstrated elsewhere (Barnes et al. 2008, Veilleux, et al. 2009, Barnes et al. 2010) how an initial version of TCoG (call it version 1.0), accounts for the influence of a variety of contour shape phenomena on the perception of tonal timing contrasts; Figures 3a & b below illustrate schematically how an approach like TCoG could account for the plateau-timing effects reported by D'Imperio (2000). In the case of a perfectly symmetrical rising-falling accent pattern with a sharp peak, TCoG will simply coincide with the accent's F0 maximum, because the elevated F0 region is distributed evenly to either side of this point. In the case of an

analogously symmetrical plateau-shaped accent, TCoG 1.0 will by a similar logic place the center-of-gravity at the plateau's midpoint. From the point of view of D'Imperio's findings, this initially seems like a welcome result: If a plateau-shaped pitch accent, all things being equal, sounds "later" to Neapolitan listeners than does a sharp peak timed at that plateau's onset (as in 3a below), this is because TCoG of the plateau necessarily falls later. By the same token, plateaux should sound earlier than sharp peaks timed at their offsets, again owing to their relatively earlier locations of TCoG (as in 3b). The model thus correctly predicts that the hypothetical peak analog point within an intonational plateau (reconceived as TCoG) falls neither at the end of the plateau, nor at its beginning, but somewhere in between. Furthermore, an account based on integrating F0 information over time, here by means of a weighted average, seems extensible to the scaling facts reviewed above as well.[3] The peculiarities of plateaux in both the alignment and scaling dimensions may therefore be amenable to explanation by a single, unified approach to the perception of F0 events.

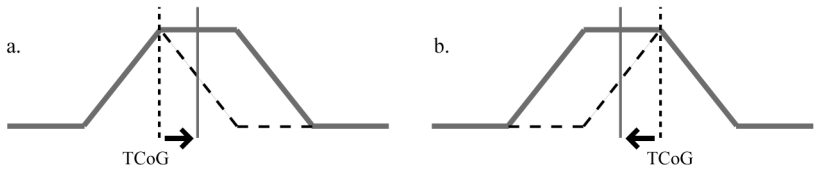

Figure 3a & b: Centers-of-gravity for sharp peaks and plateaux.

There is, however, a problem. The prediction of the TCoG model as implemented above for the location of the peak analog point is quite specific: for a perfectly symmetrical plateau, TCoG should fall precisely halfway between plateau onset and plateau offset. This is not, however, what D'Imperio reports. As D'Imperio, et al. (2010) underscore, while the analog point was statistically equivalent neither to the plateau onset nor to its offset, it was considerably closer to the latter than to the former.[4] This

3 D'Imperio 2000 in fact suggests something of this sort to handle the scaling question, and weighted averages are familiar tools in other models involving the apparent integration of pitch information over some temporal interval (e.g., d'Alessandro, et al. 1998).

4 Another study summarized in D'Imperio, et al. (2010), this one involving Pisan Italian, using similar methodology, finds no difference in terms of response patterns between plateaux and sharp peaks timed at plateau offset. We suspect, however, that this owes to a confound between the relative roles of accent alignment and scaling in the perception of the contrast in question. Specifically, in Pisan Italian, the later of the two accents under comparison is also typically scaled higher. This study furthermore demonstrates by comparing a range of scalings for both peaks and plateaux that higher perceived scaling biases listeners toward categorization of a stimulus as the "later" of the two accents. The perception of peak and plateau timing is then investigated using peaks and plateaux that

is still bad news for a TP-based approach, but also clearly a problem for TCoG 1.0.

While the results presented in D'Imperio (2000) concerning Neapolitan are clear and persuasive indeed, we were struck by how contrary to the pattern reported there our own initial attempts to create peaks and plateaux along a continuum from L+H* to L*+H in English turned out to be. In fact, informal observation suggested that for English, the opposite pattern might even obtain (i.e. a plateau might sound, in timing terms, more like a sharp peak aligned at the plateau's beginning, than one aligned at its end)[5]. D'Imperio, et al. speculate, we believe correctly, that perceptual differences between peaks and plateaux are psychoacoustic in origin, and thus might be expected to hold regardless of the languages or intonational contrasts involved. The possibility that these differences might be subject to language- or even contrast-specific variation, however, suggested to us that something was missing from our TCoG version 1.0 picture of how accent timing and scaling patterns are perceived.

A clue as to the nature of this missing element came to us from a more detailed comparison of the phonetic realization of the accents in question in both English and Neapolitan. Specifically, while these accents typically receive identical phonological representations in AM systems in both languages (that is, L+H* and L*+H), the tonal targets in question are in fact timed substantially earlier with respect to the pitch-accented vowel in Neapolitan than they are in English.[6] In English, L+H* pitch accents typically rise over the entirety of the accented vowel, frequently reaching peaks in following consonants, or even the vowels of following unaccented syllables. L*+H, on the other hand, typically rises not at all during, or only toward the very end of the accented syllable, placing the bulk of the high F0 region well beyond the bounds of the accented syllable. In Neapolitan Italian, the situation is quite different: in that

---

are matched for scaling acoustically, in terms of their maximum F0. The problem with this is that plateaux are known ('t Hart 1991, Knight 2008, and indeed this very study), to sound higher than peaks with identical maximum F0. This being the case, we would also expect plateau-shaped accents to bias Pisan listeners more toward the later accent category than would sharp peaks with identical perceived timing. Put another way, if a given plateau sounds only as "late" as a sharp peak timed at its offset, despite its higher perceived scaling, then we would expect a plateau that matched that peak in perceived scaling to sound "earlier" than this. The peak analog point, in other words, might fall near, but not quite at, the plateau's offset, a result which would be parallel to what D'Imperio finds for Neapolitan Italian.

5 That is, it seemed to us more relevant where the plateau's rise occurred, than where its fall was.

6 All our information concerning tonal implementation in Neapolitan Italian is taken from D'Imperio (2000).

language, what is transcribed L+H* is in fact timed much earlier than its English analog, so that (apparently depending to some extent on the speaker) not only the peak, but even the bulk of the fall, takes place well within the duration of the accented vowel. L*+H, by contrast, is actually in phonetic terms a great deal more like English L+H*, in that it typically rises throughout the accented vowel, reaching a peak toward vowel offset, and then falls during following segments.

