CHAPTER 9

Prosody

Erik R. Thomas and Tyler S. Kendall

9.1 Prosodic Variables

In Latino Englishes, as for any varieties of English, the study of prosody has lagged behind the study of segmental aspects of speech. Intonation and prosodic rhythm in Latino varieties have each been an object of just a few studies. Articulation rate of Latino English has been examined empirically in just a single previous project. Although overlooked by researchers, prosody pervades every utterance a speaker makes and may be one of the first variable aspects of speech that a listener notices, much like it is one of the first aspects of speech that infants recognize (e.g., Mehler et al. 1988). Moreover, as noted in Chapter 1, it certainly can contribute to the distinctiveness of speech varieties that result from language contact. The neglect of prosody in sociolinguistic studies is unfortunate. Nevertheless, prosodic studies of language variation have made considerable strides over the past two or three decades (see the discussion in Thomas 2011) and methods for studying variation in some aspects of prosody are now well established.

Here, we explore, in North Town speech, several elements of prosody that now have well-defined analytical methods: intonation, prosodic rhythm, and rate of speech. For intonation, we examine two of the numerous possible variables, peak delay of pre-nuclear pitch accents and incidence of boundary tone types. Peak delay, a measure of one aspect of pitch accents, has, in various formulations, proved one of the most useful factors for differentiating intonational variation mathematically. Our foray into types of boundary tones, which occur at the ends of phrases, is based on transcribed notations, unlike peak delay, which by its nature relies on physical measurements of timing. Prosodic rhythm has to do with the evenness of successive elements in the speech signal, and here we focus on vocalic elements, with a comparison to the Spanish rhythm of our most Spanish-dominant subjects. For rate of speech, we examine articulation rate, in syllables spoken per second (σ/s), separately for two parts of utterances that exhibit distinct rates: syllables falling before the final foot of a prosodic phrase and those falling in the final feet of prosodic phrases.

9.2 Intonation

Studies of intonational variation have expanded in recent years, particularly in Europe. Regional variation in UK English, for example, has been the subject of a number of important intonation studies (e.g., Pellowe and Jones 1978; Grabe et al. 2000; Ladd et al. 2009). Regional variation in American English intonation has received a modest amount of attention (e.g., Clopper and Smiljanic 2011; Reed 2016), as has African American English intonation (e.g., Tarone 1973; Loman 1975; Jun and Foreman 1996; Thomas 2015; McLarty 2018). Latino English intonation, however, is poorly studied. In spite of Metcalf's (1972b, 1974) early imploration for work on it, apparently only Castro-Gingrás (1972), Penfield (1984; see also Penfield and Ornstein Galicia 1985), and Callahan (2008) ever heeded the call. A thorough understanding of Latino English intonation should require some information on Spanish intonation, which potentially could provide transfer input, and fortunately Spanish intonation has enjoyed extensive study (see, e.g., Prieto, van Santen, and Hirschberg 1995; Beckman et al. 2002; Face and Prieto 2007; O'Rourke 2014). Recent work has also shed light on dialectal variation in Spanish intonation (e.g., Colantoni and Gurlekian 2004; Willis 2008, 2010; Colantoni 2011; O'Rourke 2012), although Mexican Spanish intonation appears to fall within the mainstream of Spanish intonation as a whole (Prieto, van Santen, and Hirschberg 1995).

A large fraction of the findings reported in previous studies relate to the overall shape of fundamental frequency (F_0) peaks. Intonational peaks, regardless of language, vary as to whether they appear more as an abrupt rise followed by either a gradual fall or a leveling off, or, conversely, as a gradual rise to a later peak. These peaks are known as pitch accents and are assumed to be at least somewhat salient to listeners. Each pitch accent is associated with, or "hosted" by, a stressed syllable, although its F_0 peak may fall after the end of the host syllable (but never before). Pitch accents are only part of intonational structure; also crucial are phrases and the tones that mark phrasal boundaries. Both English and Spanish are regarded as exhibiting two phrasal types, Intonational Phrases (IP) and intermediate phrases (ip), in a hierarchy, with each IP consisting of one or more ips. Every IP and ip is marked by a special F_0

contour, called a boundary tone, at its end. More often than not, boundary tones consist of a movement in F_o, falling or rising, from what immediately precedes them. Pitch accents have their own hierarchy. Ordinarily (excepting certain disfluencies), a phrase has to have at least one pitch accent. Phrases can contain more than one pitch accent, however, and the last pitch accent is called the *nucleus*. Nuclear pitch accents may be influenced by boundary tones so that their peak falls earlier than it otherwise would, which helps to distinguish nuclear pitch accents from boundary tones in running speech. As a result, nuclear pitch accents must be considered separately from pre-nuclear pitch accents in quantitative analyses, and in many studies, nuclear pitch accents are omitted altogether from analysis. Phrases are also distinguished as showing either broad focus, in which no pitch accent exhibits special emphasis, or narrow focus, in which a pitch accent is emphasized by magnifying some of its physical properties. Narrow focus is used to draw attention to new information or to contrast the word containing the focus with a word in another phrase, and in fact narrow focus is often split into two categories corresponding to these two functions.

Quantitative studies of intonation may compare frequencies of different kinds of pitch accents or boundary tones, semantic uses of particular tones, or phonetic properties of tones. Phonetic properties are based on F_o measurements, timing measurements, or a combination. F_o measurements customarily involve differences between the F_o of a peak and that of an adjacent (usually preceding) trough. Timing measurements, of which peak delay is the prime example, have to do with durations of phenomena. Peak delay describes where the F_o peak of a pitch accent falls relative to the onset and offset of its host syllable. It is useful for differentiating pitch accents that show an abrupt rise at the beginning from those showing a more gradual rise. Metrics that combine F_o and timing include measures of the rate of F_o rise or fall and measures of whether a tone shows a sharp peak (or trough) or has a plateau.

The choice of intonational features on which to focus was guided by what previous researchers had found. Castro-Gingrás (1972: 9), working with subjects from Riverside, California, described three differences between Mexican American speech indistinguishable from Anglo speech and Mexican America speech that deviated slightly from Anglo speech: the latter (1) showed a less steep fall (only a one-pitch contrast) at the end of statements; (2) showed rises at the end of *wh*-questions instead of falls; and (3) showed more "intonation peaks" (i.e., pitch accents) in simple sentences without contrastive focus. Penfield (1984) formulated a longer list of

MxAE intonational features in her study in El Paso, Texas. As enumerated in Penfield (1984: 53–56), they consist of the following:

Generalization 1.	Rising glides are used to highlight or emphasize specific words
Generalization 2.	Rising glides are maintained at the end of a declarative sentence
Generalization 3.	Sentence contours often begin above the normal pitch of voice
Generalization 4.	Rise-fall glides occurring in sentence-final contours mark the stereotypic Chicano "accent"
Generalization 5.	Final endings of declarative, neutral statements have a one-pitch contrast

One feature that stands out in Penfield's list is that rising of tones (glides) is part of three of the five generalizations. Diagrams of stylized pitch contours that Penfield and Ornstein-Galicia (1985: 37) provided clearly indicate that they were referring to gradual rises in pitch, the perceptual analog of F_0, with late peaks. Generalization 1 certainly refers to *pitch accents* and apparently to those exhibiting late peaks. Generalization 2 has to refer to *boundary tones*, probably the one labeled as L-H% in the modern Tone and Break Index (ToBI) system (see Beckman, Hirschberg, and Shattuck-Hufnagel 2005). It is important to note that Penfield used the now-antiquated transcription system of Trager and Smith (1951, based on earlier work by Pike 1945), which uses a system of four pitch levels, whereas ToBI recognizes only two pitches, high (H) and low (L), that are conceptualized as targets instead of levels. As a result, some uncertainty in how to interpret Penfield's findings in terms of current practice is inevitable. Generalization 5 may refer to a nuclear pitch accent with a late peak followed by a low (L-L% in ToBI notation) boundary tone. The recurrence of "rising glides" in Penfield's descriptions suggested to us that examining MxAE for pitch accents and boundary tones with late peaks would be a fertile endeavor. We chose not to test generalizations 3 and 5, which pertain to other features and for which it is unclear how they could be quantified. Any phrase-initial pitch accent would seem to satisfy Generalization 3 and a high nuclear pitch accent followed by a falling boundary tone would satisfy Generalization 5, but both of those configurations are widespread in English.

