10 Analyzing prosody

Best practices for the analysis of prosody

Malcah Yaeger-Dror and Zsuzsanna Fagyal

Introduction

Within sociolinguistics, the importance of socio-indexical meaning conveyed by *speech prosody*, also referred to as *suprasegmentals* (Lehiste 1970), was recognized early. Utterance-level variations of the three main acoustic correlates of prosody—fundamental frequency (F0—measured in Hz), duration (measured in msec), and intensity/amplitude (measured in dB)—provide evidence of dialect variation (Boudreault 1968; Yaeger 1979; Thomas and Carter 2006), have been recognized as possible conditioning factors for vowel shifting (LYS), and are crucial in managing interactions both within and across social groups (Gumperz 1982; Erickson 1985; Queen 1996; Schegloff 1998; Yaeger-Dror 2002; Couper-Kuhlen and Ford 2004; Fagyal 2010; Yaeger-Dror et al. 2010).

According to Thomas (2002: 168), "most instrumental sociolinguistic work [. . .] has been concentrated on variation in vowels: variation in consonants, prosody, and voice quality have received little acoustic analysis." Hay and Drager (2007: 92) concur. We hope this chapter will allow you to redress that imbalance.

Variation in F0, caused by the rate of vibrations of the vocal folds (perceived as *pitch*) cues both lexically-distinctive tone and pragmatic meaning, duration (perceived as *length*) signals quantity distinctions in vowels and consonants (Lehiste 1970) as well as emphasis or phrase boundaries (Klatt 1976), and amplitude modulations (perceived as variation in *loudness*), together with other prosodic correlates, mark stress and topic shift (Lehiste 1970) as well as cuing turn-transfer locations within a conversation (Goldberg 1978; Schegloff 1996, 1998; Local and Walker 2004). All three are interrelated (Johnson 2003): for instance high frequency sounds increasing in intensity can be perceived as rising in pitch. Fortunately, the three parameters generally covary, and often acoustic analysis of one is assumed to entail corresponding changes in the other two. Today, generally we rely on F0 which is now the easiest to measure.

The variables in play—such as ethnic or gender identity (dis)agreement, and assessment, among others—are almost impossible to elicit within even an interview corpus protocol, so analyses of segments, shifts, mergers, splits . . . etc., which are easier to analyze, are more common than studies of rhythm and intonation.

This chapter provides an overview of the much slower growing body of empirical research on socio-prosodic variation. We discuss methods and findings to provide a better understanding of the primitives and analytical tools used for phrase-, utterance-, and discourse-level analysis. First, we give an overview of empirical studies of rhythm, with special focus on quantitative studies of rhythm types, to suggest techniques for further studies. Then we offer an overview of techniques to investigate dialectal, gender-, ethnic-, and discourse-related meaning conveyed by tone alignment, tunes, utterance- and discourse-level variations of F0 and intonation.

Sociolinguistic variation and rhythm variation

F0, *duration*, and *amplitude* jointly provide cues for the parsing of the acoustic signal into more or less discrete units (Lehiste 1970; Gleitman et al. 1988; Zanto, Snyder, and Large 2006). Infants also use this information to distinguish noise from meaningful speech segments (Nazzi, Bertoncini, and Mehler 1998; Jusczyk, Houston, and Newsome 1999). In social-interactional research, *tempo*, or *speech rate*, is generally quantified as number of syllables or words per second or per minute. *Rhythm* is the ordered repetition of contrasting durational elements in the speech signal. These discrimination patterns are relevant to *syllable-timed* and *stressed-timed* languages, respectively. *Mora-timing*, with Japanese as the most well-known example, is a third option (Vance 2008).

Rhythm measures

So-called *stress-timed* languages (e.g, English, German), presumably display similar durations from one *stressed* syllable ("head" of the metrical "foot") to the next, with stressed syllables longer than unstressed syllables. So-called *syllable-timed* languages (e.g., Spanish, Italian), on the other hand, supposedly give each syllable approximately the same duration regardless of its metrical strength. The cornerstone of the distinction between stress- and syllable-timing is presumed isochrony of feet versus syllable durations. Cross-linguistic work on prosodic rhythm has capitalized on that distinction and possible influences on it although acoustic studies have not found evidence for a neat distinction (Roach 1982), since stress-like prominence patterns are even found in languages such as Italian and Greek, previously classified as syllable-timed (Dauer 1983), so stress-timing and syllable-timing are now considered ends on a continuum (M. Miller 1984).

