

Dialectal variation in the rising accents of American English

Amalia Arvaniti and Gina Garding¹

Abstract

In Pierrehumbert (1980) three rising accents are posited for English, H*, L+H* and L*+H, but this three-way contrast has often been disputed. In the present experiment, Minnesotan and Southern Californian speakers read two dialogs which included four levels of emphasis. Emphasis lengthened segmental duration, raised the scaling of all tones and delayed H alignment in L+H*. Independently of emphasis effects, in both dialects L+H* and L*+H were used in different contexts and were phonetically distinct, with L*+H showing later alignment of both tones, lower scaling of L, and higher scaling of H. In addition, there were phonetic differences between the two dialects, with Southern Californian showing later alignment than Minnesotan English. On the other hand, our data indicate that Minnesotan speakers may lack the H* : L+H* contrast: in one of the dialogs, Minnesotan speakers used L+H* irrespective of emphasis, while Southern Californian speakers switched from H* to L+H* with increased emphasis. These results suggest that Pierrehumbert's three-way distinction between H*, L*+H and L+H* is essentially correct, though it may not apply to all the dialects of American English.

1. Introduction

The high and rising pitch accents of English have been a matter of a great deal of research, speculation and controversy in both traditional frameworks, such as that of the British school (e.g., Halliday 1967), and in more recent research (for a review, see Ladd 1996, and Ladd and Schepman 2003). Within the autosegmental/metrical framework — which is the focus of the present paper — three accents are generally recognized, based on the description proposed by Pierrehumbert (1980): H*, L*+H, and L+H*.² The main difference between H* and the two bitonal accents is that the latter are said to show a rise from a low point, which is lacking in H*. The alignment of this low point, the reflex of the L tone, depends on the accent: according to Pierrehumbert (1980: 77), in L*+H, the L is aligned with the accented syllable and the H trails it, often appearing in the post-accentual syllable. In contrast, in L+H*, the L aligns before the accented syllable, while the H aligns within the accented vowel. The canonical shape of the three accents is schematically presented in Fig. 1.

The meanings associated with this three-way contrast are not discussed in The three pitch accents under investigation have been widely adopted in work on English intonation (among many, Dainora 2002;

Fletcher et al. 2002; Herman and Tevis McGory 2002; Welby 2003). In addition they are part of the tonal inventory of MAE ToBI (Mainstream American English Tones and Break Indices), a labeling standard based on Pierrehumbert (1980) which is used for the prosodic annotation of spoken corpora (Silverman et al. 1992; Beckman and Ayers-Elam 1997; Beckman, Hirschberg and Shattuck-Hufnagel 2005); in addition, they have been used in text-to-speech systems (e.g., Sproat 1998; Jilka, Möhler, and Dogil 1999). H*, L+H* and L*+H are now so widely recognized as accentual categories that what started as a notation representing a three-way accentual contrast in English has been adopted for the intonational analysis of other languages. For instance, the German rising accents, which have late peaks, are analyzed as L*+H in many descriptions of German intonation, even when no L+H* accent is posited (see Grice, Baumann, and Benz Müller 2005 and references therein).

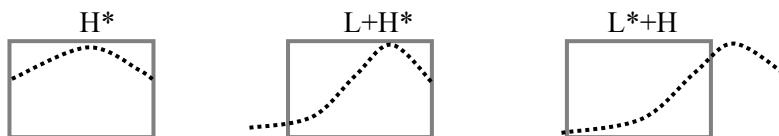


Fig. 1: Schematic representation of the phonetic realization of H*, L+H* and L*+H. The gray boxes represent the extent of the accented syllable; the broken black lines represent the course of F_0 for each accent. It is assumed that all three accents are followed by L- L% phrasal tones, i.e. a fall in pitch.

Despite its widespread recognition, little experimental support for Pierrehumbert's analysis is currently available. Some evidence for the L+H* : H* contrast comes from Dainora's (2002: 119-121) observation that accents transcribed as L+H* and H* in her corpus have different distributions and therefore are likely to be distinct categories in English intonation. The only study of the L*+H : L+H* contrast, on the other hand, is Pierrehumbert and Steele (1989) which did not provide strong evidence in favor of it. Pierrehumbert and Steele used recordings of *Only a millionaire*, uttered with L*+H or with L+H* on *millionaire*, to create a continuum of stimuli with progressively later peaks. The stimuli were presented to five speakers who had to imitate them. The authors hypothesized that if the two accents form distinct categories, the alignment of the valley and the peak in the speakers' imitations would fall into bimodal distributions; if, on the other hand, L*+H and L+H* are the ends of a continuum, the speakers' imitations would show normal distribution. The valley data did not show bimodal distribution and neither did the peak

data of one of the five speakers. The data of the other four speakers showed evidence for bimodal distribution, but this was not very strong except for one speaker (Steele herself). Nevertheless, Pierrehumbert and Steele (1989: 195) conclude that “English speakers have two rise-fall-rise intonation patterns, which differ in how they are aligned with the stressed syllable.”

The three-way contrast between H*, L+H* and L*+H has been challenged on theoretical and empirical grounds, most notably by Ladd (1983). Specifically, Ladd (1983), who does not explicitly refer to L+H*, suggested that L*+H is not a distinct accent but a version of H* that adds emphasis or insistence. Accordingly, he proposed that instead of positing two different accents, a feature [\pm delayed peak] should be used to distinguish between “scooped” ([+d.p.]) and “plain” ([-d.p.]) accents, that is between L*+H and H* respectively. In later work, Ladd acknowledged that “almost certainly” the scooped accents have a L tone, as suggested by Pierrehumbert (1980) and Beckman and Pierrehumbert (1986), and therefore that the distinction between H* and L*+H is valid for English (Ladd 1996: 104). On the other hand, it is not clear whether a principled distinction is made in Ladd (1996) between H* and L+H*. For instance, Ladd (1996: 84–85) discusses the difficulty in distinguishing between H* and L+H* utterance initially, and questions the validity of Pierrehumbert’s analysis that treats the (shallow) F_0 dip before H* as interpolation, but treats the (less shallow) dip in L+H* as the reflex of a tone. Indeed, subsequent work by Ladd and Morton (1997) suggests that Ladd treats L+H* as an emphatic variant of H* (and thus L+H* could possibly be analyzed using the [+raised peak] feature in the framework of Ladd [1983]).

Although the above mentioned works suggest that for Ladd and his colleagues there is only a contrast between H* and L*+H in English, more recently Ladd and Schepman (2003) proposed that it would be more appropriate to analyze most instances of H* in English as (L+H)*, reserving H* only for the rare occurrences of “flat hat” contours, in which there is straight interpolation between successive peaks (p. 105). Ladd and Schepman base this re-analysis on the finding that pre-nuclear accents of the type that would usually be analyzed as H* show a consistent dip before the rise. Such dips between H* accents had been analyzed by Pierrehumbert (1981) as *sagging transitions* which are not expected to have stable scaling and alignment. However, Ladd and Schepman show that the scaling and alignment of these dips are stable and thus consistent with the idea that they are the reflex of a L tone. They also suggest that the difference between Pierrehumbert’s H* and L+H* tones may be “a matter

of gradient variation between two ends of an emphasis-based continuum” (p.106). They further speculate that L+H* may be characterized by later L alignment and greater pitch excursion than H*.