Callahan (2008), using a peak delay metric, compared pitch accents in North Town MxAE, northwestern Mexican Spanish, and three ethnic groups in the Raleigh/Durham area of North Carolina (Latinos [from

Durham], African Americans, and white Anglos). Her results showed that North Town MxAE showed significantly later peaks than the three North Carolina groups, but that it fell within the range of variation of Mexican Spanish. The differences between North Town MxAE and Durham Latino English presumably result from the differing ecologies of the two communities. See Chapters 10 and 11 for more on Durham and how it contrasts with North Town.

Spanish provides a readily available source for the late peaks that Penfield (1984) and Callahan (2008) described. Spanish intonation is characterized by its recurrent use of rising tones, and recent models of Sp_ToBI – i.e., the Spanish ToBI system – even recognize three distinct rising pitch accent types (Face and Prieto 2007; O'Rourke 2014). Rising pitch accents are so ubiquitous in Spanish, at least in broad-focus declarative utterances, that dialects that use few rising tones, such as Italian-influenced Buenos Aires Spanish, stand out conspicuously (Colantoni and Gurlekian 2004; Colantoni 2011). Callahan's (2008) analysis of northwestern Mexican Spanish confirms that Mexican Spanish shows the typical Spanish pattern of frequent rising tones.

The findings reported in Penfield (1984) and Callahan (2008) led us to narrow our analysis to just two quantitative analyses: peak delay, which we will apply to pre-nuclear pitch accents, and computation of frequencies of boundary tone types. Multiple ways to calculate peak delay have been devised by various authors, depending on the landmarks within the host syllable used as reference points and on whether a subtracted distance in seconds (or ms) or a proportion is used. The fact that we were dealing with spontaneous speech influenced how we approached the problem. For spontaneous speech, it is necessary to take into account both syllabic consonants and the ability of any sonorant in an onset or coda to carry part of the F_0 contour. In addition, a subtracted value cannot account for speech rate variations. We settled on a technique that differs somewhat from previously published methods. Our approach was to calculate the temporal position of the peak of the pitch accent as a proportion of the sonorant part (not merely a vowel) of the host syllable. Thus, a proportion of 0.08 would indicate that the pitch accent peak fell very near the beginning of the sonorant section of the host syllable, whereas a proportion of 1.37 indicates that the peak fell on an unstressed syllable that followed the host syllable.

For the current analysis, transcription was performed by the first author in several stages over a period of years. The first stage involved orthographic transcription in Praat textgrids. Next, segments were demarcated, primarily for the purpose of conducting prosodic rhythm analysis, but this procedure also played a key role in intonational peak delay analysis. The third stage was the demarcation of phrases, including differentiation of IPs from ips. Well after that process was complete, ToBI notation of pitch accents and boundary tones was added. Then the necessary reference points for peak delay analysis – onset and offset of the sonorant part of the host syllable and location of the peak for all pre-nuclear H*, L+H*, and L*+H pitch accents (thus excluding L* and H+!H*, but including !H*, L+!H*, and L*+!H) – were marked on a separate tier of the textgrid. Autocorrelation pitch tracks superimposed on narrowband spectrograms (which can reveal F_0 patterns when pitch tracking fails) were utilized for this purpose, but always in conjunction with auditory judgment. Pitch accents were identified as peaks or troughs in F_0 that were both visible in the acoustic displays and impressionistically audible, and which could not be accounted for either by a boundary tone or phrasal tone or by the effects of consonant voicing on F_0. Figure 9.1 illustrates how the landmarks were used to determine peak delay for a pitch accent. Peak delay calculations were performed using a Praat script designed specifically for this purpose by Eric Wilbanks. No fewer than 40 pre-nuclear pitch accents were transcribed for each speaker. The last stage was to determine impressionistically which pitch accents were associated with narrow focus. This process was necessarily subjective; criteria for identification of narrow focus were that a) the pitch accent had to have a significantly greater F_0 value at its peak, often augmented by greater amplitude, than nearby pitch accents and b) the host word had to present new information or contrastive information. Assignment of narrow focus was somewhat conservative, with the result that only 240 pre-nuclear pitch accents (11.1 percent of the total of 2,157) were labeled as showing narrow focus. An important advantage of dividing the whole process into several stages is that it facilitated checking of notations from earlier stages at each step of the process. Intonational labeling is typically less definitive than most other kinds of linguistic analysis, with lower reliability of transcriptions, so checking and re-checking of transcriptions is virtually essential.

Figure 9.2 shows year of birth against mean peak delay of pre-nuclear pitch accents for each speaker in the sample. The general trend is reasonably clear – most Mexican Americans show higher mean proportions than most Anglos, which is to say that Mexican Americans tend to have later peaks. The two ethnicities do overlap, but two of the anomalous speakers are familiar from Chapter 4. The highest mean among Anglos is the youngest Anglo, the one who had grown up in a school class composed

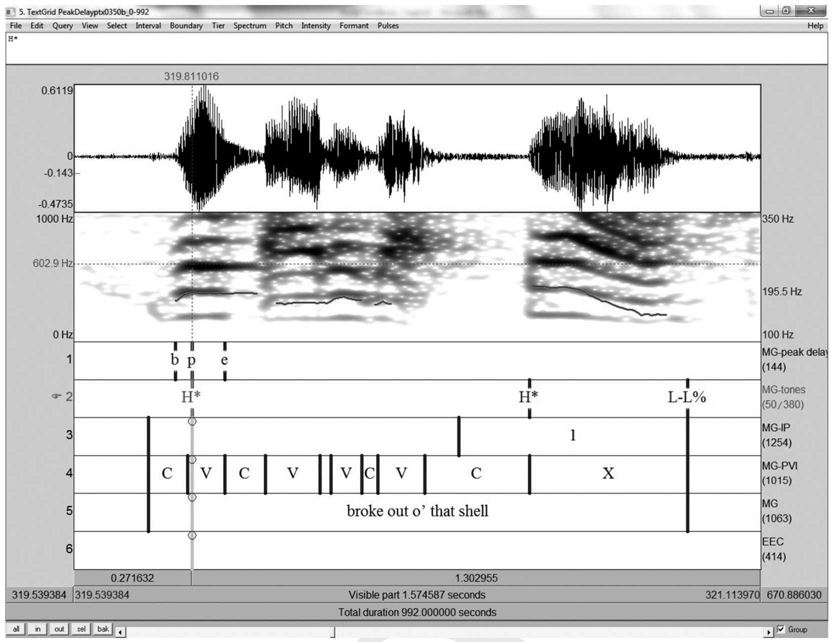

Figure 9.1 An illustration of peak delay measurement. A narrowband spectrogram of an utterance with an autocorrelation pitch track superimposed on it is shown. The pre-nuclear pitch accent is the first H* in the "MG-tones" tier, with "p" directly above it in the "MG-peak delay" tier to mark the point of peak F_0 associated with it. "b" in the "MG-peak delay" tier marks the onset, or beginning, of the sonorant part of the syllable (the onset of the /r/ in *broke*) and "e" marks the offset, or end, of the sonorant section, corresponding in this case to the offset of the vowel. The peak delay represents the proportion of the distance from "b" to "e" at which "p" falls, which for this example is 0.343.

almost entirely of Mexican Americans. The lowest mean of all the speakers is the Mexican American woman whose vowel configuration resembled configurations typical of young urban Anglos.