For present purposes, what this means is that, if one were to construct a continuum of pitch accent alignments between these two extremes in each language, the territory over which this continuum would be realized relative to the pitch accented vowel would be very different in the two cases. In English, only the earliest steps of such a continuum would see much of the high region associated with the H tonal target realized during the pitch accented vowel, with the bulk of the high F0 occurring only later. In Neapolitan, on the other hand, the continuum that D'Imperio constructed began with the sharp peak's F0 maximum approximately 35% of the way through the pitch-accented vowel, and ended with the F0 maximum about 88% of the way through same. In other words, for most of the continuum steps the bulk of the high F0 region (i.e. more than 50% of its "mass") coincided with the pitch-accented vowel.

Why this could be important is as follows: we have known since the beginning of the TCoG research program that the model contained (at least) one assumption about the perception of F0 events that was unlikely to prove correct in the long run. Specifically, as was reviewed above, TCoG 1.0 takes time values from within the window of integration, and then weights them by their measured F0, such that higher F0 matters more for the location a High tone's TCoG than lower F0. Importantly, however, beyond this the model treats all samples within the window of interest equivalently: no higher or lower perceptual influence is posited for samples on the basis of, for example, how they coincide with the segmental string over which the F0 event is realized. But this is almost certainly a mistake.

In fact, there is abundant reason from the literature to believe that tone is not perceived equally robustly over all segment types, or within all sections of a given segmental string. For example, the preferential crosslinguistic licensing of tonal contrasts within higher sonority regions (e.g., the syllable rhyme, and more sonorous rhymes in particular) has been attributed to increased perceptibility of tone during more sonorous segments (Gordon 2001, Zhang 2002, *et seq.,* Flemming, to appear). In addition, other models of tone perception interested in the integration of F0 information over time (exclusively, as far as we know, in the frequency

domain) have noted that, even over the course of a single vowel, not all regions of the F0 contour contribute equally to the distillation of a perceived target scaling. House (1990), for example, models the endpoints of accentual tone movements by averaging F0 over the final 32 ms. of the accented vowel.[7] d'Alessandro et al. (1998), likewise, model the perception of the endpoint of a pitch rise realized on a synthetic vowel by averaging over the final region of the vowel, with sample weights increasing toward the end of the specified window. Both of these models suggest the existence of something like a perceptual "sweet spot" within the accented vowel, likely over and above the effects of sonority and signal level referenced above.

It seems clear to us, based on the above, that introducing into TCoG an additional weighting factor, one that is sensitive to the nature of the segmental backdrop against with the relevant F0 samples are realized, is already necessary for independent reasons. It also seems likely to us that potential differences in the perceived timing of peak- and plateau-shaped accents in English and Neapolitan Italian could be, relatedly, the result of differences in the way the accents in question are timed with respect to the segmental string in those languages. That is, a model of TCoG augmented in this way would account for differences in the perceived timing of identically-shaped plateaux in the two languages as arising from different degrees of overlap (for the same F0 shape) with regions of higher and lower salience for tonal perception. The resulting differences in weight applied to what are otherwise 'the same' F0 samples would yield different TCoG alignments, and with them the documented perceptual patterns.

We are currently in the process of implementing such a revision to the TCoG model. As a first step, we must verify the perceptual facts themselves experimentally, and then consider the extent to which a model of the type just outlined is likely to succeed in accounting for the perceptual patterns we document. The following study aims to do just this for two pairs of contrasting pitch accents in American English, using perceptual experiments based heavily on those found in D'Imperio's investigation of Neapolitan.

---

7 House's approach rests on the idea that specific properties of the signal (e.g., spectral stability or instability) are responsible for local alterations in the robustness of tonal perception. His model, however, locates target windows for extracting average F0 values not directly via detecting those properties themselves, but rather using a temporally fixed window, during which the properties in question tend to be found.

3. An Experiment.

In order to examine potential differences in the perceived timing of pitch accents realized as high plateaux and those realized as sharp peaks, we first selected two American English accentual contrasts to form the basis of two alignment continua constructed à la D'Imperio (2000). These contrasts were 1) relatively early-aligned (with respect to the pitch-accented vowel) peaks (AM H+!H*) vs. more-or-less medially-aligned peaks (AM L+H*), and 2) medially-aligned L+H* vs. much-later-aligned peaks (L*+H). (See Figure 4 for examples.) On the basis of these contrasts, we constructed two synthetic tonal alignment continua: early-to-mid, and mid-to-late. Each of these continua was rendered in two variants: as sharp-peaked accents and as plateau-shaped accents. These contours were resynthesized using straight line approximation from a recording of a native English speaker using Praat (Boersma & Weenink 2009).

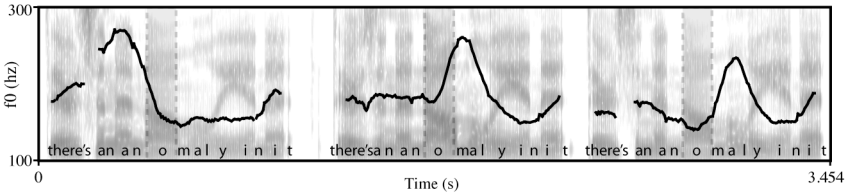

Figure 4. Natural productions of the three pitch accents under investigation, realized on the carrier sentence there's an anomaly in it. *Left to right, these are H+!H*, L+H*, and L*+H*

3.1.1. Stimulus creation.

Our alignment continua were constructed using as a model 10+ renditions of a ToBI-trained male speaker producing the sentence "*There's an anomaly in it"* with each of the 3 relevant pitch accents followed by a final rise. The average alignment with respect to segmental boundaries and average values of F0 turning points were used to construct the endpoints in our synthetic alignment continua. A recording of the same male speaker saying *"BOB said there's an anomaly in it?!",* with the target phrase deaccented, was used as the base for resynthesis. The precise shapes of the synthetic contours (i.e. the alignment and scaling of the F0 points connecting straight line segments) for each of the two alignment continua were determined separately, based on measurements of the elicited productions of the endpoint pitch accents of each continuum.