Inspired by findings on infants' ability to detect basic speech segments in their native languages and Dauer's (1983) proposal of a continuous unidimensional model of rhythm, several studies have proposed protocols for the analysis of *rhythm metrics* (e.g, Ramus, Nespor, and Mehler 1999; Low, Grabe, and Nolan 2000; Deterding 2001; Ramus 2002), three of which are discussed here: %V, ΔC, and ΔV (see Table 10.1 on the book website).

"Vocalic intervals" correspond to vowels measured from left-to-right from the vowel or glide onset to the offset (chapter 8) and contrast with "intervocalic" (consonantal) intervals (regardless of the number of consonants included) (chapters 6–7). While Ramus, Nespor, and Mehler (1999) considered glides with preceding consonants (in words like queen /kwi:n/), subsequent studies group glides with vowels (e.g, Grabe and Low 2002), while the status of devoiced segments and acoustically ambiguous sounds (like liquids), has rarely been addressed. The best recommendation for analysis is to follow the segmentation protocol of a previous study of the same language (Arvaniti 2009).

Obviously, to carry out these rhythmic analyses correctly, consonant (chapters 6–7) and vowel (chapter 8) durations must be measured accurately. Several programs now claim to provide preliminary automated alignment of the speech stream,for example, Easy Align (Goldman 2010), a plug-in for Praat (Boersma and Weenink 2009a; chapter 8. Such programs save time, but ultimately the measurements must be corrected by hand.

Ramus, Nespor, and Mehler's (1999) first measure shown on Table 10.1, %V, is the sum of vowel durations divided by the total duration of an utterance, capturing the ratio of vocalic portions in a signal which is expected to be lower in languages with complex onsets and codas. The second, ΔC, is the standard deviation of consonantal durations, indicating varieties of syllable types, such as heavy onsets and codas: the higher ΔC, the greater the variation of consonantal durations. The third measure, ΔV, i.e. the standard deviation of vowel durations, was expected to be low in syllable-timed languages with no diphthongization and/or vowel reduction.

Prototypical stress-timed languages, such as English and Dutch, with both complex codas and onsets, and wide variation in vowel durations at different points in a phrase had low %V and high ΔC and ΔV values, despite the fact that English consonants do not vary with "stress" and sentence position as much as vowels (Klatt 1976); languages presumed to be syllable-timed, such as French, Italian, and Spanish showed the opposite tendency, with voiceless stops more variable in different positions than occurs in English. Japanese patterned separately from both types: it showed low ΔC and high %V values, i.e. indicating the prevalence of onset and coda structures with no diphthongization. Surprisingly, ΔV was found to be the least discriminating of the three measures.

While the idea of gauging phonotactic differences independently of language-specific phonological categories was innovative, the high sensitivity of interval measures to speech rate and interactive factors turned out to be a stumbling block, as Klatt (1976) had foreseen. Ramus, Nespor, and Mehler (1999: 269) "solved" this by measuring decontextualized utterances or short texts of comparable speech rate pronounced under tightly controlled conditions. Subsequent studies normalized durations of these controlled speech samples in order to further reduce the impact of speech rate and, thus, hoped to permit the use of less artificial corpora. Other measures have been designed to accomplish this task.

The normalised Pairwise Variability Index (nPVI-V), for instance, corresponds to the mean of the differences between successive intervals in the acoustic signals, divided by the sum of the same intervals. It was devised to capture the durational differences between adjacent stressed and unstressed vowels, which tend to be greater in so-called stress-timed (English, Dutch) than in so-called syllable-timed (French, Italian) languages (Grabe and Low 2002). Other PVI measures, such as rPVI-C (raw consonantal Pairwise Variability Index), are used in conjunction with other normalized scores to express the mean of the differences between successive consonantal versus vocalic intervals (cf. Nolan and Asu 2009 for a review).