Mixed-effects linear regression was applied to the peak delay data using R. Ethnicity of speaker, speaker's year of birth or generation (birth cohort), speaker's number of years of education, sex of speaker, and whether the pitch accent carried narrow focus were coded as the independent variables. As in previous analyses, parents' average number of years of education was substituted for minors. The results, shown in Table 9.1, provide a clear picture. Penfield's (1984) Generalizations 1 and 4 are confirmed if our interpretation of them is correct. Ethnicity was significant at $p = 0.0034$

Table 9.1 *Mixed-effects linear regression results for peak delay of pre-nuclear pitch accents, with speaker as a random effect. Statistically significant results are marked with an asterisk. Akaike Information Criterion = 4145.6, Bayesian Information Criterion = 4191.0, log likelihood = -2064.8.*

Random effects:					
Speaker	(Intercept)		Residual		
Standard dev.	0.0976		0.6205		
Fixed effects:					
	Value	Standard error	DF	t value	p value
(Intercept)	1.3489	1.9608	2115	0.6879	0.4916
ethnicity	0.1756	0.0560	36	3.1347	0.0034
sex	0.0135	0.0492	36	0.2739	0.7857
year born	−0.0004	0.0010	36	−0.3728	0.7115
education	−0.0104	0.0067	36	−1.5397	0.1324
focus	0.1900	0.0428	2115	4.4424	<0.0001

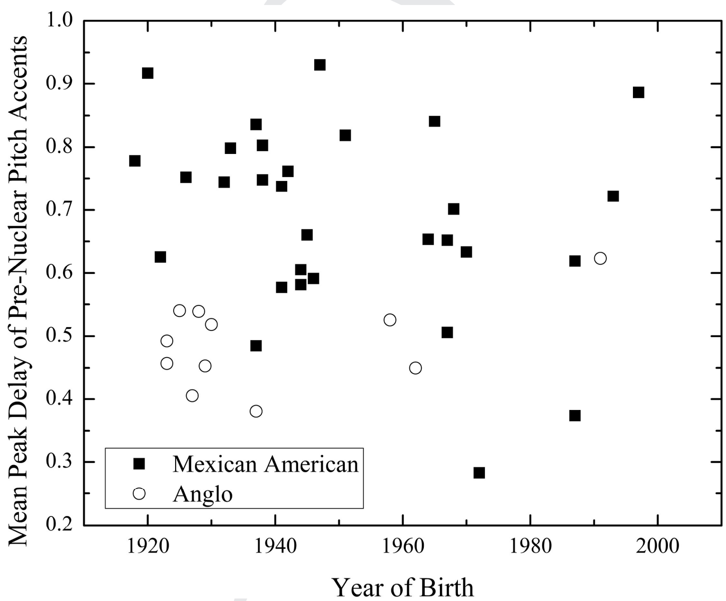

Figure 9.2 Year of birth against mean peak delay of pre-nuclear pitch accents for North Town subjects.

and whether the pitch accent carried narrow focus was significant at p <0.0001. Narrow focus tended to delay the peak, with narrow focus pitch accents showing a mean peak delay proportion of 0.801 and other pitch accents showing a mean proportion of 0.639. No other independent variables reached significance. The lack of significance for year of birth suggests that long peak delay is a stable part of the MxAE dialect of North Town, although it would be desirable to have a larger sample of the youngest speakers to confirm the stability.

Boundary tones were also examined in order to test Penfield's (1984) Generalization 2. Unlike for boundary tones, however, phonetic correlates were not used. The four boundary tones of English, L-L%, L-H%, H-L%, and H-H%, are differentiated for the most part by the kind and amount of change in F_0 associated with them. L-L% shows a falling F_0, L-H% shows a moderately rising F_0, H-L% shows a roughly level F_0, and H-H% exhibits a sharply rising F_0.[1] In theory, change in F_0 (known as pitch excursion) should be a simple phenomenon to calculate – the difference between a peak and a trough in the vicinity of the boundary tone could be computed, preferably in ERB units instead of Hz to normalize speaker differences. In practice, pitch excursion of boundary tones is quite problematic to determine. Boundary tones, of course, mark the end of intonational phrases, but ends of phrases commonly show non-modal phonation, either vocal fry or breathiness, for both of which F_0 values can be difficult or impossible to obtain. In addition, the four boundary tones are relatively easily distinguished by visual scrutiny of spectrograms, unlike the pitch accents H*, L+H*, and L*+H, which form a continuum without sharp differentiation.[2] As a result, we found it practical to base the analyses on counts of the ToBI transcriptions of the four boundary tones. Analysis was restricted to non-questions. Both utterances with broad focus and those with narrow focus were included, as focus did not show a clear-cut influence on usage of different boundary tones. This decision left only the demographic factors as independent variables in the statistical comparisons, which were conducted using logistic regression.

Figures 9.3–9.5 show the general trends. Note that only the first birth cohort (referred to here for convenience as a "generation") of Anglos, those born before 1940, are shown in Figures 9.3–9.5, although all Anglos are included in the statistical calculations. Figure 9.3 shows the four ethnicity/sex combinations. The most obvious trend is that males show a higher percentage of L-L% than females, whereas females show more L-H% than males. Ethnic differences are slight, but it can be discerned that Mexican Americans produce higher percentages of H-L% than Anglos.

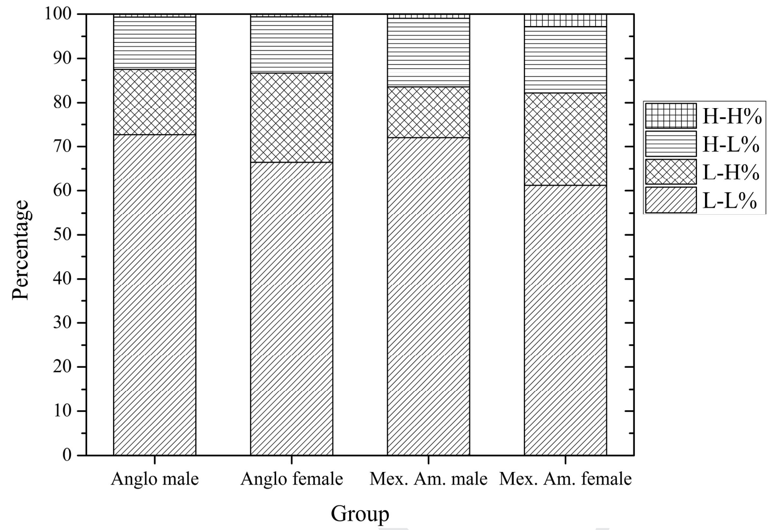

Figure 9.3 Percentages of the four boundary tones for each ethnicity/sex combination.

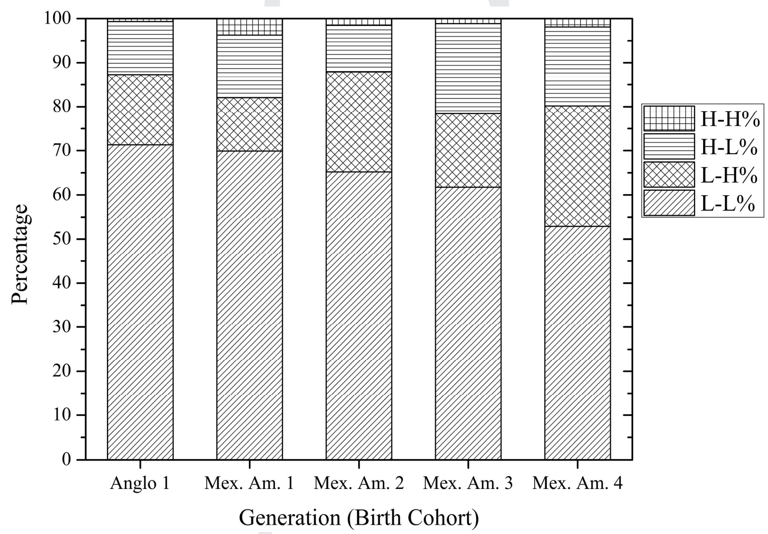

Figure 9.4 Percentages of the four boundary tones by generation.