3.1.1.1. The early-to-mid continuum

For the early-to-mid continuum populated by sharp-peak-shaped pitch accents, the shape of the peak was determined by averaging F0 values and alignment at the rise onset, peak and fall offset across productions of pooled H+!H* and L+H* recorded models. This procedure yielded a shape that was a slightly asymmetrical peak, with a sharper fall than rise. The F0 rise began at 103 hz, peaked at 146 Hz. and then fell to 100 Hz. The alignment of the peaks in the continuum began at 70 ms before the vowel onset and ended at 37 ms after the end of the vowel (vowel duration: 147 ms). This resulted in a peak-to peak distance across the continuum of 250 ms., divided into 10 steps of 25 ms, giving 11 sharp peak stimuli. (See Figure 5.)

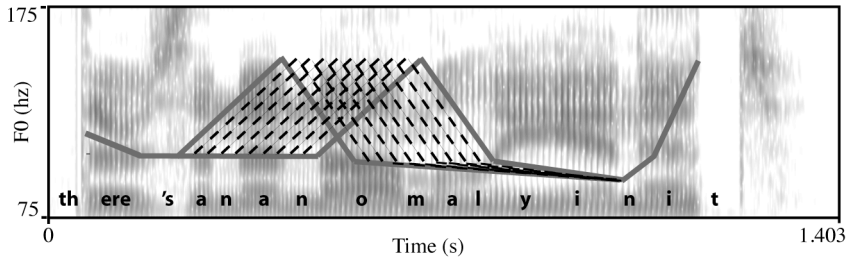

Figure 5: The early-to-mid continuum of sharp peaks.

The plateau version of the early-to-mid continuum was created using the rise and fall shapes from the sharp peak continuum, but "walking them apart" 3 steps to create a 75 ms plateau between rise and fall. The endpoints of the plateau continuum were chosen such that the fall of the leftmost plateau corresponded with the same segmental alignment as the fall of the leftmost sharp peak, while the rise of the rightmost plateau coincided with the rise of the rightmost sharp peak. As with the sharp peaks, the step size in the continuum was 25 ms, giving 14 plateau steps. (See Figure 6.)

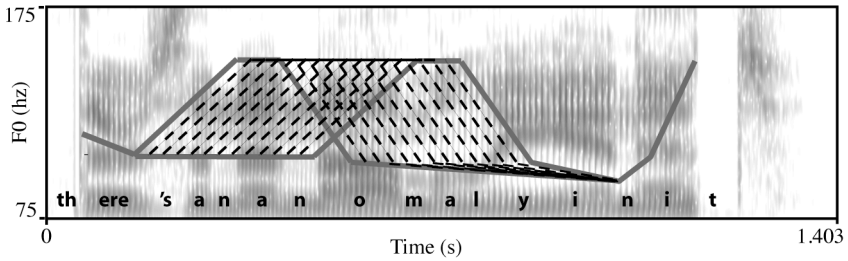

Figure 6: The early-to-mid continuum of plateaux.

3.1.1.2. The mid-to-late continuum

For the mid-to-late continuum consisting of sharp peaks, peak shapes were determined much as above, by averaging the relative alignments and F0 values for the elicited L+H* and L*+H models. Here, the implemented rise onset began at 95 Hz, 165 ms before the peak, rose to 141 Hz, and then fell to a 95 Hz elbow 119 ms after the peak. The leftmost peak was aligned 16 ms after the end of the pitch-accented vowel and the rightmost 150 ms later. This interval was divided into steps of 25 ms. giving 7 continuum steps. (See Figure 7.)

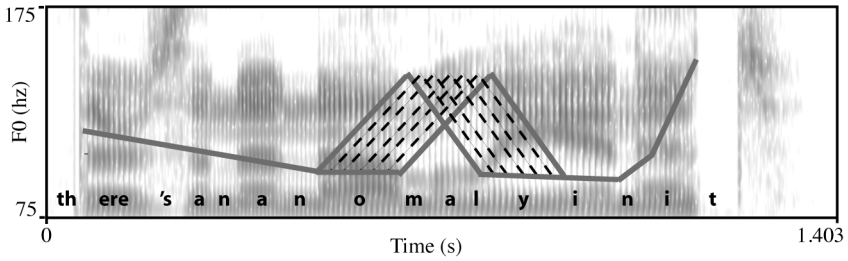

Figure 7: The mid-to-late continuum of sharp peaks.

The plateaux of the mid-to-late continuum were created as described above for the early-to-mid continuum, by "walking apart" the rise and fall of the sharp peaks to create a 75 ms plateau between the rise and fall. The endpoints of the continuum were chosen, again as above, such that the fall of the leftmost plateau coincided with the fall of the leftmost sharp peak, and that the rise of the rightmost plateau coincided with the rise of the rightmost sharp peak. As with the sharp peaks, the step size in the continuum was 25 ms, giving 10 plateau steps. (See Figure 8.)

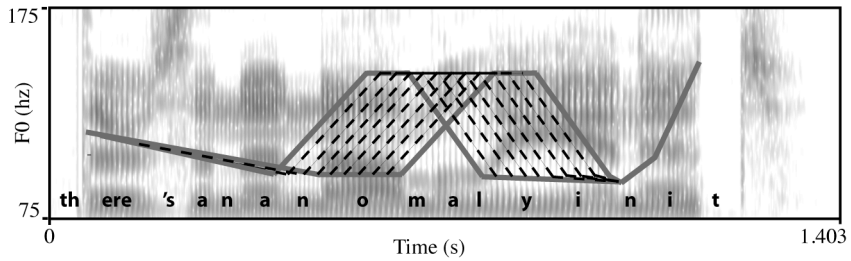

Figure 8: The mid-to-late continuum of plateaux.

3.1.2. Subject Selection and Training Procedure

50 subjects were recruited, of whom 28 reached performance criteria set for inclusion (on which see Section 3.2 below). Subjects were all native speakers of American English, and naïve as to the purpose of the experiment. As an initial screening procedure, subjects were required to successfully complete a short training session using natural, unambiguous productions of the three contours in question. This session was also designed to introduce the AXBX task described below. Only 2 subjects were excluded from the study at this point.