When the various views on the phonetics, phonology and pragmatics of H*, L+H* and L*+H are considered together, it is clear that different pictures emerge: according to Pierrehumbert and colleagues, English has three accents, H*, L+H* and L*+H, which are phonologically, phonetically and pragmatically distinct; L+H* is the accent that is most suited for use with emphasis due to its pragmatics. According to Ladd and his collaborators, it is possible that all these accents represent instances of the same phonological category (H or L+H) which may show later peak alignment and higher scaling with added emphasis. If his analysis is correct, then one would expect accents which most closely fit the phonetic realization of Pierrehumbert’s L*+H category to be used with emphasis, as L*+H has the latest peak alignment.

However, it is also the case that any comparison between the above mentioned studies is fraught with difficulties due to two reasons. First, experimental work has focused either on the scaling of the H under emphasis (e.g., Liberman and Pierrehumbert 1984; Ladd and Morton 1997) or on the alignment of the tones in non-emphatic renditions (e.g., Pierrehumbert and Steele 1989; Ladd and Schepman 2003), although it is clear that both the scaling and the alignment of the tones play a part in this three-way contrast and both are affected by emphasis. For example, the figures of Pierrehumbert and Steele (1989), who manipulated only tonal alignment, indicate that the scaling of the H tone may be higher in L*+H than in L+H*. Similarly, Ladd and Morton (1997), who manipulated only tonal scaling, note that they “made no adjustment for the fact that the peak of the emphatic contours was invariably aligned later in the accented syllable than the peak of the normal contours” (Ladd and Morton 1997: 321–22).

Second, in much of this work, English is treated as a homogeneous language when in fact in most cases the research involved speakers of quite distinct varieties. The fact that the five speakers in Pierrehumbert and Steele (1989) were speakers of American English can only be inferred from their paper. Ladd and Morton (1997) mention that they used “several native speakers of British English” (p. 318), but do not provide any further details. Similarly, Ladd and Schepman (2003) compare their results to Pierrehumbert (1980), although their data come from one English and one Scottish speaker, while Pierrehumbert’s data come from American speakers. Such situations are fairly typical of much research on intonation,

but it is still of interest here that differences related to the linguistic variety spoken by the subjects are not expected to arise and are not taken into consideration when the results of different studies are compared.

In the experiment presented below we try to address some of these shortcomings by examining data from speakers of two distinct varieties of American English – Southern Californian English, one of the Western varieties, and Minnesotan English, a variety from the North Central region (Labov, Ash, and Boberg 1997) – in order to assess what accentual contrasts they have and how these are affected by emphasis. Specifically, the aim of our experiment was to shed light to the following questions:

- Is it possible that some dialects of English lack one of the accentual categories under investigation, while others have all three accents?
- Are H*, L*+H and L+H* realized in the same way across dialects of American English that have this three-way accentual contrast?
- Are all three accents equally good choices when emphasis is needed in discourse?
- If so, what is the effect of emphasis on each accent's realization, and can it tell us anything about the tonal composition of the accents?

On the basis of the above questions, our own observations, and a pilot experiment, we formulated the following hypotheses:

- The H* : L+H* : L*+H contrast may not be present in all dialects of English; for example, Pierrehumbert and Steele (1989: 190) suggest that their speaker D.T.T. “lacks the L*+H pitch accent,” a lack that could pertain to his (unspecified) dialect. Within this context, we chose to investigate Minnesotan and Southern Californian English because of the different reactions that ToBI L*+H examples elicited from speakers of these two varieties during an intonational phonology seminar: the Californian speakers claimed they found these examples odd; in contrast, the second author, who comes from Minnesota, thought they were perfectly acceptable. On the basis of this difference and pilot data, we expected that Minnesotan English was more likely to have the three-way accentual contrast.
- Furthermore, we expected that the same accents could be realized with different phonetics across the two dialects, although we did not have any specific predictions on this point. We based this expectation on recent results showing that there are subtle cross-varietal distinctions

even when the same phonological analysis can be used (e.g., Atterer and Ladd, 2004; Arvaniti, submitted). Indeed, Prieto, D'Imperio and Gili Fivela (to appear) and Arvaniti, Ladd and Mennen (submitted) show that even within the same linguistic variety small differences in tonal alignment may result in pragmatic interpretations that are so distinct they cannot possibly be seen as a continuum of pragmatic meanings.

- Based on the pragmatics of the accents, as discussed in Pierrehumbert and Hirschberg (1990), we expected that L+H* would be more natural in emphatic renditions across dialects. Specifically, we expected that in appropriate contexts speakers would use H* in non-emphatic renditions and switch to L+H* when using emphasis; we also expected that they would have more difficulty using L*+H with emphasis.
- We further expected that emphasis would affect L and H targets differently. Specifically, it has often been noted that Hs tend to rise with emphasis, and thus we expected that all Hs, starred and unstarred, would get higher with increased emphasis (Pierrehumbert 1980; Liberman and Pierrehumbert 1984).
- The effect of emphasis on Ls, on the other hand, is less straightforward and may depend of starredness. Thus, Pierrehumbert (1980: 75) remarks that leading Ls would rise if the scaling of the starred H increases, while starred Ls would get lower.³ Support for this position comes from production and perception data (Liberman and Pierrehumbert 1984, and Gussenhoven and Rietveld 2000, respectively). On the basis of these studies, we expected that starred Ls would get lower with increased emphasis, while unstarred Ls would stay the same or rise. On the other hand, non-tonal dips – that is dips that are not phonetic reflexes of L tones (in the cases where H* was used instead of a bitonal accent) – would be shallower and more variable than L tones proper.
- On the basis of Ladd and Morton's (1997) results, we expected that all Hs would align later with increased emphasis, but had no specific predictions about the alignment of the Ls.
- Finally, we expected that segmental lengthening would accompany increasing emphasis (Klatt 1976; Cummins 1999).

2. Method

2.1 Materials

Two test dialogs were created in which the speakers' part consisted of the same response given at four emphasis levels. In the first dialog (henceforth *phone dialog*), the speakers read the part of a person who answers repeated inquiries as to who was on the phone. This dialogue was expected to elicit answers with a H* L- L% tune without emphasis, and a L+H* L- L% tune with emphasis. In the second dialog (henceforth *date dialog*), the speakers read the part of a sibling trying to convince a brother in a sly, cajoling way that his/her friend Raymona would be the perfect date for an upcoming event. This dialogue was expected to elicit answers with a L*+H L- H% tune, though several speakers used L*+H with H- L% phrasal tones instead (see Fig. 2).⁴ The complete dialogs in the format in which they were presented to the speakers are given in Table 1. As can be seen, boldface font, underlining and capitals were used to indicate increasing levels of emphasis within each dialog.