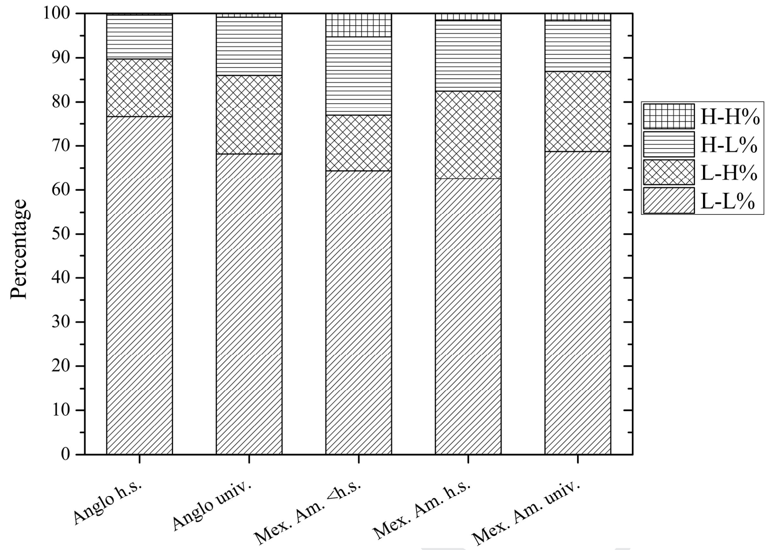

Figure 9.5 Percentages of the four boundary tones by educational attainment.

Figure 9.4 shows differences across generations. A trend is evident for L-L% – the oldest generation of Mexican Americans shows approximately the same percentage as elderly Anglos, but there is a steady decline in the L-L% across later generations. There are corresponding, but unsteady, increases in the percentages of L-H% and H-L%.

Percentages of boundary tones according to educational attainment are depicted in Figure 9.5. Oddly, increased education is associated with less L-L% among Anglos and more L-L% among Mexican Americans. Trends for the other boundary tones also follow irregular patterns.

Mixed-effects logistic regression analyses of the boundary tone distributions are shown in Tables 9.2–9.5. First, as shown in Table 9.2, the L-L% tone, the usual boundary tone for declaratives, was significantly more common among males and older speakers, but did not show a significant effect for ethnicity or education. Conversely, Table 9.3 shows that the L-H % tone, to which Penfield's (1984) Generalization 2 seems to refer, was significantly more common among Anglos and females, with no significant effect for education; for generation, it was significantly more common among the second and fourth generations and less common among the first and third. The ethnicity results for L-H% thus contradict Penfield's (1984: 54) finding that "[r]ising glides are maintained at the end of a declarative

Table 9.2 *Best-model logistic regression results for incidence of L-L% in non-questions, with speaker as a random effect. Reference settings are ethnicity = Anglo, sex = male, generation = 1st, education = less than high school. Akaike Information Criterion = 6151, Bayesian Information Criterion = 6190, log likelihood = −3069. Ethnicity and education were not significant in the initial run and are thus excluded here.*

Random effects:		
Group	Variance	Standard dev.
Speaker (Intercept)	0.21804	0.46695

Fixed effects:	Estimate	Standard error	z value	p value (>|z|)
Intercept	−1.2285	0.1533	−8.013	<0.0001
sex: female	0.5298	0.1640	3.230	0.0012
generation: 2	0.1781	0.2031	0.877	0.3806
generation: 3	0.4641	0.2159	2.150	0.0316
generation: 4	0.8912	0.2575	3.461	0.0005

Table 9.3 *Best-model logistic regression results for incidence of L-H% in non-questions, with speaker as a random effect. Reference settings are ethnicity = Anglo, sex = male, generation = 1st. Akaike Information Criterion = 4385, Bayesian Information Criterion = 4431, log likelihood = −2185. Education was not significant in the initial run and is thus excluded here.*

Random effects:		
Group	Variance	Standard dev.
Speaker (Intercept)	0.2689	0.5186

Fixed effects:	Estimate	Standard error	z value	p value (>|z|)
Intercept	2.0132	0.1945	10.349	<0.0001
ethnicity: Mex. Am.	0.5363	0.2554	2.100	0.0358
sex: female	−0.7505	0.2161	−3.472	0.0005
generation: 2	−0.5892	0.2415	−2.440	0.0147
generation: 3	−0.3311	0.2637	−1.256	0.2093
generation: 4	−1.0069	0.2990	−3.367	0.0008

Table 9.4 *Best-model logistic regression results for incidence of H-L% in non-questions, with speaker as a random effect. Reference settings are generation = 1st, education = less than high school. Akaike Information Criterion = 4009, Bayesian Information Criterion = 4054, log likelihood = −1997. Ethnicity and sex were not significant in the initial run and are thus excluded here.*

Random effects:				
Group	Variance		Standard dev.	
Speaker (Intercept)	0.2793		0.5234	
Fixed effects:				
	Estimate	Standard error	z value	p value (>|z|)
Intercept	1.5645	0.2573	6.080	<0.0001
generation: 2	−0.0227	0.2579	−0.088	0.9299
generation: 3	−0.7138	0.2670	−2.673	0.0075
generation: 4	−4.4418	0.3245	−1.362	0.1733
educational level: high school	0.5372	0.3245	1.656	0.0978
educational level: university	0.7168	0.3325	2.156	0.0311

sentence" for MxAE. H-L%, the level boundary tone, shown in Table 9.4, was most common among the third generation and less educated speakers. A further analysis using the fourth generation as the reference, however, showed no significant difference between the third and fourth generations. Finally, the H-H% tone, which is relatively rare in declaratives, is shown in Table 9.5 to be used significantly more often by speakers with less than a high school education than by speakers with a high school or university education.

9.3 Prosodic Rhythm

Prosodic rhythm is customarily viewed as separating languages into "stress-timed" and "syllable-timed" camps (Lloyd James 1940; Pike 1945; Abercrombie 1967). In stress-timed languages, individual syllables vary considerably in duration, but the duration between lexical stresses, which is to say the duration of prosodic feet, is idealized as relatively constant. In syllable-timed languages, each syllable is supposed to have approximately the same duration, regardless of stress placement. In recent decades,

Table 9.5 *Best-model logistic regression results for incidence of H-H% in non-questions, with speaker as a random effect. Reference settings are sex = male, education = less than high school. Akaike Information Criterion = 954.5, Bayesian Information Criterion = 987, log likelihood = −472.2. Ethnicity and generation were not significant in the initial run and are thus excluded here.*

Random effects:				
Group	Variance		Standard dev.	
Speaker (Intercept)	0.2438		0.4938	
Fixed effects:				
	Estimate	Standard error	z value	p value (>|z|)
Intercept	3.7225	0.4362	8.533	<0.0001
sex: female	−0.6029	0.3208	−1.880	0.0602
educational level: high school	0.9072	0.3723	2.436	0.0148
educational level: university	0.9608	0.4242	2.265	0.0235

this opposition has come to be seen as a gradient, not as an absolute difference (Miller 1984). One constant, however, is that English and most other Germanic languages are regarded as falling on the stress-timed end of the spectrum, while Spanish and most other Romance languages are considered to fall on the syllable-timed end of the continuum. Hence rhythm is an object of interest in studies of varieties of English with a Spanish substrate.