A second series of measures—VarcoC and VarcoV—express the standard deviation of consonantal and vocalic interval durations divided by the mean consonantal and vocalic duration, respectively (Dellwo and Wagner 2003, Dellwo 2006). Each of these measures has been extensively applied to over 30 different languages and dialects in first and second language studies. Their success in discriminating between languages thought to be prototypically syllable versus stressed timed varies considerably, as does their ability to classify new languages along this dimension. Gut et al. (2002) and White and Mattys (2007) have tested different methods against each other but others have expressed doubts that even normalized durational variability can express mental representations of periodicity (Barry et al. 2003, Kohler 2008, Arvaniti 2009).

Relevant sociolinguistic studies

Since Low, Grabe, and Nolan's (2000) PVI-based study of British and Singapore English revealed more stress-timed tendencies in the latter, subsequent studies have used rhythm metrics on isolated sentences, short texts read in controlled conditions, guided interviews, and unscripted speech samples from various databases to investigate the possibility of sub- or adstratal influences on specific ethnolects. *Substratal* influence on rhythm can be inferred if the language is still in use elsewhere, but is difficult to prove when the purported substratum language is locally defunct. A stronger case can be made for *adstratal* influences, where the influencing language or dialect is in use; if you wish to make such claims, you must collect equivalent corpora of both the L1 and the L2 spoken in the same social situation by the same speakers, and measure the durations of both, following the same protocol as Udofot (2003) did for Nigerian English, Holm (2006) did for L2 Norwegian, or Frota and Vigario (2001) used for comparison of Brazilian and Old World Portuguese perhaps caused by substratal African language influences. Western dialects of Arabic permit more vowel reduction than Eastern varieties (Ghazali, Hamdi, and Barkat 2002). Thomas and Ericson (2007) investigated rhythm-type differences between Hispanic (Chicano) and White North Carolina English; Thomas and Carter (2006) compared African American and White Southern English, and Coggshall (2008) studied two Native-American English varieties from the Smoky Mountains and the Coastal Plain of North Carolina. O'Rourke (2008) compares two varieties of Peruvian Spanish that exhibit more stress-timed characteristics than expected with the speakers' Quechua dialect use. Fagyal (2010) compares adolescent peer-groups in contact with immigrant languages in France and finds a clear effect of speech rate but not of rhythm type.

Our conclusion is that rhythm-based analyses can be useful as possible evidence of *ad-* or *substratal* influences, as well as accommodative tendencies. Such analyses necessitate measuring vowel durations, consonant durations, and coding them for features known to influence duration. The choice to carry out such comparative studies is in your hands.

Prepausal lengthening

Prepausal lengthening is another critical feature of some stress-timed languages. Klatt (1976) determined that while prefinal lengthening is robust, weaker influences—vowel aperture, number and manner of consonants within the syllable, focal stress, and number of syllables in a word (Klatt 1976; Tauberer and Evanini 2009)—all are simultaneously in play even in tightly controlled corpora. We know that the more cells there are to fill, the more tokens are needed, so the amount of data required for a study of comparative durations is not negligible, even if your "corpus" is read, and your preliminary durational measures can be "automatically aligned." One might suspect such an effect to be washed out in more casual speech, but even though French is said to be a prototypical syllable-timed language, in sociolinguistic interviews of Quebec French prepausal lengthening was robust, despite competing influences (Yaeger 1979); this feature may be traceable to English adstratum influence, and certainly confirms Klatt's understanding that comparative durational measurements yield new understandings.

Silent and filled pauses, hesitations, and other timing factors

Both perceptual and instrumental methods have been used to identify communicatively-relevant silent intervals, or pauses, in speech.[1] This variable can be easily measured, once pauses are distinguished from voiceless stops (chapter 6) which are shorter (O'Connell and Kowal 2009: 99–114). In the ToBI system (discussed below), which was formulated based on trained NPR readers' rereading of news, pauses serve as "break indices" of hierarchically super-imposed syntactic/prosodic units which are carefully defined. In large conversational corpora, pauses are typically segmented as silent intervals of some minimum duration (Jefferson 1974, 1989). A variety of so-called cut-off points have been applied to isolate word searches, from "predisagreements," or other hesitation phenomena, or pauses signaling turn-completion (Goldman-Eisler 1968; Jefferson 1974, 1989; Carlson, Gustafson, and Strangert 2006), with pause durations now easily measured with your software package.