Since the meanings of these English contours are, in our experience, substantially less directly accessible to subjects than the narrow focus statement vs. question contrast of Neapolitan Italian, this study relied on a matching-to-sample task to investigate the perception of tonal timing within the stimuli from our alignment continua. In this task, a variant on the familiar 4IAX design that we might call AXBX, subjects are played two pairs of stimuli. In each AXBX presentation, A and B (the standard exemplars) were the sharp-peak versions of the relevant continuum's endpoints, and X was a test item from somewhere in the continuum. The AX and BX pairings were separated by a 500 ms pause. Subjects were seated in a sound-attenuated room facing a computer monitor and wearing headphones; they were asked to indicate whether it was the first pair (AX) or the second (BX) that sounded like a "match" (i.e. in which the intonation patterns sounded "more similar").

The experimental session was broken into 2 parts, one focused on the mid-to-late contrast, the other on the early-to-mid contrast. Order of presentation was balanced. Stimuli for both continua were presented in 6 blocks, each of which contained all steps of the relevant alignment continuum in both peak and plateau variants in a randomized order. In

three of these blocks, the A sample in the AXBX task was the left endpoint contour of the continuum, and in the other three, A was the right endpoint contour. In all, subjects heard 252 experimental stimuli: 150 from the early-to-mid continuum, and 102 from the mid-to-late.

3.2. Results.

Results from the AXBX perception tests for our two synthetic alignment continua are as follows. Before proceeding to analyze results pooled over subjects, we first reviewed each subject's results individually. Because it was clear that some subjects, including those that had passed the initial screening procedure, were nonetheless finding the task quite difficult, we imposed an additional criterion for inclusion in the study on all subjects. This criterion involved analyzing each subject's results individually using a binary logistic regression model, in which response category (e.g., L+H* or L*+H) was the dependent variable, and continuum step was the sole predictor. The responses to (only) the peak-shaped stimuli of the two continua were analyzed separately for each subject. To be included in the study, subjects needed to display a significant effect (at $p < .05$) of continuum step in their response patterns. The rationale for this criterion was as follows: The effectiveness of accent alignment continua in eliciting roughly sigmoid shaped response curves as part of a categorization task has by now been demonstrated in a number of languages by many different researchers (e.g., Kohler 1987, D'Imperio 2000, Niebuhr 2007a, *inter alia*). If a particular listener, however, does not show a significant tendency for the percentage of "late timing" responses to increase as a sharp peak is moved later along one of our continua, then it makes little sense to inquire further as to how changing that peak to a plateau affects the categorization pattern. As noted above, imposing this criterion winnowed our pool of subjects down to 28.

3.2.1. Early-to-mid continuum.

Taking first the early-to-mid continuum, Figure 9 displays the percentage of "mid" or L+H* timing judgments (averaged across listeners) as a function of continuum step and shape. Continuum steps are numbered so that plateaux share step numbers with sharp peaks timed at their offsets. If the timing of plateaux is perceived as identical to that of sharp peaks aligned at plateau offset, à la D'Imperio, et al. (2010), we would expect to see no difference in the response patterns for peaks and plateaux; that is,

we would expect a statistically significant effect of continuum step, but no significant effect of F0 contour shape. This expectation is robustly disconfirmed. First, as Figure 9 shows, there is a clear difference in the response patterns for peaks and plateaux, with plateau-shaped accents biasing listeners toward earlier timing judgments than sharp peaks.

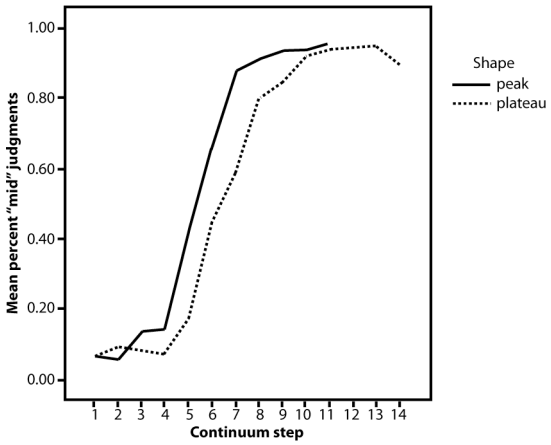

Figure 9: Mean percent L+H judgments as a function of continuum step and accent shape, pooled across listeners, for the early-to-mid continuum. At any given continuum step, the peak is aligned at the plateau's offset.*

Modelling these results using binary logistic regression, with listener response ("early" or "mid") as an independent variable, and continuum step number and accent shape as predictors, tells a similar story.[8] Specifically, a forward stepwise (LR) model selects continuum step as its first predictor (Chi-square (1) = 2132.235, $p < .001$, Nagelkerke $r2$ = .594). In its second step, however, the model adds accent shape as a predictor (Chi-square (2) = 2202.839, $p < .001$, Nagelkerke $r2$ = .608). The difference in fit between the two models is highly significant (Change in -2 Log Likelihood = 70.604, $p < .001$).

In sum, for the early-to-mid continuum the timing of plateaux is not judged purely with reference to the timing of the F0 fall. Nor, however, is it perceived purely with reference to the timing of its rise: Figure 10 below shows the same data, reconfigured this time in such a way as to match plateau continuum step numbers with peaks timed at plateau onset, rather

8 When carrying out such analyses, throughout, we have only included data for plateaux at continuum steps for which analogous sharp peak data exists. That is, for the early-to-mid continuum, we include only steps 1 through 11.

than offset.[9] Again, the two lines representing shape are clearly distinct, this time with plateau-shaped accents crossing the 50% categorization shift threshold earlier than peak-shaped accents.

As before, a logistic regression model underscores the validity of this generalization: the forward stepwise (LR) model selects continuum step as its first predictor (Chi-square (1) = 1819.239, $p < .001$, Nagelkerke $r2$ = .539). In its second step, the model again adds accent shape as a predictor (Chi-square (2) = 1943.801, $p < .001$, Nagelkerke $r2$ = .567), and once again the difference in fit between the two models is highly significant (Change in -2 Log Likelihood = 124.562, $p < .001$).