The shift to viewing rhythm as a continuum led to the development of several mathematical formulas for assessing it. Cummins (2002) notes that they can be divided into those (e.g., that of Ramus, Nespor, and Mehler 1999) based on the notion that the phonotactics of a language determine rhythm and those (e.g., by Low, Grabe, and Nolan 2000) that do not assume that phonotactics can account for all aspects of rhythm. English, with its variation in intrinsic vowel duration (Peterson and Lehiste 1960) and its variation in syllable durations due to stress variation, is well-suited to the latter approach. This reasoning led us to favor the Normalized Vocalic Pairwise Variability Index (conventionally abbreviated as nPVI_V or simply nPVI) formula devised by Low and Grabe (1995) and Low et al. (2000). This formula divides the difference in durations of vowels in adjacent syllables by the mean of the durations of the two syllables. All the resulting quotients for an individual speaker can then be assessed as a bloc, using a mean or median value. For instance, of the first two vowels in Figure 9.1, the vowel in *broke* has a duration of about 86 ms and the diphthong in *out* (treated as a single vowel) has a duration of 126 ms. The quotient for these two durations is thus, with rounding, 0.38.

For studies of dialectal differences in rhythm, previous authors have used not only the nPVI formula and a companion formula for consonantal intervals, but also formulas from Ramus et al. (1999) and related formulas from Dellwo (2006), often in combination. A number of studies of English, including Low and Grabe (1995) and Low et al. (2000) – both of which examined Singapore English – have focused on forms of English with recent influence from other languages; see also Deterding (2001) on Singapore English, Szakay (2008) on New Zealand Māori English, and Torgersen and Szakay (2012) on Multicultural London English. Fought and Fought (2003), Carter (2005b), and Thomas and Carter (2006) have examined Latino English rhythm. Typically, English varieties with recent substrate influence, including Latino Englishes, precipitate as more syllable-timed than most longstanding forms of English. Fewer studies have examined rhythm in varieties of English with only old substrate influence or no demonstrable influence from other languages (see White and Mattys 2007 for several regional British dialects; Clopper and Smiljanic 2015 for regional US dialects). The latter study has important implications for the North Town analysis because it finds that Southern US English, which would include the Anglo contact dialect in North Town, shows more extreme degrees of stress-timing than other regional dialects of US English. In fact, they found that the South stood out from other US regions on most of their metrics, showing, in general, a greater degree of variation in its vowel interval durations, a smaller degree of variation in its consonantal interval durations, and a greater percentage of the duration of utterances taken up by vowels.

Because the nPVI formula was originally designed for read speech, some adjustments were necessary to make it suitable for spontaneous speech. We followed Thomas and Carter (2006) in this matter. The key adaptation is to exclude utterance-final feet, which show prolongation, from the calculations. This exclusion is extended to feet occurring immediately before a filled pause – a filled pause consisting of a hesitation marker such as *uh* or *um* – or before a disfluency. This change is vitally necessary for analysis of spontaneous speech because speakers differ greatly in the number of breaks they produce – some individuals habitually produce frequent breaks, causing a larger proportion of their syllables to fall in final feet, where they would be subject to phrase-final lengthening (see §9.5 below), whereas other individuals characteristically produce many long, uninterrupted utterances and thus fewer syllables that are subject to final lengthening. Excluding final feet removes this skewing effect. The lower limit for unfilled pauses was set at 50 ms when the silent period could not be accounted for as a voiceless stop occlusion. This limit is lower than in most other treatments, but pauses as short as 50 ms still sounded impressionistically like pauses to our ears and were commonly associated with phrase-final lengthening in the preceding foot. We also dispensed with a feature of the nPVI formula that we considered superfluous, the multiplication of the mean (or, if we had applied the multiplication to our analysis, median) value by 100. Syllabic consonants, such as the [n̩] in *button*, were treated as vocalic intervals. Coda /r/ and /l/, which behave much like diphthongal glides, were treated as part of the duration of the preceding vowel, but onset /r/ and /l/, which are more clearly consonantal, were not. Successive vowels falling in different syllables, such as in *doing*, were treated as separate vocalic intervals; demarcation of the boundary between them was based on a combination of audition and scrutiny of formant patterns, F_0 patterns, and timing in spectrograms and pitch tracks.

For each speaker, we obtained a minimum of 250 adjacent syllable comparisons for which neither syllable was part of an utterance-final foot, for a total of 11,325 comparisons. Figure 9.6 shows the median values for every speaker. Medians are shown instead of means because distributions of nPVI values are strongly skewed, but for the most part means differ merely in being a few hundredths higher than medians and have a minimal impact on the sequencing of speakers along the continuum. For comparison, we have included median nPVI scores for Spanish spoken by ten elderly North Town subjects, nine of whose English scores are also shown.[3] Higher scores represent relatively stress-timed speech, lower scores relatively syllable-timed speech. An obvious pattern emerges. All but the youngest Anglo exhibit high median values, ranging from nearly 0.55 to over 0.70, and seven are over 0.60. Spanish, conversely, shows uniformly low median values, under 0.42. Nearly all the MxAE values fall between 0.40 and 0.60, thus representing an intermediate state between Spanish and Anglo English. Younger Mexican Americans, moreover, do not show a movement toward the high values of the majority of Anglos.

The prolongation that applies to final feet occurs not only in the last foot before a pause (filled or unfilled) or a disfluency, but also in the last foot of any IP, including those not followed by a pause, and possibly in the last feet of ips. Figure 9.7 shows the median nPVI values when the final feet before

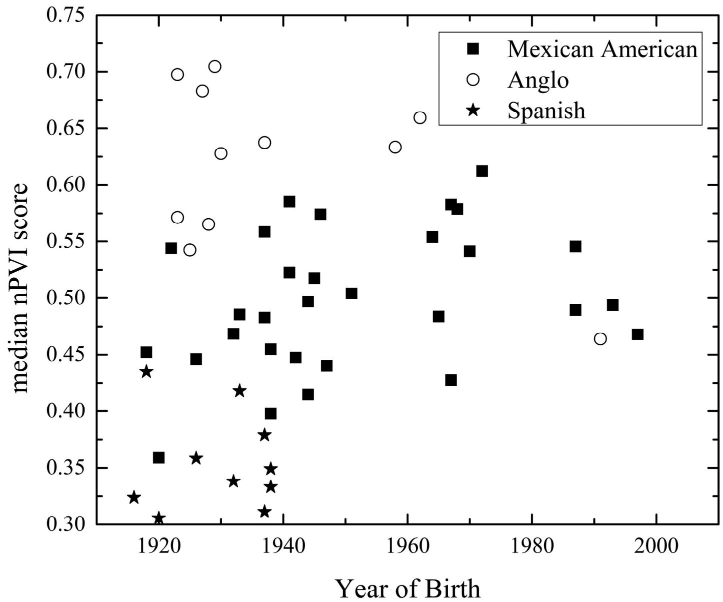

Figure 9.6 Median nPVI scores for North Town subjects.

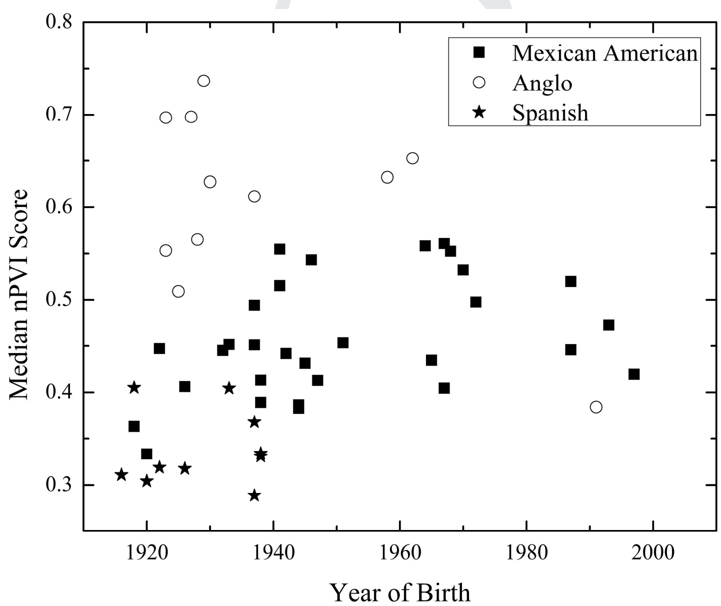

Figure 9.7 Median nPVI scores for North Town subjects, with final feet before utterance-internal Intonational Phrase and intermediate phrase breaks excluded.

utterance-internal IP and ip breaks are excluded. The median values are slightly lower, but the overall pattern is maintained: the highest values are those of Anglos, Spanish comprises most of the lowest values, and the majority of MxAE values fall in the middle.