Table 1. The two test dialogs

	H*/L+H* phone dialog	or	L*+H or date dialog
Investigator:	Who was on the phone?		I need someone who would hold her own.
Speaker:	Raymona.		Raymona would...
Investigator:	Who?		But someone who would understand politics.
Speaker:	Raymona.		Raymona would...
Investigator:	It was who?		But someone who would be really funny, too.
Speaker:	<u>Raymona.</u>		<u>Raymona would...</u>
Investigator:	Raymond?		But someone who would look totally hot!
Speaker:	<u>RAYMONA!</u>		<u>RAYMONA WOULD...!</u>

The name *Raymona* was chosen because its stressed syllable is preceded by a syllable with a full vowel. This ensured that L tones would not be truncated (Beckman and Ayers-Elam 1997; Grice 1995) and if present, they would not be difficult to measure (as would be the case if they aligned with a reduced initial syllable). In addition, the name *Raymona* lent semantic naturalness to the phone dialog, as it is not a common name and was more likely to be questioned or misheard.

Ideally, we would have liked to use *Raymona* as a response to both dialogs, but in our pilots the speakers had difficulty producing the L*+H contour on the word *Raymona* alone. Although it is possible that the

presence of the extra syllable has resulted in less tonal pressure in this dialog, thereby exaggerating the difference between L*+H and L+H*, we believe that this condition could not have significantly affected our results, since *Raymona* has a post-nuclear syllable on which the phrasal tones can be fully realized, as evidenced by the low F_0 reached at the end of *Raymona* (see Fig. 2). In addition, to our knowledge, there is no literature, indicating that L- L% phrasal tones may be truncated in any variety of American English (though truncation is possible in some British varieties, according to Grabe et al. 2000). On the contrary, there is literature suggesting that utterance-final low tones in American English always reach a consistent speaker-specific low pitch level (Menn and Boyce 1982; Liberman and Pierrehumbert 1984).

2.2 Speakers

The results of six speakers of Southern Californian English (henceforth *SoCal*), and seven speakers of Minnesotan English (henceforth *MN*) are reported. To qualify, the speakers had to be born and raised in monolingual families in SoCal or MN, and either still living there, or having lived away for less than 2.5 years. Of the SoCal speakers, two were from Orange County (both females), two were from San Diego (both females), and two were from the coastal part of Los Angeles county (both males). Of the MN speakers, three were from Twin Cities or their suburbs (one male, two females), and the other four were from West Central rural Minnesota (three males, one female). The speakers ranged from 19-29 years of age (except for two MN male speakers who were in their fifties) and were paid \$10 for their participation. They reported no speech or hearing problems. They were all naïve as to the purposes of the experiment and had no linguistic or phonetic training of any sort.

2.3 Procedures

The speakers were presented with 4×6 note cards each of which had either one of the two target dialogs or one of two similarly structured filler dialogs printed on it. The second author explained the context for each dialog, including the fillers, and demonstrated – by reading each dialog aloud once – how to consistently use the same contour within each dialog while increasing the level of emphasis. The speakers were asked to read the dialogs at their normal speaking rate, as if they were talking to a friend. They then listened once to one recorded sequence of their part in each target dialog, in order to hear the increasing levels of emphasis with no

intervening material. These examples were extracted from the data; specifically, the speakers heard one example from another native speaker of their dialect, except for the first SoCal speaker who heard an example from a MN speaker; for obvious reasons, the first MN speaker did not hear recorded examples. This procedure was used for two reasons: first, pilots suggested that our speakers had difficulty producing four distinct levels of emphasis (unlike phonetically trained speakers, such as those used in Liberman and Pierrehumbert [1984] for instance, who could produce ten different levels of emphasis), and we wanted to provide them with an illustration of what increasing emphasis in steps might sound like. On the other hand, we wanted them to hear intonation patterns from their own linguistic variety, since those were more likely to sound natural to them and less likely to influence their own production away from their natural patterns.

All SoCal and three MN speakers were recorded in the UCSD Phonetics Laboratory studio; the other MN speakers were recorded in a quiet room in Minnesota using a DAT recorder. The dialogs were counterbalanced across speakers: half of the speakers read first several repetitions of the phone dialog alternating with a similarly structured filler, while the other half read first several repetitions of the date dialog alternating with a similarly structured filler. On average the speakers read twelve repetitions of each dialog (the exact number depended on their fluency and ease with the task), but some speakers had to read the date dialog a few more times in order to achieve a consistent increase in emphasis, a point to which we return in section 4.2.

Since the speakers did not receive explicit instructions regarding the intonation they should use, except implicitly in the examples they heard, many did not produce the contours we had anticipated or did so inconsistently. In the phone dialog in particular, several speakers, especially those from Southern California, used a rising contour (L* H-H%) for the neutral and lower emphasis levels.⁵ In such cases, the second author demonstrated to them the expected falling contour and asked them to use it *if they felt comfortable doing so*. After this, most speakers did not have problems producing falling contours except occasionally. When this happened, dialogs with rising declaratives at any level of emphasis were discarded. In the same dialog, some speakers had occasionally had to be asked not to “segment” *Raymona* at higher levels of emphasis (RAY-MOna!), that is to avoid two-accent renditions (see Pierrehumbert [1993] on words realized with two pitch accents). In general, the speakers found the task manageable.

2.4 Measurements

All recordings were digitized at 44.1 kHz and 16-bit resolution and were analyzed using Pitchworks. Measurements of duration and F_0 were obtained. The duration measurements were made by simultaneous inspection of the waveform and broadband spectrograms. F_0 was extracted using the Pitchworks cepstral algorithm with a 10 ms. time window. Specifically, the following measurements were taken (see Fig. 2 for an illustration):

- The duration of the segments in the accented syllable, [m] and [ou].
- The scaling of H (in Hz); this was taken to be the highest point reached in the vicinity of the accented syllable; in the case of a series of equally high points (a rare occurrence), the first in the series was selected. Although points were occasionally missing where the F_0 rise was expected, the H was generally not subject to microprosodic perturbations, as it did not occur close to a segmental boundary.
- The alignment of H with respect to the offset of the accented vowel, [ou]; this landmark was chosen over the vowel onset because initial inspection showed that H appeared after the accented vowel, especially in the date dialog.
- The scaling of the L: for female speakers, this was taken to be the point where there began a sustained rise of at least 4 Hz per consecutive pitch point that lasted for at least four points in the pitch track; for male speakers, this was modified to be a sustained rise of at least 3 Hz per point. In very flat contours the definition was modified to be a sustained rise of 1 Hz per point or greater. Such very flat contours were encountered primarily in the phone dialog of the SoCal speakers, a point to which we return in section 4.3. For many speakers, L showed microprosodic perturbations due to the nasal offset; such pitch points were ignored when calculating the increase in F_0 , as illustrated in Fig. 2. Ls were measured whether we expected them to be targets, as in the date dialog, or not, as in the non-emphatic level of the phone dialog.
- The alignment of L with respect to the onset of the accented vowel, [ou]; [ou] chosen over the onset of [m], with which previous descriptions suggest L aligns, because L appeared rather late within the accented syllable.
- The scaling difference and temporal interval between L and H.