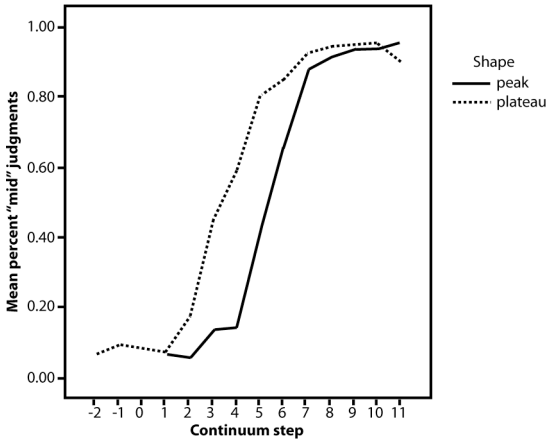

Figure 10: Mean percent L+H judgments as a function of continuum step and accent shape, pooled across listeners, for the early-to-mid continuum. At any given continuum step, the peak is aligned at the plateau's onset.*

Interestingly, however, the magnitude of the change in fit is greater than in the analysis above. Shape, in other words, is doing more work in a model comparing plateaux with sharp peaks timed at their onsets. This pattern is mirrored in the graphs; the magnitude of the difference between the lines representing the two accent shapes is clearly greater in the onset- or rise-based analysis than it is in the offset- or fall-based analysis. Under the assumption that it makes sense to talk about these patterns in terms of a unique peak analog point within plateaux-shaped accents that is stable across contexts (a view which we repudiate below), the results so far indicate that, at least for the early-to-mid continuum in English, this point

9 In this case, we take as a standard the original numbers (1-11) assigned to sharp peaks. Numbers less than 1 assigned to the first three steps of the plateau continuum reflect the fact that our study does not include sharp peaks aligned with the rises of the earliest three plateaux.

lies somewhere between plateau onset and plateau offset, a bit closer to the latter than to the former. This is similar to what D'Imperio originally reported for peaks and plateaux in Neapolitan Italian.[10]

3.2.2. Mid-to-late continuum

Results for the mid-to-late continuum tell a similar story, but are suggestively different as well. Figure 11 first gives the percentage of "late" (L*+H) timing judgments averaged across listeners as a function of continuum step and shape. As before, continuum steps are numbered so that plateaux share step numbers with sharp peaks timed at their offsets. It is immediately clear that the difference in the lines for the two accent shapes is substantially greater than it was for the early-to-mid continuum. Here, the line representing sharp peaks crosses the 50% threshold far earlier than that for the plateaux with analogously timed offsets (falls), suggesting that these plateaux sound much "earlier" than the peaks.

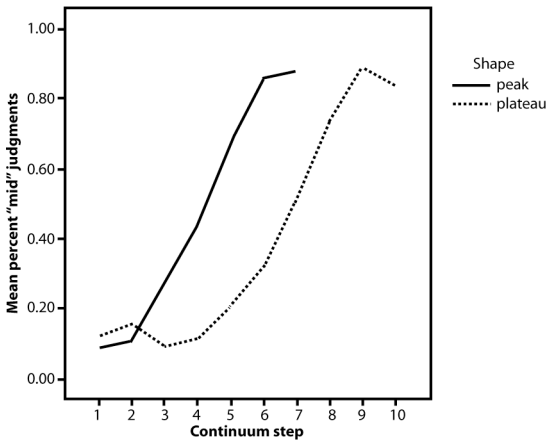

Figure 11: Mean percent L+H judgments as a function of continuum step and accent shape, pooled across listeners, for the mid-to-late continuum. At any given continuum step, the peak is aligned at the plateau's offset.*

Logistic regression analysis again bears this out: the forward stepwise (LR) model selects continuum step as its first predictor (Chi-square (1) = 483.163, $p < .001$, Nagelkerke $r2 = .261$). In its second step, the model adds accent shape as a predictor (Chi-square (2) = 696.469, $p < .001$, Nagelkerke $r2 = .359$). Once again the difference in fit between the two

10 And different from what D'Imperio, et al. 2010 and Gili-Fivela and D'Imperio 2010 report for Pisan Italian, likely for the reasons detailed above in footnote 4.

models is highly significant (Change in -2 Log Likelihood = 213.306, p < .001). Furthermore, both -2 Log Likelihood and the change in Nagelkerke *r2* support the impression given by the graph, i.e. that the difference in perceived timing between peaks and plateaux is much greater in this instance than in any of the preceding analyses.

Recasting these results to reflect a comparison between plateaux and peaks timed at plateau onset (i.e. coinciding rises, rather than falls), the picture that emerges in Figure 12 is somewhat less clear than those we have seen so far. The lines representing peaks and plateaux do differ somewhat, in that the plateau line does cross the 50% threshold marginally earlier than does the peak line. This suggests that plateaux sounded slightly later to listeners than did sharp peaks timed at plateau onset. Logistic regression supports this conclusion: the forward stepwise (LR) model again selects continuum step as its first predictor (Chi-square (1) = 895.509, $p < .001$, Nagelkerke *r2* = .429). In its second step, the model adds again accent shape as a predictor (Chi-square (2) = 902.334, $p < .001$, Nagelkerke *r2* = .432). The difference in fit between the two models remains highly significant (Change in -2 Log Likelihood = 6.836, p = .009), but, though significant, note that it is also extremely small. The change in Nagelkerke *r2* is likewise tiny, just as the difference in response patterns shown in Figure 12 would suggest.

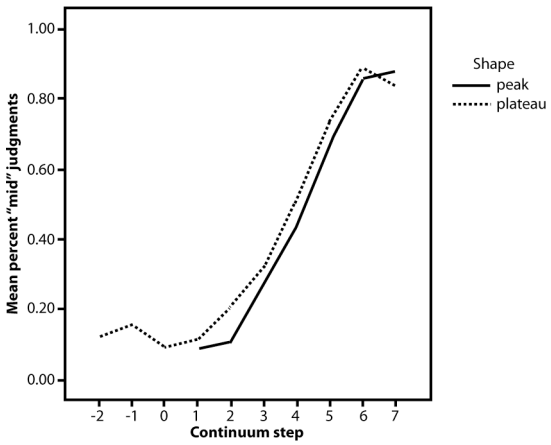

Figure 12: Mean percent L+H judgments as a function of continuum step and accent shape, pooled across listeners, for the mid-to-late continuum. At any given continuum step, the peak is aligned at the plateau's onset.*

If there is an analog to the sharp peak somewhere in the plateau, these data from the mid-to-late continuum again support the conclusion that this point falls somewhere between plateau onset and plateau offset. But this time, in contrast to what we saw with the early-to-mid continuum, the

analog point is extremely close to the onset of the plateau. This is somewhat in line with what our informal observations of English L+H* and L*+H suggested, but precisely the opposite of what is reported by D'Imperio, et al. 2010, and Gili-Fivela and D'Imperio 2010 for the Italian contrasts.