In statistical analyses of prosodic rhythm, even when only vowels and syllabic consonants are considered in the analysis, as for nPVI_V, it is important to take into account the fact that segmental durations are sensitive to several influences. The normalization part of the formula (i.e., dividing by the mean of adjacent vowel durations) counteracts one of those factors, the rate of speech. Nevertheless, other influences remain. Thus, vowels are shorter before voiceless obstruents than elsewhere; low vowels tend to be longer than corresponding mid vowels, which average longer than corresponding high vowels; stressed vowels are longer than unstressed vowels; intrinsic vowel length gives some vowels greater durations than others; and vowels in the final feet before any IP boundary (including, but not limited to, positions before pauses) and perhaps before ip boundaries are prolonged. All of these influences were factored into the statistical analysis. For intrinsic vowel length, a division was made between the so-called "lax" or short vowels (BIT, BET, BAT, BOT, BUT, and BOOK) and all other vowels, including diphthongs. For phrasal boundaries, feet falling before a pause or disfluency were excluded, as they were in Figure 9.7, but feet falling before utterance-internal IP and ip boundaries were included. Because nPVI compares adjacent vowels, the coding addressed the differential in the contexts of the two vowels. For example, if the first vowel was followed by a voiceless obstruent and the second by a voiced sound, the voicing differential was coded as 1, but if both were followed by a voiceless obstruent or both were followed by a voiced sound, the voicing differential was coded as 0.

Table 9.6 shows the results of the mixed-models linear regression analysis. As noted in the caption, speaker sex, speaker's year of birth, the voicing differential of following segment, and the vowel height differential did not reach statistical significance at $p < 0.05$ or even $p < 0.10$. However, ethnicity was strongly significant and educational level barely reached statistical significance. Increased education was associated with a greater degree of stress-timing. Among the linguistic factors, stress, position in the last foot before a phrasal boundary, and intrinsic vowel length were all strongly significant.

The overall results for prosodic rhythm appear, at first glance, to indicate that the Spanish substrate has pulled MxAE rhythm in North Town in a syllable-timed direction, at least moderately so. Ethnicity is statistically

Table 9.6 *Best-fit mixed-models linear regression results for prosodic rhythm in North Town. Akaike Information Criterion = 6796.2, Bayesian Information Criterion = 6854.8, log likelihood = −3390.1. Sex, year born, voicing differential of following segment, and vowel height differential were not significant in the initial run and are thus excluded here.*

Random effects:					
Speaker	(Intercept)		Residual		
Standard dev.	0.0489		0.3246		
Fixed effects:					
	Value	Standard error	DF	t value	p value
(Intercept)	0.2848	0.0394	11281	7.2245	<0.0001
Ethnicity	−0.1045	0.0191	38	−5.3035	<0.0001
Education	0.0052	0.0026	38	2.0462	0.0477
final foot of IP/ip	0.0871	0.0065	11281	13.4588	<0.0001
Stress	0.3748	0.0064	11281	58.8155	<0.0001
intrinsic vowel length	0.0603	0.0062	11281	9.8108	<0.0001

significant, with Anglos showing higher (and thereby more stress-timed) nPVI values than Mexican Americans, and the Mexican American nPVI median values fall fairly neatly between Anglo English values and Spanish values. However, in this case, a presumption of Spanish influence may be premature. Clopper and Smiljanic (2015) noted that Southern English stands out from other US values in that it falls farther on the stress-timed end of the continuum on all vowel-related measures. The Anglo contact dialect in North Town, as previous chapters have demonstrated, is quite strongly marked as a Southern variety. It is possible, then, that North Town MxAE is not really more syllable-timed than most US dialects, but instead that North Town Mexican Americans simply rejected the extreme stress-timed rhythm of North Town Anglos just as they rejected other Southern Anglo variants. It would be instructive to compare North Town MxAE rhythm with that of Anglo dialects in other regions. Unfortunately, Clopper and Smiljanic (2015) is the only currently available study that has compared American regional varieties for rhythm and their data are not perfectly comparable with the North Town data because they used read speech instead of spontaneous speech and because of some computational differences – they did not exclude pre-pausal feet; they coded successive vowels, as in *doing*, as a single vowel interval; and they showed mean, not median, values. Thus, the nPVI values that they found (means of 60.85 for Southerners and 56.62–58.95 for other US dialects, with the multiplication by 100 from Low et al.'s 2000 original formula) cannot be used without qualification to gauge North Town MxAE scores. In fact, the North Town MxAE *mean* nPVI values average slightly lower than those of any dialect reported by Clopper and Smiljanic – factoring in the 100x transformation, the mean of these means is 54.61, with a standard deviation of 5.01 – but Clopper and Smiljanic's inclusion of pre-pausal feet should have inflated somewhat the values that they obtained. Hence there is no way to know, without appropriately comparable analysis of non-Southern varieties, how plausible it is that Spanish influence had an impact on MxAE rhythm in North Town.

9.4 Speech Rate

Speech rate encompasses several potential variables having to do with the overall tempo of speech. For instance, speech rate can be a measure of the overall rate of speech, in terms of the number of words or syllables produced over some unit of time, for a speaker over an entire recording or recording excerpt. This typically includes silences and disfluencies, as well as fluent speech, and is termed *speaking rate*. Speech rate can also be defined as a pause-exclusive measure of speech as it is produced in individual utterances by speakers. This is termed *articulation rate* and it is the specific variable investigated here.

Articulation rate, and speech rate more generally, have been examined in various ways by researchers interested in language production patterns as well as scholars of language variation and languages in contact. Kendall (2013) provides a detailed overview of work on speech rate and pause from a sociolinguistic perspective. He also presents an empirical investigation of ethnic, regional, age-based, and sex-based variation in articulation rate in US English by comparing articulation rates across over 150 speakers sampled from the Sociolinguistic Archive and Analysis Project (SLAAP; Kendall 2007) and identifies that these social factors (as well as factors involving the interlocutors) significantly affect articulation rates. Particularly relevant for a focus on articulation rate in North Town, Kendall (2013) included in his analysis a subset of the speakers examined here and found that Latino males are slightly faster talkers than Anglos and Latina females were slightly slower talkers than Anglos. In Texas specifically, Mexican American speakers were approximately 0.5 σ/sec faster than Anglos, on average.

Little other research has examined speech rate differences between Anglo and Latino English varieties, although De Johnson, O'Connell, and Sabin (1979) compared Spanish and English rates between two groups, native speakers of Spanish from San Luis Potosi, Mexico, and native speakers of English from St. Louis (N = 45 in each) in a narrative retelling task. They found faster speech for Spanish (6.08 σ/sec) than English (5.00 σ/sec) across the groups (although they also found that the Spanish speakers produced longer (in σs) narratives than the English speakers). Evidence that Mexican Spanish is spoken faster than American English and that Mexican American English is spoken faster than Anglo American English suggests that articulation rate may be affected by persistent language contact.

In addition to choosing a specific speech rate measure, research examining speech rate also has to consider its unit of analysis. Articulation rate, as a measure of rate exclusive of pauses, is typically implemented as a measure of rate for individual utterances (rather than, say, larger stretches of speech). However, as §9.5 explores further, articulation rates are not stable within utterances but are known to vary as a function of other prosodic factors, particularly phrase-final lengthening. The analysis undertaken here examines articulation rates as a function of individual IPs and further examines articulation rate as a function of both the length (in σs) of the pre-final and final feet of each utterance. (See Dankovičová 2001 for an inquiry into the "domain" of articulation rate and support for a focus on rate in the "domain" of IPs.)