Overall, between four to nine repetitions could be used for measurement, depending on the speaker. Repetitions were discarded if the tune used varied across emphasis levels or if emphasis, visible in the pitch track and amplitude of the waveform, did not consistently increase within a dialog. Further, several repetitions could not be used because Ls or (much more rarely) Hs were missing from the pitch track. Missing Ls were particularly prevalent in the data of the male speakers.

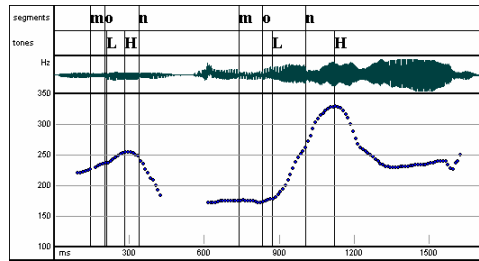


Fig. 2: An illustration of measurements in the phone dialog (left) and the date dialog (right).

2.5 Statistical analysis

The data were analyzed statistically by means of analyses of variance in which DIALECT (MN, SoCal) was the between-subjects factor, and ACCENT (L+H*, L*+H) and EMPHASIS (levels 1–4) the within-subjects factors. Note that we statistically treated the data as if they involved only two categories, L+H* and L*+H, by subsuming the H* tokens under L+H*. We focus on the differences between L+H* and H* in section 4.3. of the discussion. All statistical analyses were based on mean values for each speaker, which in turn were based on five measurements for each dependent variable (when more than five repetitions were available those used to derive the mean were randomly selected; this reduction of the data was used so that all means be based on the same number of observations). Space considerations do not allow for a presentation of individual results, but F0 and duration variances were not notably different across speakers, dialects or dialogs. For the scaling data, the datapoints used for analysis were converted to their natural logarithms in order to normalize the sample which included values from both male and female speakers (cf. Xu 1999). The means and standard errors were re-converted to Hz for illustration purposes (see Fig. 3). Differences between levels of emphasis and expected interactions were examined using planned comparisons; unexpected interactions (e.g., dialect × emphasis × accent) were examined using Tukey HSD tests. Due to space

limitations, interactions that do not affect the validity of the main effects are not discussed. Similarly, it should be noted that the analyses did not show three-way interactions unless otherwise stated.

3. Results

3.1 L scaling

L was scaled lower in L*+H than in L+H* [$F(1,11)=36.020$, $p<0.0001$]. L was also scaled higher with increased emphasis [$F(3,33)=92.3$, $p<0.0001$], and planned comparisons showed that each additional level of emphasis increased its scaling [$p<0.0001$ in all cases]. These effects can be observed in Fig. 3a.

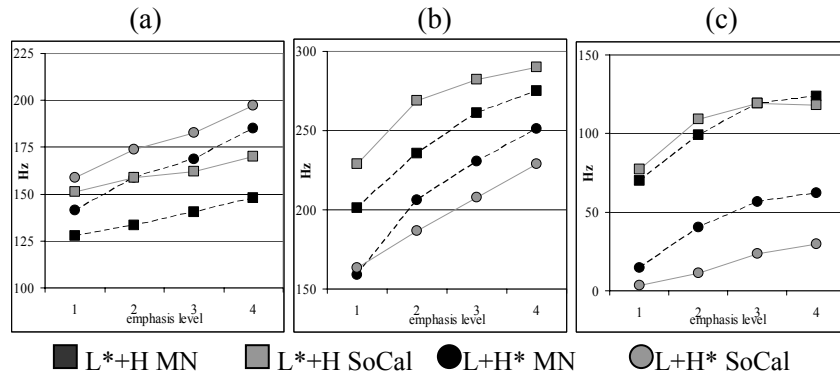


Fig. 3: Means and standard errors (not visible) for L scaling (panel a), H scaling (panel b), and for the scaling difference between L and H (panel c). Data presented as a function of ACCENT and EMPHASIS, and separately for MN and SoCal.

3.2 H scaling

H was scaled higher in L*+H than in L+H* [$F(1,11)=36.5$, $p<0.0001$]. In addition, H scaling was also affected by EMPHASIS [$F(3,33)=128.9$, $p<0.0001$], and planned comparisons showed that each additional level of emphasis increased the scaling of H [$p<0.001$ for all comparisons]. These effects can be observed in Fig. 3b.

3.3 Scaling difference between L and H

The scaling difference between L and H was smaller in L+H* than in L*+H [$F(1,11)=107.7$, $p<0.0001$], and increased with added emphasis [$F(3,33)=100.05$, $p<0.0001$]. However, the data also showed a three-way

interaction [$F(3,33)=3.3$, $p<0.03$], illustrated in Fig. 3c. Tukey HSD tests showed that the difference between L+H* and L*+H holds across all emphasis levels and both dialects, but the emphasis effects are not consistent across accents and dialects. In MN, the emphasis effect is very small in both accents: for L*+H, level 1 of emphasis shows a smaller scaling difference than levels 3 and 4 [$p<0.05$ for level 1 vs. 3; $p<0.02$ for level 1 vs. 4], but all other pair-wise comparisons are not significant; for L+H*, level 1 shows a smaller scaling difference than the other levels [$p<0.0001$ in all cases], but all other comparisons are not significant. In SoCal, there are no differences across emphasis levels of L*+H; for L+H*, in contrast, the scaling difference increased at each emphasis level, except for level 3 vs. 4 [for level 2 vs. 3, $p<0.003$; for all other comparisons, $p<0.001$; level 3 vs. 4, n.s.].

3.4 L alignment

As can be seen in Fig. 4a, L was aligned earlier with respect to the onset of the accented vowel in MN than in SoCal [$F(1,11)=7.9$, $p<0.02$]. L also aligned later in L*+H than in L+H* [$F(1,11)=31.4$, $p<0.0001$], occurring after the onset of [ou] in L*+H, but before it in L+H*. There was no effect of EMPHASIS or any statistically significant interactions.