3.3. Discussion.

Taken together, results from the perception of peaks and plateaux in the early-to-mid and mid-to-late continua seem to undermine the notion of a crosslinguistically stable pattern of timing perception in plateau-shaped accents for English and Neapolitan Italian. Moreover the difference in the two English alignment continua shows that even within a language, there may be no single, stable way of relating the perceived timing of plateaux to that of sharp peaks.

This does not mean, however, that D'Imperio, et al. (2010) (or Knight (2008), and others) are wrong in attributing the perceptual differences between peaks and plateaux to universal, psychoacoustic mechanisms. Indeed, we think such mechanisms are precisely the source of these perceptual differences. The problem, we believe, lies in the very notion of a single "peak analog" point within plateau-shaped accents that remains stable across different patterns of alignment with respect to the accented syllable, i.e. in the assumption (shared by TCoG 1.0) that this reference location can be derived without referring to the segmental string upon which the tones in question are realized.

As noted above, we believe this problem can be addressed by modifying TCoG to weight samples taken within the integration window not just by their measured F0, but according to their position relative to the segmental string as well. This modification seems capable of accounting for precisely the differences in timing perception we have documented above in English. To illustrate, let us assume a very simple version of such a model: one in which samples taken from anywhere within the accented vowel are weighted equally heavily, and samples taken from outside the accented vowel are severely discounted.[11]

11 The simple approach we adopt here is reminiscent in some ways of Kohler's (1987, *et seq.*) view of pitch-accent timing contrasts, in which the location of the rise-fall complex relative to the accented vowel in particular is of paramount importance. More complex versions of such a model might incorporate the notion of a "sweet spot" or spots within the accented vowel (possibly its final third, give or take), within which weights would be at their maximum, and outside of which they would decline gradually, even within the vowel itself. Weighting by the sonority of the segments involved is another possibility that we have been investigating. This approach would introduce differences between the accented vowel

To understand how this step would affect the perception of our two English alignment continua, we must think a bit more about the results we and D'Imperio, et al. have amassed. Because of the sigmoid shape of the response curves, with near-floor/ceiling effects at continuum edges, the most reliable information about perceived timing differences comes from the stimuli located in the vicinity of the category boundary. And in fact, comparison of peak and plateau-shaped accents from this critical region begins to make sense, in light of the modified procedure for TCoG sample-weighting sketched above. For example, beginning with the early-to-mid continuum, and taking sharp peak accents as a baseline, Figure 9 above suggests that the category boundary (i.e. the 50% crossover point between H+!H* and L+H* responses) must lie somewhere between steps 5 and 6 of our alignment continuum. Figure 13 below compares the sharp peak aligned at continuum step 5 to a plateau timed such that its rise coincides with that sharp peak's rise.

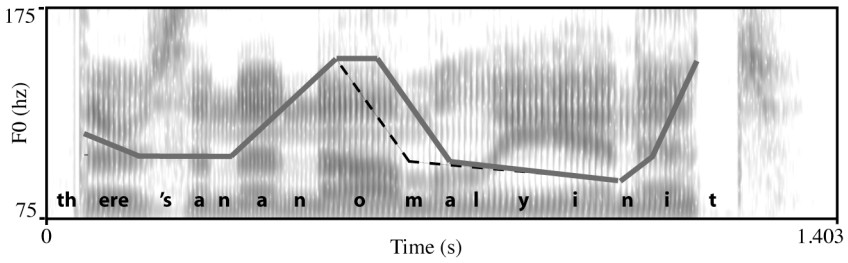

Figure 13: Early-to-mid sharp peak at continuum step 5, and plateau with the same rise.

As can be seen, the sharp peak at step 5 of the continuum rises to an F0 maximum somewhat inside the onset of the accented vowel and falls over most of the vowel. Contrast this with the plateau with a coincident rise. The two accents are identical up to onset of the plateau, but thereafter all of the extra high F0 "bulk" found in the plateau vs. the peak is concentrated within the vowel. Our new system of sample weighting will place comparatively heavy weights on this region, so that the TCoG of the plateau will be pulled dramatically later in comparison to the TCoG of the peak. This comparatively late TCoG should make the plateau sound quite different in terms of its timing from the sharp peak timed at its onset, and this is exactly what we found.

---

and surrounding consonants, as well as, potentially, preceding or following unaccented material (see, e.g., Parker 2002), and is particularly interesting in light of Niebuhr's (2006, 2007b) ideas concerning the contribution of segmental intensity to pitch accent perception.

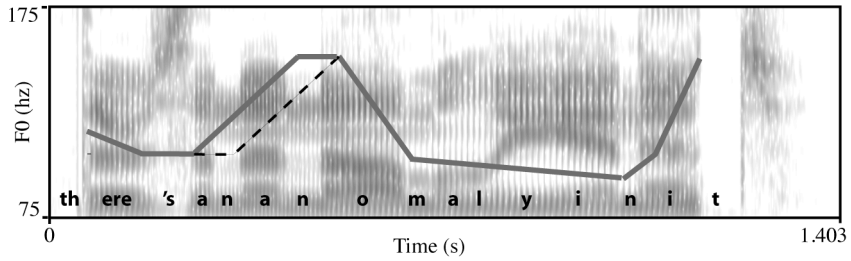

Figure 14: Early-to-mid sharp peak at continuum step 5, and plateau with the same fall.