There is now an entire bibliography of references to substantiate claims that conversationalists slow down or speed up their speech rate in order to gain or retain access to the conversational floor (e.g., Schegloff 1996, 1998). While slowing speech rate may only enhance perception of possible breaks and boundaries, speeding up is often used specifically to maintain the floor by avoiding pauses which signal turn completion. Benus et al. (2006) found that the use of silent and filled pauses correlates more with truthfulness than with intended deceit, inferring that—at least in experimental situations—people tend to monitor their speech more carefully when lying. The conversational fine-tuning of the use of silent and filled pauses and the adjacent prefinal lengthening, not to mention techniques like "rush throughs" and "abrupt joins" (Local and Walker 2004) demonstrates again that while conversational material may not be ideal for rhythm studies, the analysis of conversation remains an almost untapped resource for sociophonetic analysis.

Pitch and intonation

This section will sketch out analytical techniques for empirical work on intonational cues for social-demographic (dialect, gender, ethnicity) membership as well as conversation analytic work.

The rate of vibrations of the vocal folds, measured as *fundamental frequency* (F0), is perceived as *pitch* (Lehiste 1970; Lehiste and Peterson 1961a); the term *intonation* refers to variation in a speaker's fundamental frequency, organized in grammatically meaningful *peaks* or *melodies* within some larger phrasal domain. Until fairly recently, perceptual evidence was considered sufficient for such studies, but more recently, most analyses are based directly on acoustic evidence, although often the terminology used implies a pitch-based/perceptual analysis. Thus variations in a speaker's fundamental are typically referred to as being within that speaker's characteristic *pitch range*, i.e. the degree of systematic deviation from the average pitch of their voices.

Two larger prosodic phrases—the intonation phrase and the utterance—correspond to the domain of *intonation contours*. *Falling* and *rising tonal contours* refer to the direction of F0/pitch movements at the end of a phrase. First let us note how earlier studies have used fundamental frequency (F0) data, and then review techniques used to measure and code F0/pitch variation.

Average pitch and pitch range

Like analysis of duration, fundamental frequency is increasingly simple and reliable with most software packages. (To be discussed further in chapter 11.) Obviously, one distinction that everyone recognizes is the fact that men have, on average, lower F0, and narrower pitch ranges, than women, and that children have higher F0s and broader ranges than either (Henton 1985, 1989). While mean F0 is said to be primarily limited by physiology—that is, the size of our vocal chords (Ladefoged 2005a), F0 range has also been correlated with gender identity, ethnicity, and region. Comparing men's and women's speech in the US and England shows, for instance, that American men appear more likely to manipulate their F0 to "index," say, manliness or "toughness" than an equivalent group of British speakers (Henton and Bladon 1985, 1987; Yuasa 2008).

Podesva (e.g., 2007) has documented the use of falsetto and broader prosodic contours by American males not only to index their sexual identity, but to adapt their use of prosodic features to reflect the relative importance of their gender identity to the interaction at hand. Japanese women use falsetto systematically to index politeness, or girlishness (L. Miller 2004), while a woman's avoiding this "burriko" use of falsetto is said to index sexual identity as well (Cameron and Kulick 2003; Camp 2009).

Most speech software will now track speakers' F0 automatically and determine the vocal *range* of speakers over the *span of speech chosen by the analyst* (see website). F0 and range seem transparent, but one must attend to variation across the entire speech-range. For example, Goodwin, Goodwin, and Yaeger-Dror (2002) found that in arguments that arise during game-playing behavior, Los Angeles Latino pre-teen girls' F0 varied across a very wide F0 range fairly evenly (as in Figure 10.1a, "You out!" from that paper), while