3.5 H alignment

H showed earlier alignment with respect to the offset of the accented vowel [ou] in L+H* than in L*+H [$F(1,11)=187.02$, $p<0.0001$], aligning after the vowel in L*+H but within it in L+H*. The alignment of H was affected by EMPHASIS as well [$F(3,33)=12.4$, $p<0.0001$], but also showed interactions between DIALECT and ACCENT [$F(1,11)=9.98$, $p<0.009$] and between ACCENT and EMPHASIS [$F(3,33)=12.55$, $p<0.0001$] (see Fig. 4b). The investigation of these interactions showed first, that both SoCal and MN speakers aligned H later in L*+H than in L+H*, confirming the ACCENT main effect. In addition, the two groups showed similar alignment of the H in L+H*; in L*+H, however, the SoCal speakers showed much later H alignment than the MN speakers [$p<0.03$]. In addition, the two accents showed different effects of EMPHASIS: the H of L+H* aligned increasingly later with added emphasis [$p<0.008$ at least for all comparisons, except for level 3 vs. 4, n.s.]; in L*+H, on the other hand, most pair-wise comparisons did not reach statistical significance, with the exception of level 4 which showed earlier alignment than levels 2 [$p<0.003$] and 3 [$p<0.02$] for both dialects.

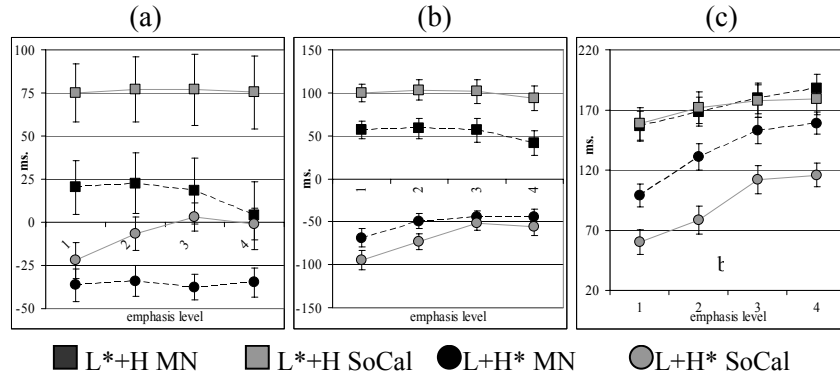


Fig. 4: Means and standard errors of L alignment (panel a), of H alignment (panel b), and of the temporal interval between L and H (panel c). In panel (a), 0 on the y axis represents the onset of [ou], and negative values represent alignment during the preceding [m]. In panel (b) 0 on the y axis represents the onset of [n], and negative values represent alignment during the preceding [ou]. All data presented as a function of ACCENT and EMPHASIS, and separately for MN and SoCal.

3.6 The temporal interval between L and H

The temporal interval between L and H was affected by ACCENT [$F(1,11)=36.2$, $p<0.0001$] and EMPHASIS [$F(3,33)=51.8$, $p<0.0001$], but there was also interaction between EMPHASIS and ACCENT [$F(3,33)=8.7$, $p<0.0002$]. The investigation of this interaction showed that the interval between L and H was consistently longer in L*+H than in L+H*, confirming the main effect of ACCENT; on the other hand, the EMPHASIS effect was less consistent in L*+H than in L+H*: in L*+H the difference between levels 2 and 3 and between levels 3 and 4 was not significant, while this was the case only for level 3 vs. 4 in L+H* [$p<0.001$ for all other pair-wise comparisons]. These effects can be observed in Fig. 4c.

3.7 Segmental durations and proportional alignment of L and H targets

As mentioned, in addition to effects on the scaling and alignment of the tones, we expected to find emphasis effects on the duration of segments. Further, we wanted to see whether there were any differences in duration across dialects. Such differences would cast doubt on the later alignment observed in the SoCal data, as they would make it possible to argue that this is not a difference in tonal realization per se, but it is, trivially, the result of generally longer durations in this dialect (see Silverman and Pierrehumbert 1990 for a discussion of absolute vs. proportional measures

of tonal alignment). Similarly, any delay in tonal alignment under emphasis could be attributed to longer segmental durations.

Analyses of variance on [m], [ou] and syllable ([m]+[ou]) duration showed increasing lengthening under levels of EMPHASIS, with [ou] showing the most consistent increase [for [ou], $F(3,3)=69.4$, $p<0.0001$; $p<0.0001$ for all pair-wise comparisons; the results for consonant and syllable durations were similar]. On the other hand, none of the three intervals showed any statistically significant effects of ACCENT or DIALECT, suggesting that any alignment differences across accents and dialects reflect real distinctions and are not an epiphenomenon of durational effects. The validity of these alignment differences and also of the effects of emphasis on alignment – to the extent they were present – was confirmed by the results of analyses of variance on the proportional alignment of H and L (distance of H from the onset of [ou] divided by the duration of [ou], and distance of L from the onset of [m] divided by the duration of [m]); these analyses yielded the same pattern of results as the analyses based on absolute target alignment.⁶

4. Discussion

4.1 The L*+H : L+H* contrast

In brief our results showed that the speakers used two L+H accents with distinct phonetics, such that in one accent both tones aligned early with respect to the accented syllable, while in the other they aligned late. In addition, the L and H tones generally aligned later in SoCal than in MN. Further, previously unreported differences were found in the scaling of the two bitonal accents: the L of L+H* was scaled higher than the L of L*+H, while the reverse holds for the H tones. Emphasis increased the scaling of both Ls and Hs and also increased the duration of the accented vowel (and to a lesser extent of the consonant and the entire syllable). Finally, emphasis delayed the alignment of H in L+H*, but had little effect on the alignment of H in L*+H, and no effect on the alignment of the L in either accent.

Taken all together these results provide clear answers to some of the questions we set out to investigate. First, our hypothesis that SoCal speakers lack this contrast was not borne out. Both our MN and our SoCal speakers consistently used the expected accent in each of the two test dialogs we constructed, suggesting that L+H* and L*+H are not extreme points in a continuum, but distinct phonological categories with distinct

pragmatics; if this were not the case, we should have found much greater variability across accent categories and small or no statistically significant differences between the two. In contrast, the phonetic differences between the L*+H and L+H* are very robust and leave no room for confusion between the two accents. Thus our results support Pierrehumbert's original contention that American English has two distinct L+H accents, L*+H and L+H*.

On the other hand, we did find differences between SoCal and MN regarding the phonetic realization of the accents, with the SoCal speakers showing later alignment of the targets overall (with the exception of H in the phone dialog). Such small but significant differences between related linguistic varieties are expected when discussing segmentals, but have just begun to be investigated in intonation (see e.g., Atterer and Ladd [2004] on peak alignment in varieties of German, and Arvaniti [submitted] on differences between Mid-Western American English and British RP in the realization of final lowering and declination). This type of research, including the present study, shows that it is important to look at phonetic details of tonal target scaling and alignment, rather than assume that intonational categories will be realized in the same way in a wide spectrum of related linguistic varieties because the same phonological analysis may apply to all of them.