Contrast this now with the two accents pictured in Figure 14: the same sharp peak from continuum step 5, now compared with a plateau with a coinciding fall. Again, the sharp peak has a maximum F0 near the onset of the accented vowel. In this case, however, because the sharp peak's F0 maximum is timed at plateau offset, both accents are falling within the heavily-weighted region of the accented vowel. Only a short stretch of the plateau's level high F0 is realized within the accented vowel, differentiating the two shapes in this region. On the other hand, the region within which the two accents do differ substantially actually precedes the pitch-accented vowel, and is therefore predicted to receive lower weights in the calculation of TCoG. TCoG of the plateau-shaped accent would thus be relatively close to that of the sharp peak, so that plateaux realized in the vicinity of the category boundary would sound (in terms of their timing patterns) more like sharp peaks timed at plateau offset, than like sharp peaks timed at plateau onset. This is the essence of the pattern we reported above for the early-to-mid continuum.[12]

12 Note that the pattern in this region, to the extent that the plateau nonetheless clearly sounds earlier to listeners than does the sharp peak timed at plateau offset, seems to conflict with a pattern identified in a similar experiment involving German peaks and plateaux by Niebuhr (forthcoming). In Niebuhr's study, a plateau, the offset of which is timed around the category boundary for early vs. medial accent alignment (or H+L* vs. L+H* in GToBI), actually increases listener judgments of medial alignment relative to the sharp peak, rather than decreasing them, as we observe here. Niebuhr attributes this pattern to the enhancement effect that the long high region of the plateau has on the salience or prominence of the H tone, an enhancement that mirrors what is achieved by aligning an accent's high F0 region with the higher sonority pitch periods of the accented vowel, much as Niebuhr's contrast theory of pitch-accent perception predicts it should (Niebuhr 2007b). The contrast theory seems both to embody many of the same intuitions that have driven the development of TCoG, and to differ from TCoG in terms of the predictions it makes in ways that we find fascinating, and worthy of further investigation. It is so far unclear to us why the results obtained here appear to differ from those of Niebuhr (forthcoming) to the extent that they do.

For the less sigmoid response pattern of the mid-to-late continuum, the situation is in some ways even simpler. Because the continuum proceeds from a canonical sounding alignment for English L+H* to a similarly canonical L*+H, the earliest continuum step for the sharp-peak-shaped accents rises throughout the duration of the accented vowel, reaching a peak just after the boundary between that vowel and the following consonant. Comparing this with a plateau whose rise coincides with that of the first sharp peak, as in Figure 15, we see that the portions of the two accents overlapping the accented vowel are perfectly identical, with all of the bulk of the plateau falling outside the accented vowel. Assuming again that F0 samples from within the accented vowel are weighted heavily, while those following it are severely discounted, we would expect the extra bulk of the plateau to exert comparatively little influence on the location of that accent's TCoG, so it should sound, in terms of its timing, similar to the sharp peak aligned at its onset, and all successive steps will continue this pattern. Thus, the close perceptual resemblance between plateaux and peaks timed at plateau onset in this continuum, illustrated in the results reported above, falls out quite naturally from the modified TCoG weighting scheme.

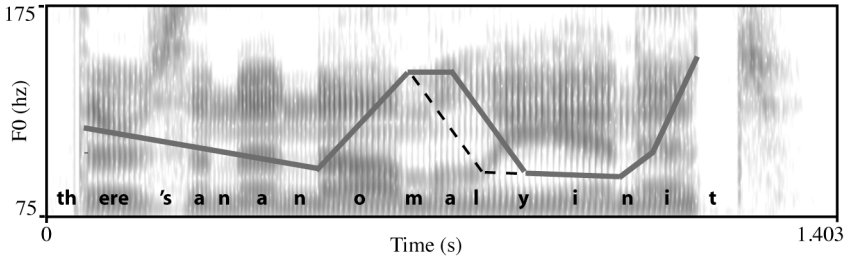

Figure 15: Mid-to-late sharp peak at continuum step 1,and plateau with the same rise.

Finally, Figure 16 compares that same sharp peak at step 1 of the mid-to-late continuum to a plateau timed such that the falls of the two accents align with one another. Here, both the entire flat top the plateau-shaped accent, and most of its rise, takes place within the pitch-accented vowel. The sharp peak, by contrast, only rises during this interval, beginning at its minimum F0. The prediction then, given the heavy weights applied to samples from this region, is that these two accents should sound extremely dissimilar in timing. Since both of these tokens were identified as L+H* nearly 100% of the time, there is little we can conclude from listener responses concerning the details of their perceived timing. However, as we begin to move along the continuum, the results reported above show that the frequency with which sharp peaks are identified as L*+H begins to rise almost immediately as they are shifted

later with respect to the pitch-accented vowel, while plateaux linger for several more steps at essentially floor levels before finally beginning to rise, reaching their own crossover point long after the sharp peaks. Again, this is to be expected, since as we move along the alignment continuum, the high F0 bulk of the plateau will remain within the accented vowel, assuring TCoG alignments that stay comparatively early relative to those for analogously fall-aligned sharp peaks. Throughout this continuum, then, plateaux should sound very little like peaks aligned at plateau offset, which is precisely the pattern we document in English.

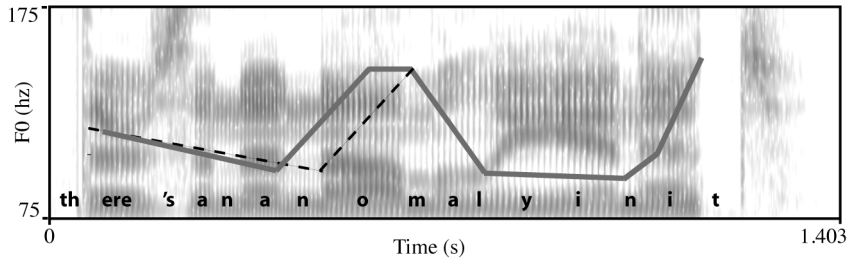

Figure 16: Mid-to-late sharp peak at continuum step 1, and plateau with the same fall.

4. Conclusion

A model based on this augmented weighting-by-segmental-affiliation approach to TCoG, we believe, can account for perceptual differences not only of the kind we have documented, i.e. between peaks and plateaux in the context of different alignment patterns within a single language, but also for the differences reported between English and other languages, like Neapolitan Italian.[13] One consequence of weighting F0 samples by position relative to the accented syllable, however, is that it now becomes impossible to identify any single point within plateau-shaped accents (as a class) that should perform across the board as a "peak analog". Rather, the perceptually relevant point, i.e. the Tonal Center of Gravity, changes its location within the plateau as a function of its alignment with respect to the accented vowel. Put another way, as the plateau is shifted through a continuum of alignments as in the experiments described above, TCoG

13 To the extent that the Neapolitan contrast between phonological L+H* and L*+H is actually in phonetic terms a bit more like the English contrast between H+!H* and L+H* (meaning, the general disposition of the starred tones with respect to the accented vowel are earlier rather than later) it is not surprising that the results D'Imperio (2000) reports for peaks and plateaux along a continuum in that region should be similar to what we find in English for H+!H* and L+H* as well.

also shifts its position within the plateau. (The same should occur for sharp peaks as they are shifted through a continuum of alignments, although because their high region is shorter, the differences in relative TCoG location should be smaller.)