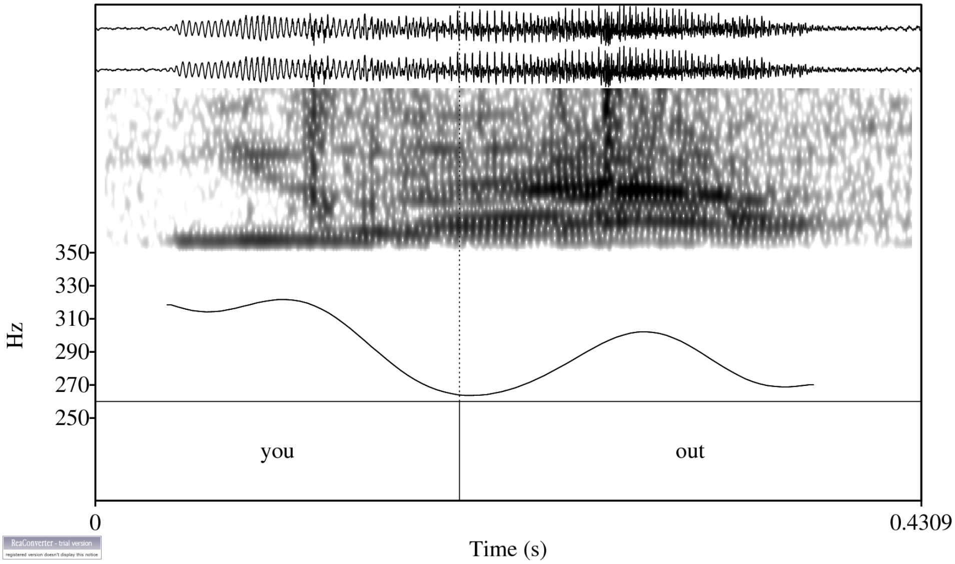

Figure 10.1a Latina girl calling "you out" during a game on the playground (from Goodwin, Goodwin, and Yaeger-Dror 2002). The range on both words is from 320 Hz (near the speaker's falsetto range) to 260 (in the lower part of the speaker's pitch range), back to 300 Hz and down again to 270 Hz.

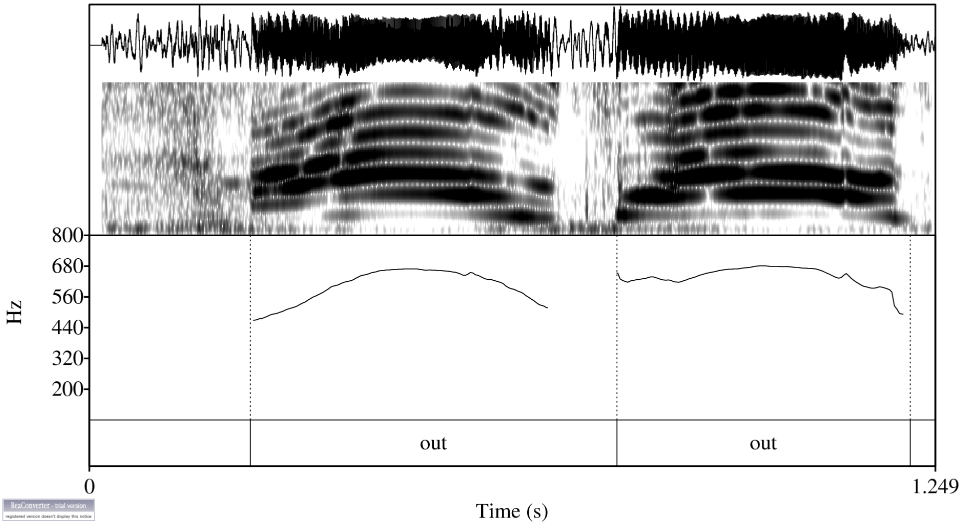

Figure 10.1b African American girl calling "out out!" during a game on the playground (from Goodwin Goodwin, and Yaeger-Dror 2002). The F0 on both words peaks at 680 Hz, which is in the speaker's falsetto range.

in an equivalent corpus of African American pre-teens, there was a rather narrow range of F0, but realizations of narrow focus were in falsetto (as in Figure 10.1b, "out, out" from that paper), giving an automatic pitch-tracker the erroneous sense that the two groups' pitch range(s) were fairly equivalent (Goodwin, Goodwin, and Yaeger-Dror 2002). Future studies of F0 range should adapt the program for the automated analysis of F0 range to take this into consideration. (Compare Figures 10.1a and 10.1b and their sound files which can be found on the book website, URL = XXXXXXXXXXXXXXX.) Such measurements are used for analysis of gender or ethnic indexing, or for analysis of speaker-displays of excitement, "deception" (Enos et al. 2007), or other emotion (Gravano et al. 2007).