Similarly our data differ somewhat in phonetic detail from those of previous studies (Pierrehumbert 1980; Beckman and Pierrehumbert 1986; Pierrehumbert and Steele 1989). In our L+H* data, H aligns roughly in the middle of the accented vowel (on average at 60% of the vowel's duration), as suggested in those studies, but the L appears within the accented syllable (typically during the consonantal onset), not in the syllable preceding the accent as Pierrehumbert (1980), Beckman and Pierrehumbert (1986), and Pierrehumbert and Steele (1989) suggest. Regarding L*+H, our data agree with previous studies, showing L alignment with the accented syllable (in fact with the accented vowel), and H alignment with the post-nuclear syllable. However, our L*+H data also differ somewhat from those reported in Pierrehumbert (1980) regarding the temporal distance between L and H: our data show an average distance of 172 ms., while Pierrehumbert (1980: 77) reports a distance of 200 ms.⁷ This discrepancy between the older studies and ours could again be due to dialectal differences, similar to those observed here between our SoCal and MN speakers.

In addition, our data showed differences in tonal scaling which are not easy to accommodate within the autosegmental-metrical model of

intonational phonology. Specifically, the higher scaling of the unstarred L in L+H* when compared to L*+H is predicted in Pierrehumbert (1980), who suggests that starred H tones in bitonal accents raise the scaling of associated L tones, but starred Ls are not affected by H tones associated with them. However, in our data the H of L+H* was also scaled lower than the H of L*+H, and this difference cannot be easily accounted for. A possible explanation could be the somewhat greater tonal crowding in our L+H* dialog, but this seems unlikely in light of the fact that similar differences in scaling are apparent (though not quantified) in Pierrehumbert and Steele (1989) where tonal crowding was not an issue. It is possible that this difference in the scaling of the two H tones is somehow related to the marked status of the L*+H, a point to which we turn.

4.2 The marked status of L*+H

Although our speakers clearly distinguished L*+H from L+H* in their productions, it is also the case that they found the date dialog more challenging than the phone dialog. First, as mentioned, pilot data showed that speakers could only produce L*+H in a longer phrase. In addition, the speakers varied in their choice of phrasal tones, some using L- H% while others used H- L%, a configuration that is phonetically realized as sustained mid-level pitch described in Ladd (1978) as *stylized*. These observations together testify to the marked status of L*+H. This was first noted in Pierrehumbert and Steele (1989:188), and has since been supported by the rarity of this accent reported in Dainora (2002: 111). Although Dianora's results should be treated with caution, since they are based on the professional speaking style of two female news readers, they do suggest that L*+H is much rarer than L+H* and H*: according to her corpus, the probability of H* occurring as the final accent in an utterance is 60% (the probability of the entire configuration in our data, H* L- L%, is 33%), the probability of L+H* is 25% (the probability of L+H* L- L% is 9%), while that of L*+H is 0.5% (the probability of L*+H L- H% is 0.02% and that of L*+H H- L% is 0.09%). The relative rarity of L*+H may partly explain its more exaggerated scaling and tonal alignment, although this is clearly a point that requires further research preferably with a varied corpus of natural data.

4.3 The L+H* : H* contrast

As mentioned in the introduction, one of our hypotheses was that neutral renditions of the phone dialog, that is those at level 1 of emphasis, would

show use of the H* accent, rather than L+H* and therefore a smaller scaling difference between L and H than the other levels, and possibly a less stable L. This was partially borne out by the data, as shown in sections 3.3 and 3.6, although any differences related mostly to changes in the alignment and scaling of the H under emphasis, while the L appeared to be consistently scaled and aligned.

However, this general conclusion regarding the behavior of Ls does apply to all of our data. During labeling it became apparent that there were substantial differences between the MN and SoCal speakers regarding the realization of the accent in the non-emphatic level of the phone dialog. In the MN data it was easy to locate a L target even in the non-emphatic renditions (see e.g., Fig. 2). In contrast, in the SoCal data, identifying a L proved much more problematic because in many cases the contours at level 1 of the phone dialog were very flat (see e.g., Fig. 9), while L was more shallow and variable than the L in the MN data. This is an effect we had expected to see in the non-emphatic levels of the phone dialog in both dialects, but it was present only in the SoCal data. It is also revealing that the SoCal speakers switched contours depending on emphasis: in non-emphatic answers in the phone dialog they used flat or falling contours or resorted to rising intonation (L* H- H%); when their speech became more emphatic, however, they always switched to falling intonation (L- L%) preceded by an accent with an easily identifiable initial dip, i.e. to an accent that fits the L+H* description. In contrast, the MN speakers used the same rise-fall contour in all levels of emphasis in the phone dialog, increasing tonal scaling and delaying the alignment of H for emphasis.

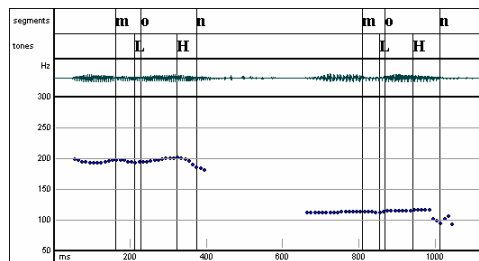


Fig. 4: Typical non-emphatic contours from the phone dialog as realized by two SoCal speakers, a female (left) and a male (right).

These impressionistic differences were confirmed statistically by a series of one-way analyses of variance with DIALECT as a between-subjects factor and the measurements from level 1 of emphasis in the phone dialog as the dependent variables. The results showed that the scaling difference

between L and H was larger for MN than SoCal [$F(1,11)=10.35$, $p<0.008$; MN $\bar{X}=14$ Hz, s.d.=2; SC, $\bar{X}=6$ Hz, s.d.=2], and so was the distance between these two targets [$F(1,11)=7.72$, $p<0.02$; MN $\bar{X}=99$ ms., s.d.=33; SC, $\bar{X}=60$ ms., s.d.=10]. These differences contrast with the level 1 data of the date dialog which did not show similar differences between the two dialects.

These results suggest that while the contrast between L+H* and H* is present in SoCal, it may be absent from MN. This suggestion finds support in the general impression that Minnesotans speak with a “sing-song” intonation (Jeff Good p.c.; Ben Munson p.c.). However, it is also possible that the context of the phone dialog was not appropriate for H* in their intonational system. Although this possibility should be investigated further, a system without a plain H* would be in line with recent results from other dialects of English, such as those of Ladd and Schepman (2003) who show that in British English, accents typically analyzed as H* have a consistent dip preceding the peak. The lack of a contrast in some dialects could also offer a partial explanation for the fact that H* and L+H* are considered similar and are often confused in MAE ToBI transcriptions (Ladd 1996; Herman and Tevis McGory 2002).

4.4 The effects of emphasis

The use of emphasis helped us clarify the pragmatics of the accents. First, the fact that all speakers, including those who used “uptalk” (L* H-H%) in the non-emphatic levels of the phone dialog, switched to L+H* in the emphatic levels, coupled with the fact that emphasis was more difficult to elicit in the L*+H dialogue (a difference also noted by the speakers themselves), clearly shows that the choice par excellence for emphasis is L+H*.⁸ This preference supports Pierrehumbert and Hirschberg’s (1990) analysis of the pragmatics of L+H* and L*+H according to which the use of L+H* indicates that the speaker and hearer should select the accented item as the one to be believed among a series of alternatives.