The methods employed follow those used in pilot work on a subset of these North Town speakers (Kendall and Thomas 2010, Kendall 2013: 138–48). As was described in the sections above, IPs as well as ips and final feet were delimited for samples of speech from 41 speakers from North Town. Following with the treatment of the previous sections of this chapter, speakers were coded for ethnicity, sex, year of birth, and number of years of education, and the language of the IP was also coded as English or Spanish (all mixed-language utterances were excluded from the speech rate analysis). A total of 8,583 IPs were then analyzed statistically. Before proceeding to the statistical analysis, we illustrate the general patterning in Figure 9.8, which plots speakers' year of birth against the speaker-level articulation rate mean values. Speaker ethnicity (Mexican American and Anglo) and language (English unless depicted as Spanish) are indicated on the plot. As with previous treatments throughout this volume, Spanish data come from the older Mexican American speakers and, thus, nine of those ten speakers are represented twice on the plot, once for each

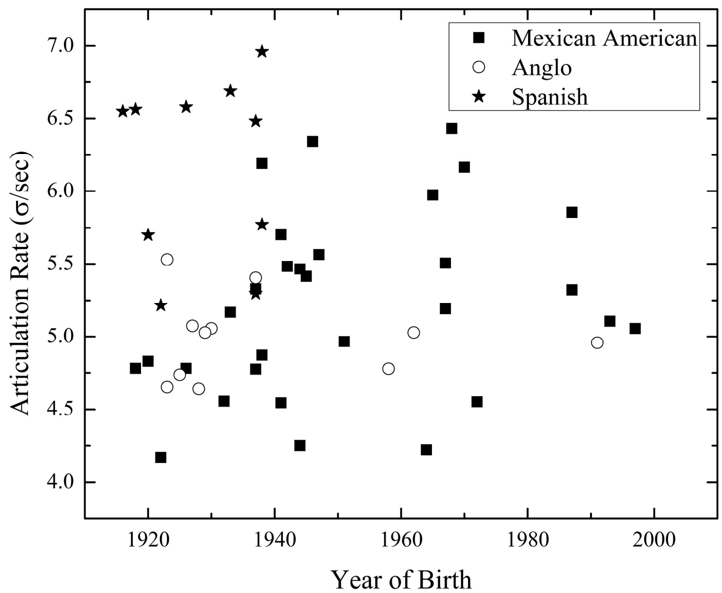

Figure 9.8 Articulation rates for North Town subjects.

language. In comparison to the data for prosodic rhythm, the data show somewhat similar clustering by ethnicity and language but not nearly the level of differentiation seen earlier.

In addition to the primary social factors of interest, IP length in number of syllables (excluding final-foot σs) and the length (in σs) of each IP's final foot were also coded as factors for each IP. Syllable counts were generated manually for final feet during the process of delimiting the final feet. Syllable counts for the rest of the IP were generated from processing scripts in the R language (R Core Team 2016). An English syllable counting function was applied to the English utterances from Kendall (2013), while a Spanish syllable counting function was developed for this project and applied to the Spanish language utterances.[4] In both cases, the functions are rule-based (based on language-specific spelling conventions) with exemption sets for special cases. The IP-level data were then submitted to mixed-effects linear regression with the dependent variable the articulation rate (σ/sec). Speakers were included in the model as a random intercept and IP length (in σs, excluding final-foot σs) and final foot length (in σs) were included as random slopes by speaker. In line with the fact that syllable length effects on articulation rate are known to be non-linear (see Kendall 2013), IP length was modeled using restricted cubic splines (with 5-knots). These splines allow the model to account for non-linear effects (cf. Baayen 2008: 176–81) and substantially improve model fit. All of the factors of interest were tested as main effects in the modeling, along with various two-way interactions. The best model, including only factors significant as main effects or as interactions that improve the model, is presented in Table 9.7.

As the model in Table 9.7 indicates, a number of factors significantly impact articulation rates in North Town. This is generally in line with Kendall's (2013) findings that articulation rate variation is highly patterned from a sociolinguistic perspective. Here we find significant effects for ethnicity and sex, with Mexican Americans realizing articulation rates 0.54 σ/sec faster than Anglos and males 0.59 σ/sec faster than females. There is no interaction for ethnicity and sex. Year of birth is the only demographic factor tested that did not emerge as significant. We also find a highly significant effect of language, as indicated in Figure 9.8, for which Spanish is spoken substantially faster than English. The significant interaction between language and years of education indicates that this difference decreases for the speakers with more years of education; speakers with higher levels of education have slower rates for their Spanish (and slightly faster rates for their English). The effects of IP length and final foot length are complex with the non-linear terms but are in line with patterns reported in Kendall (2013). These are not treated further here.

In sum, articulation rates in North Town are significantly correlated with speakers' ethnicity and sex. North Town Mexican Americans speak faster than Anglos and males speak faster than females in these interviews.

9.5 Phrase-Final Lengthening

The final prosodic feature considered here is phrase-final lengthening (PFL), the tendency for syllables preceding prosodic boundaries to be prolonged. PFL is a well-known aspect of prosody and has been described for numerous languages, including English (e.g., Klatt 1975; Byrd, Krivokapic, and Lee 2006; Turk and Shattuck-Hufnagel 2007; see also Fletcher 2010). Studies such as Byrd et al.'s (2006) articulatory examination of PFL timing have noted variability across speakers in PFL; however, little research has explored this phenomenon for sociolinguistic variation. PFL also interacts with articulation rate, as was shown in §9.4, in that the number of syllables in an IP's final foot interacts with the IP's length as a significant factor in the IP's overall articulation rate. (As mentioned

Table 9.7 *Best-fit mixed-model linear regression results for articulation rate in North Town. Akaike Information Criterion = 37005, Bayesian Information Criterion = 37161, log likelihood = −18481. Year born was not significant and is thus excluded here. IP length is length of IP (in σs) not including the final foot.*

Random effects:

Speaker	(Intercept)	IP length (random slope)	Final foot length (random slope)	Residual
Standard dev.	0.42023	0.02577	0.19498	2.07076

Fixed effects:

Factor	Value	Standard error	DF	t value	p value
(Intercept)	1.7771	0.3447	50	5.156	< 0.0001
IP length (non-lin.)	2.3531	0.2005	8478	11.735	< 0.0001
IP length (non-lin.)'	−81.0228	8.1275	8482	−9.969	< 0.0001
IP length (non-lin.)"	129.9155	13.3287	8487	9.747	< 0.0001
IP length (non-lin.)'''	−53.8609	5.9486	8499	−9.054	< 0.0001
final foot length	0.7998	0.1063	717	7.527	< 0.0001
ethnicity: MxA	0.5378	0.1744	35	3.085	0.0040
sex: male	0.5872	0.1571	34	3.739	0.0007
years of education	0.0243	0.0209	37	1.164	0.2519
language: Spanish	1.4647	0.2260	4012	6.48	< 0.0001
IP length (non-lin.) X final foot length	−0.7711	0.1366	8343	−5.644	< 0.0001
IP length (non-lin.)' X final foot length	30.3321	5.4238	8366	5.592	< 0.0001
IP length (non-lin.)" X final foot length	−49.1714	8.8627	8374	−5.548	< 0.0001
IP length (non-lin.)''' X final foot length	20.9496	3.9146	8399	5.352	< 0.0001
years of education X language: Spanish	−0.0559	0.0256	3839	−2.184	0.0291

earlier, Dankovičová 2001, focusing on Czech, examines the relative slowing of IP final syllable syllables, demonstrating again the interrelation between final lengthening and IP-level articulation rate.)