Indeed, in some ways this may be the central lesson of the perceptual differences documented between peaks and plateaux in phonologically analogous accents of English and Neapolitan, as well as between different pairs of phonologically contrasting accents within a single dialect of English: Without attention to the details of how a given pitch accent is realized phonetically, in terms, for example, of its timing with respect to the pitch-accented syllable, it is impossible to state *a priori* what the effect of contour shape differences such as those between sharp peaks and plateaux ought to be on the perception of the timing and scaling of the accents in question. Beyond this, additional factors that may influence both the timing and scaling of TCoG, such as plateau duration, will also need to be considered. (Here we find points of contact with many of the ideas pursued in Niebuhr's contrast theory, as well as in other works that appreciate the complex interplay of timing and scaling in intonational perception, such as Segerup and Nolan (2006), or Gussenhoven (2004)). In light of all this, we believe a TCoG model, revised in the direction indicated above to provide different perceptual weights to samples taken from different portions of the speech signal, is best positioned to account for the intricately patterned perceptual differences both within and across languages that are currently being uncovered through careful experimental work.

Acknowledgments.

We gratefully acknowledge the support of NSF grants 1023853, 1023954, and 1023596. We are also gratefully to Oliver Niebuhr for feedback and editorial assistance with this paper.

References.

Arvaniti A, Ladd DR, & Mennen I. 1998. Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* **26**(1): 3-25.

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. *Proceedings of Speech Prosody 2010*.

Barnes J, Veilleux N, Shattuck-Hufnagel S, & Brugos A. Tonal Center of Gravity. *In*

prep.
Boersma P, Weenink D. *Praat: doing phonetics by computer*. Available at: http://www.praat.org [Accessed May 1, 2009].
d'Alessandro, C., Rosset, S., & Rossi, J-P. 1998. The pitch of short-duration fundamental frequency glissandos. *Journal of the Acoustical Society of America* **104**(4): 2339-2348.
D'Imperio M, Gili-Fivela B, & Niebuhr O. 2010. Alignment perception of high intonational plateaux in Italian and German. *Proceedings of Speech Prosody 2010*.
D'Imperio, M. 2000. The Role of Perception in Defining Tonal Targets and their Alignment. Ph.D. Dissertation, The Ohio State University.
Flemming E. The grammar of coarticulation. *To appear*. In: Embarki M, Dodane C, eds. *La Coarticulation: Indices, Direction et Representation*.
Gili-Fivela B & D'Imperio M. 2010. High peaks versus high plateaux in the identification of two pitch accents in Pisa Italian. *Proceedings of Speech Prosody 2010*.
Gordon M. 2001. A typology of contour tone restrictions. *Studies in Language* 25: 405–444.
Gussenhoven, Carlos. *The Phonology of Tone and Intonation*. Cambridge University Press, 2004.
House D. 1990.*Tonal perception in speech*. Lund, Sweden: Lund University Press.
Knight R. 2003. Peaks and plateaux: the production and perception of intonational high targets in English. Ph.D. Dissertation, University of Cambridge.
Knight R. 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.
Knight R & Nolan F. 2006. The Effect of Pitch Span on Intonational Plateaux. *Journal of the International Phonetic Association* **36**(01): 21-38.
Kohler KJ. 1987. Categorical pitch perception. *Proceedings of the 11th International Congress of Phonetic Sciences*. Vol 5. Tallinn: 331-333.
Ladd DR. 1996/2008. *Intonational Phonology*. 2nd ed. Cambridge University Press.
Ladd DR & Schepman A. 2003. "Sagging transitions" between high pitch accents in English: experimental evidence. *Journal of Phonetics* **31**(1): 81-112.
Ladd DR, Faulkner D, Faulkner H, & Schepman A. 1999. Constant "segmental anchoring" of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America* **106**(3): 1543-1554.
Niebuhr, Oliver. 2006. "The role of the accented-vowel onset in the perception of German early and medial peaks." In *Proceedings of the 3rd International Conference of Speech Prosody*, 109-112. Dresden, Germany.
Niebuhr O. 2007a. The Signalling of German Rising-Falling Intonation Categories – The Interplay of Synchronization, Shape, and Height. *Phonetica* **64**(2-3): 174-193.
Niebuhr, O., 2007b. *Perzeption und kognitive Verarbeitung der Sprechmelodie. Theoretische Grundlagen und empirische Untersuchungen*. Berlin/New York: Mouton de Gruyter.
Niebuhr, O. Forthcoming. Alignment and pitch-accent identification – implications from F0 peak and plateau contours. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung*.
Parker S. 2002. Quantifying the sonority hierarchy. Ph.D. Dissertation, University of Massachusetts, Amherst.
Pierrehumbert J. 1980. The Phonetics and Phonology of English Intonation. Ph.D. Dissertation, MIT.

Segerup, My, and Francis Nolan. "Gothenburg Swedish word accents: a case of cue trading?." In *Nordic Prosody: Proceedings of the IXth Conference*, edited by Gösta Bruce and Merle Horne, 225-233. Frankfurt am Main, Germany: Peter Lang, 2006.

't Hart J. 1991. F0 stylization in speech: straight lines versus parabolas. *Journal of the Acoustical Society of America* **90**(6): 3368-3370.

Veilleux N, Barnes J, Shattuck-Hufnagel S, & Brugos A. 2009. Perceptual robustness of the Tonal Center of Gravity for contour classification. Poster presented at the Workshop on Prosody and Meaning, Barcelona, Spain.

Zhang J. 2001. The effects of duration and sonority on contour tone distribution: Typological survey and formal analysis. Ph.D. Dissertation, UCLA.