In addition, we had hoped that emphasis would help us clarify the status of the tones involved in the three accents under investigation. In particular, we expected, following Pierrehumbert (1980), Liberman and Pierrehumbert (1984) and Gussenhoven and Rietveld (2000), that starred Ls would have lower scaling under emphasis, but non-starred Ls would have higher scaling (or would not be affected by emphasis). This hypothesis was not borne out, as all L tones were raised with emphasis, just as Hs were, although the effect on the scaling of Ls was much less

pronounced: on average, the scaling of L increased by 24 Hz from level 1 to level 4 of emphasis, while the corresponding increase in the scaling of H was 50 Hz. Overall, these results suggest that emphasis may not be a good diagnostic of tonal composition, at least not in a straight-forward manner. This is related to the observation that emphasis may not only increase *pitch span* but may raise *pitch level* as well, thereby increasing the scaling of Ls (Lieberman and Pierrehumbert 1984; Ladd 1996; Erickson et al. 1994). Our results clearly corroborate those of earlier studies and provide additional evidence regarding the scaling of starred Ls, which were not previously examined.

Regarding alignment, our results provide partial support for our predictions, based on Ladd and Morton (1997), that peaks would align later with emphasis. Peak delay was indeed observed in our data, but only in the phone dialog, suggesting that peak delay is not associated with emphasis in general but is a characteristic of the L+H* accent. In contrast, increasing levels of emphasis in L*+H showed earlier alignment of H, suggesting that there may be a limit to how far a peak can be from the syllable it is associated with and that the average peak delay in L*+H is close to this limit.

On the other hand, L alignment was not affected by emphasis, a result that does not support the prediction of L delay in L+H* made by Ladd and Schepman (2003) (the only prediction regarding L alignment under emphasis that we are aware of). In our study the alignment of Ls was more consistent than that of Hs, independently of their (putative) phonological status. In this respect, our results are in line with the consensus emerging from the literature, namely that L and H tones have different properties: L tones are more stable in their alignment than Hs, the alignment of which appears to be affected by various parameters, such as emphasis (as the present results show), metrical factors, and speaking rate (Silverman and Pierrehumbert 1990; see also Prieto [submitted] and references therein). These traits appear to be characteristics of Ls and Hs in general and to have little to do with the starredness of the tones.

Overall then, emphasis was not as helpful as we had expected in distinguishing between possible phonological analyses for the bitonal accents we investigated, supporting the view of Arvaniti, Ladd and Mennen (2000) that starredness does not have concrete phonetic reflexes and can better be seen as an abstract phonological property of tones (on this point, see also Atterer and Ladd 2004; Prieto et al. to appear; Arvaniti et al. submitted). It remains of course to be seen whether our production results can be supported by perceptual data, or whether perceptual experiments

will provide results closer to those of Gussenhoven and Rietveld (2000), suggesting that speakers expect starred L tones to lower with emphasis, even if they themselves raise pitch level under the same conditions.

Finally, our results show that the most consistent effects of emphasis were on segmental duration, particularly the duration of the accented vowel. The observed differences, however, were barely large enough to be perceptible, as they averaged 12 ms. and just-noticeable differences in duration are expected to be in 10-40 ms. long (Lehiste 1970: 13). This suggests that, in this task at least, speakers relied more on F_0 and less on duration to indicate emphasis.

5. Conclusion

In conclusion, our experiment confirmed that American English has a three-way accentual contrast between H^* , $L+H^*$ and L^*+H , though our results also suggest that Minnesotan English may lack the $H^* : L+H^*$ contrast. Overall, however, our data confirmed Pierrehumbert's (1980) original analysis with respect to the contrast, and Pierrehumbert and Hirschberg's (1990) analysis of the pragmatics of these accents. This in turn shows that analyses according to which peak delay and the presence of a L tone are phenomena solely related to emphasis are not supported by the data. Further, generally held assumptions about the effects of emphasis on L and H tones were not supported by our data; rather both Ls and Hs appear to increase with emphasis, and this effect does not depend on a tone's starredness. The picture that emerges from these results is important in that it clearly demonstrates the need to examine linguistic varieties in more detail: our results showed that quite possibly one of the two dialects we examined lacks an accentual contrast that is often taken for granted in American English. They also showed that even when two dialects have the same contrast they may differ in how it is realized phonetically. Results like these, coupled with similar results emerging from work in other varieties of English and work on other languages show clearly that tonal alignment is crucial for differentiating intonational categories and that more detailed and larger-scale sociolinguistically informed research on intonation is crucially needed, and can help us understand (and possibly settle) disagreements between competing analyses and experimental results.

Notes

¹ We would like to thank Bob Ladd for comments on an earlier version of this manuscript, and Mary Beckman, Ben Munson, and the audiences at LabPhon 9 and the “Typology of Tone and Intonation” Workshop of the Network Tone and Intonation in Europe for helpful comments and suggestions. Thanks are also due to Dan Brassil and David Peterson for listening to our data and providing us with their native SoCal intuitions as to their naturalness. Any inaccuracies and inconsistencies remain of course ours.

² A fourth high accent, H*+H-, was eliminated as a possible pattern in Beckman and Pierrehumbert (1986) and will not be discussed further.

³ Trailing Ls in bitonal H*+L accents are not phonetically realized; their function is to trigger downstep of the following accent.

⁴ This is not a configuration that is discussed at length in the literature; at present we have no good explanation for it, apart from the fact that these contours are impressionistically more sly than the L*+H L- H% configuration would be. In any case, we do not think the difference in phrasal tones materially alters our results.

⁵ This contour is often referred to as *uptalk* and is considered a feature of *Valspeak*, supposedly the accent of young women from the San Fernando Valley in North West Los Angeles. Our data, however, suggest that this intonation pattern is most probably the default for declaratives in the speech of young Southern Californians of all genders and living well beyond the San Fernando Valley. A discussion of these unexpected data is beyond the scope of this paper.

⁶ The proportional alignment data were based on [m] for L and [ou] for H because in this way all the values on which the calculations were performed were positive and thus meaningful proportional alignment measures could be obtained.

⁷ No data on the distance between L and H for L+H* are given in Pierrehumbert (1980) and we are not aware of any other studies on this point.

⁸ This difference between the two dialogs may also be partly attributed to the fact that the investigator’s part in the date dialog was longer, so the speakers had greater difficulty keeping the emphasis levels increasing; however, this is unlikely to be the only reason for the differences observed between the date and the phone dialog with respect to emphasis.