Given that we find articulation rate differences (§9.4) and prosodic rhythm differences (§9.3) across ethnicities, we might also expect that ethnolects implement PFL differently. In this section, we briefly consider whether ethnicity, or any other social factors in the North Town data, correlate with PFL. The measure of PFL implemented here is based on articulation rate and is the ratio of the articulation rate of the final foot over the articulation rate of the entire IP. This will be referred to as Final Lengthening Ratio (FLR). The calculation of FLR is demonstrated in (9.1), an IP uttered by ptx012, a Mexican American male born in 1938, in which the final foot is uttered at a rate of 0.815 of the full IP, i.e. FLR = 0.815.

(9.1) "when I came to work <u>here</u>"

Final foot (underlined)	= 1 syllable in 0.247 sec	= 4.049 σ/sec	= 0.815
Intonational Phrase	= 6 syllables in 1.208 sec	= 4.967 σ/sec	

The data were heavily trimmed to exclude IPs with FLR greater than 1.0 (which primarily arose in IPs containing disfluencies in the pre-final feet) or rare (N = 28) cases in which the final foot contained more than 3 syllables. This left N = 6,314 FLR measurements, which were logit transformed (using the logit function in R from Fox and Weisberg 2011) for modeling and then subjected to mixed-effects linear regression. The best regression model for these data, including a random intercept for speaker and random slopes, by speaker, for the length of the final foot (in σs) and the length of the pre-final foot of the IP (in σs), yields significance for only three main effects. These are the length of the pre-final foot of the IP (in σs), the length of the final foot (in σs), and the language. These effects are shown in Table 9.8.

Longer (in σs) final feet lead to higher FLRs, which indicate less final lengthening, while longer pre-final feet lead to lower FLRs, indicating relatively greater final lengthening. Spanish IPs have lower FLRs or greater relative final lengthening. This language difference appears peculiar at first glance but actually belies a language-specific prosodic difference in the data: the placement of lexical stress in Spanish words leads to a high proportion of final feet with 2 σs. This is shown in Table 9.9, which displays the relative proportion of final foot lengths for the Anglos,

Table 9.8 *Best-fit mixed-model linear regression results for final lengthening ratio (logit-transformed) in North Town. Akaike Information Criterion = 14932, Bayesian Information Criterion = 15007, log likelihood = −7455.2. Sex, ethnicity, year born, and years of education were not significant and are thus excluded here. IP length is length of IP (in σs) not including the final foot.*

Random effects:				
Speaker	(Intercept)	Final foot length (random slope)	IP length (random slope)	Residual
Standard dev.	0.1303	0.0842	0.0097	0.7834

Fixed effects:					
Factor	Value	Standard error	DF	t value	p value
(Intercept)	0.0464	0.0359	36.1	1.292	0.2050
final foot length	0.6787	0.0225	38.7	30.125	< 0.0001
IP length	−0.0312	0.0027	25.3	−11.718	< 0.0001
language:Spanish	−0.2425	0.0438	1347.0	−5.541	< 0.0001

Table 9.9 *Percentage of IPs with 1, 2, and 3 σ length final feet*

Ethnicity	Language	% 1 σ	% 2 σ	% 3 σ
Anglo American	English	64.5	30.4	5.1
Mexican American	English	69.5	26.3	4.2
Mexican American	Spanish	31.0	60.6	8.4

Mexican Americans speaking English and Mexican Americans speaking Spanish. Recall that the Mexican Americans speaking Spanish are nine of the same (older) Mexican Americans whose English is represented in the table.

The fact that language emerges as a significant factor in the statistical model for FLR, in addition to final foot length (and there is no significant interaction between language and final foot length), indicates that the language difference is nonetheless a factor in final lengthening – Spanish in these data realizes more final lengthening (as measured by lower FLR values) than English. Of greatest relevance to the present project, however, is the observation that none of the main demographic factors of interest – ethnicity, sex, education, or year of birth – significantly impact phrase-final lengthening.

9.6 Diverse Variables, Diverse Patterns

Prosody covers a range of variables that are not necessarily related to each other. It is thus not surprising that they showed strikingly diverse correlations with social variables in our analyses. Peak delay of pitch accents and prosodic rhythm were both strongly correlated with ethnicity. Even here, however, it is unclear whether the provenance of the MxAE pattern was the same for both variables. For intonation, a causal link from Spanish, which shows a high proportion of pitch accents with late peak delays in most dialects, to the long peak delay values of MxAE seems likely, particularly because the MxAE pattern does not match that known for any white Anglo dialect of American English. A connection with Spanish is less secure for prosodic rhythm, for which – according to the meager amount of previous work – MxAE resembles several regional US dialects. In this case, a reaction against the local Anglo English in North Town is at least as likely as an explanation as interference from Spanish. Of course, a combination of those two potential influences is also possible.

Quite a different picture emerged for boundary tones and articulation rate. Sex of speaker played a major role for both of these variables. This result contrasts with those of segmental variables, for which, as seen in earlier chapters, sex was not, in general, one of the stronger correlates. Three of the four boundary tones showed significant correlations with sex. Males showed an especially strong preference for the L-L% in non-questions, whereas females of both ethnicities were more prone to employing L-H%, suggesting potential gender connotations for those two tones. The oldest Mexican Americans showed the greatest similarity to the Anglo contact dialect in their distribution of boundary tones of any age group, which probably eliminates Spanish influence as an explanation for boundary tone patterning. The proportion of L-L% dropped with each generation. Sex of speaker was strongly correlated with articulation rate, males showing faster rates than females. Beyond the evident gender constraint, Spanish plausibly could have exerted an influence on MxAE articulation rates because there were significant correlations with both ethnicity and language – Mexican Americans showed faster articulation rates than Anglos, and, in turn, Spanish showed faster articulation rates than English, suggesting a carryover effect from Spanish to MxAE. As with prosodic rhythm, however, the possibility of Mexican American reaction against the local Anglo dialect lurks in the background. It is possible that North Town Anglo English, as a Southern variety, shows longer vowel durations than other regional dialects; see Clopper and Smiljanic (2015) for a review of the controversy over whether Southern dialects show longer vowel durations than other dialects, as well as an empirical analysis comparing durations across regions.

Finally, MxAE and Anglo English showed no significant differences in phrase-final lengthening, even though Spanish showed greater final lengthening than English. L2 learners of English in the oldest generation were able to assimilate English final lengthening patterns readily. Variables behaving in this manner should not be neglected; if they are, a skewed picture of an ethnolect will emerge, one representing it as more divergent than it actually is.

One crucial lesson to be taken from the prosodic analyses is how radically variables, even seemingly related variables, can diverge in their social correlations. No two variables considered here showed identical social patterning. Moreover, sex of speaker played a more important role here than it had in previous chapters. The prosodic results, as they contrast with each other and with those of segmental variables, amplify the point that the overall sociolinguistic patterning of a group cannot be determined with any certainty from just a few variables.

Notes

1. The boundary tones are also differentiated by their customary uses. L-L% is associated with declarative statements and wh-questions but may also be used in yes/no questions to denote an accusatory or demanding attitude. L-H%, the "continuation rise," is commonly used to indicate that the speaker wishes to continue speaking. Speakers often use H-L% when giving lists of items, but they occur in other situations as well, such as for expressing uncertainty, and at times appear in situations in which L-H% might be expected. H-H% is usually associated with yes/no questions, but it can occur occasionally elsewhere, such as for connoting excitement.
2. As with boundary tones, pitch accents are supposed to be differentiated semantically, not just phonetically. However, dialectal (including ethnolectal) differences, particularly where influence from another language is possible, may render semantic differentiation unreliable.
3. The tenth speaker opted to be interviewed only in Spanish, so no English is available for her.
4. R scripts for syllable counting are available at http://lingtools.uoregon.edu/scriptsetc.php.