References

- Arvaniti, Amalia
submitted On the presence of final lowering in British and American English. In *Tone and Intonation*, Carlos Gussenhoven, and Tomas Riad (eds.). Berlin: Mouton de Gruyter.
- Arvaniti, Amalia, D. Robert Ladd, and Ineke Mennen
2000 What is a starred tone? Evidence from Greek. In *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, Michael B. Broe,

- and Janet B. Pierrehumbert (eds.), 119–131. Cambridge: Cambridge University Press.
- Arvaniti, Amalia, D. Robert Ladd, and Ineke Mennen
submitted Tonal association and tonal alignment: evidence from Greek polar questions and contrastive statements. *Language and Speech*.
- Atterer, Michaela, and D. Robert Ladd
2004 On the phonetics and phonology of “segmental anchoring” of F_0 : evidence from German. *Journal of Phonetics* 32: 177–197.
- Beckman, Mary E., Julia Hirschberg, and Stefanie Shattuck-Hufnagel
2005 The original ToBI system and the evolution of the ToBI framework. In *The Phonology of Intonation and Phrasing*, Sun-Ah Jun (ed.), 9-54. Oxford: Oxford University Press.
- Beckman, Mary E., and Gayle Ayers-Elam
1997 Guidelines for ToBI Labelling. Columbus, OH: The Ohio State University Research Foundation.
- Beckman, Mary E., and Janet B. Pierrehumbert
1986 Intonational structure in Japanese and English. *Phonology Yearbook* 3: 255–309.
- Cummins, Fred
1999 Some lengthening factors in English speech combine additively at most rates. *The Journal of the Acoustical Society of America* 105: 476-480.
- Erickson, Donna, Kiyoshi Honda, Hiroyuki Hirai, Mary Beckman, and Seiji Niimi
1994 Global pitch range and the production of low tones in English intonation. *International Conference on Spoken Language Processing* 3: 651–654.
- Grabe, Esther, Brechtje Post, Francis Nolan, and Kimberley Farrar
2000 Pitch accent realization in four varieties of British English. *Journal of Phonetics* 28: 161-185.
- Grice, Martine
1995 Leading tones and downstep in English. *Phonology* 12: 183–233.
- Grice, Martine, Stefan Baumann, and Ralf Benz Müller
2005 German intonation in autosegmental-metrical phonology. In *The Phonology of Intonation and Phrasing*, Sun-Ah Jun (ed.), 55-83. Oxford: Oxford University Press.
- Dainora, Audra
2002 An empirically based probabilistic model of intonation in American English. Ph.D. diss., Department of Linguistics, University of Chicago.
- Fletcher, Janet, Lesley Stirling, Ilana Mushin, and Roger Wales
2002 Intonational rises and dialog acts in the Australian English Map Task. *Language and Speech* 45: 229–253.
- Gussenhoven, Carlos, and Toni Rietveld

- 2000 The behavior of H* and L* under variations in pitch range in Dutch rising contours. *Language and Speech* 43: 183–203.
- Halliday, Michael A. K.
1967 *Intonation and Grammar in British English*. The Hague: Mouton.
- Herman, Rebecca, and Julia Tevis McGory
2002 The conceptual similarity of intonational tones and its effects of intertranscriber reliability. *Language and Speech* 45: 1–36.
- Jilka, Matthias, Gregor Möhler, and Grzegorz Dogil
1999 Rules for the generation of ToBI-based American English intonation. *Speech Communication* 28: 83–108.
- Klatt, Dennis H.
1976 Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America* 59: 1208–1221.
- Labov, William, Sharon Ash, and Charles Boberg
1997 A national map of the regional dialects of American English.
http://www.ling.upenn.edu/phono_atlas/NationalMap/NationalMap.html
- Ladd, D. Robert
1978 *The structure of Intonational Meaning*. Bloomington, Indiana: Indiana University Press
- Ladd, D. Robert
1983 Phonological features of intonational peaks. *Language* 59: 721–759.
- Ladd, D. Robert
1996 *Intonational Phonology*. Cambridge: Cambridge University Press.
- Ladd, D. Robert, and Rachel Morton
1997 The perception of intonation emphasis: continuous or categorical? *Journal of Phonetics* 25: 313–342.
- Ladd, D. Robert, and Astrid Schepman
2003 “Sagging transitions” between high pitch accents in English: experimental evidence. *Journal of Phonetics* 31: 81–112.
- Lehiste, Ilse
1970 *Suprasegmentals*. Cambridge, MA: The MIT Press.
- Liberman, Mark, and Janet Pierrehumbert
1984 Intonational invariance under changes in pitch range and length. In *Language Sound Structure*, Mark Aronoff, and Richard Oehrle (eds.), 157–233. Cambridge, MA: The MIT Press.
- Menn, Lise, and Suzanne Boyce
1982 Fundamental frequency and discourse structure. *Language and Speech* 25: 341–383.
- Pierrehumbert, Janet

- 1980 The phonology and phonetics of English intonation. Ph.D. diss., Department of Linguistics and Philosophy, MIT.
- Pierrehumbert, Janet
1981 Synthesizing intonation. *The Journal of the Acoustical Society of America* 70: 985–995.
- Pierrehumbert, Janet
1993 Alignment and Prosodic Heads. *Proceedings of the Tenth Eastern States Conference on Linguistics*, 268–286. Ithaca, NY: DMLL Publications.
- Pierrehumbert, Janet, and Julia Hirschberg
1990 The meaning of intonational contours in the interpretation of discourse. In *Intentions in Communication*, Philip R. Cohen, Jerry Morgan, and Martha E. Pollack (eds.), 271–311. Cambridge, MA: The MIT Press.
- Pierrehumbert, Janet B., and Shirley A. Steele
1989 Categories of tonal alignment in English. *Phonetica* 46: 181–196.
- Prieto, Pilar
submitted Word edge tones in Catalan *Italian Journal of Linguistics* 17.
- Prieto, Pilar, Mariapaola D’Imperio, and Barbara Gili Fivela
to appear Pitch accent alignment in Romance: primary and secondary associations with metrical structure. *Language and Speech*.
- Silverman, Kim, Mary Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert, and Julia Hirschberg
1992 ToBI: a standard for labeling English prosody. *Proceedings of the 2nd International Conference on Spoken Language Processing*, 867–870.
- Silverman, Kim A., and Janet B. Pierrehumbert
1990 The timing of prenuclear high accents in English. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, John Kingston, and Mary E. Beckman (eds), 72–106. Cambridge: Cambridge University Press.
- Sproat, Richard (ed.)
1998 *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach*. Dordrecht: Kluwer Academic Publishers.
- Ward, Gregory, and Julia Hirschberg
1985 Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* 61: 747–776.
- Welby, Pauline
2003 Effects of pitch accent position, type and status on focus projection. *Language and Speech* 46: 53–81.
- Xu, Yi
1999 Effects of tone and focus on the formation and alignment of F_0 contours. *Journal of Phonetics* 27: 55–105.