Tunes and tones

There has been much debate whether intonation should be analyzed in terms of tunes (Liberman and Sag 1974) or local F0 minima (valleys) and maxima (peaks) (Bolinger 1978). While most American approaches to intonation currently favor the phonological analyses of individual and sentential contours (Hirschberg 1994/2008; Jun 2005; Ladd 2008), both are represented in recent sociophonetic studies. All studies now rely on both auditory and acoustic analyses of F0 movements in hypothesis formation and segmentation of the sound files into perceptually relevant F0 events. For any analyst, listening to the tracked contour is indispensable, but there is ample evidence to prove that the best researchers can be fooled if this "listening" isn't backed up by acoustic measurement.

F0 movements associated with strong syllables are referred to as *pitch accent*, while those associated with superimposed prosodic phrases are referred to as *boundary tones* (Shattuck-Hufnagel, Brugos, and Veilleux 2007). These papers are cast within a specific theoretical framework referred to as ToBI, to be discussed below. If you wish to adopt—or respond to—the intonational phonology theory you must be conversant with this framework and its tenets.

An influential line of research in English intonation gave rise to the labeling system referred to as MAE-ToBI, i.e. mainstream English tone and break indices (Beckman, Hirschberg, and Shattuck-Hufnagel 2005), which was originally developed from the seminal work of Pierrehumbert (1980). Her model has been elaborated in numerous subsequent studies and partially reconceptualized (see Jun 2005) for other languages, with a rich body of literature analyzing its efficiency in labeling large speech corpora (e.g., Syrdal et al. 2001).

The working hypothesis within ToBI studies of intonation is that each dialect has its own phonological intonational system that also contains fine-grained phonetic differences in alignment and pitch range variations (cf. Post, D'Imperio, and Gussenhoven 2007). Among the labeling conventions used in the papers are MAE-ToBI for American English (Shattuck-Hufnagel, Brugos, and Veilleux 2007), Sp-ToBI for Spanish (Jun 2005; Estebas-Vilaplana 2007), G-ToBI for German (Queen 1996), J-ToBI for Japanese (Venditti 2005) and I-ToBI for Italian (D'Imperio 2002). In each case, the speech materials are transcribed by teams of trained researchers, using symbols corresponding to different pitch contours associated with the metrically most prominent, primary stressed, syllable in a phrase. For each language, the ToBI researchers have proposed which *nuclear pitch accents* (symbolized with *) can express emphasis, and what forms the (syntactic) boundary tones can take. Within that model of intonational phonology, discrete prominent F0 and boundary tone labels are assigned, and inter-coder reliability is checked. The notational conventions of this system will not be discussed in detail here, but the interested student should follow the online tutorial (Shattuck-Hufnagel, Brugos, and Veilleux 2007). Both this chapter's website and the Shattuck-Hufnagel tutorial provide examples of common focal contours. These, so-called, *pitch accents* can be "simple" or "complex," and correspond to a local peak (H*) or valley (*L) or a combination of the two (L + H*), marked by one or two capital letter symbols reflecting the prosodic pattern expressing focal stress. In addition, four *intonational phrase* (IP) *boundary tones* are used to label major discontinuities in the speech signal. These various conventionalized transcriptions can then be used for a rough and ready comparison between the intonational patterns of different speech communities. Future research will tell how helpful these models have been to the study of socio-prosodic variation.

The ToBI system was devised using a sample corpus of news items read by NPR professionals. Later studies have shown that reliability ratings for perceived tones are variable (Syrdal et al. 2001), and that the original corpus contained patterns that had been initially considered ungrammatical for American English (Shattuck-Hufnagel, Brugos, and Veilleux 2007). However, ToBI transcription conventions can be useful for coding, if you do not lose track of its limitations, and supplement the coding options as needed.

Sociolinguistic applications

Conversation analytic and sociophonetic studies both generally adopt the view that F0 variation is continuous; most conversational analytic approaches refer to intonation holistically (cf. Couper-Kuhlen and Selting 1996, Selting 2003), others code for specific tone-tunes (cf. Golato and Fagyal 2008). Nevertheless, ToBI conventions are often followed: Considerable work has focused on defining geographical variations in intonation (e.g., Jarman and Cruttenden 1976; Gussenhoven and Aarts 1999; Sosa 1999; Kaminskaia and Poire 2004; Estebas-Vilaplana 2007; Post, D'Imperio, and Gussenhoven 2007), whether of the production of various intonational contours (Grabe et al. 2000; Grabe 2004; Gilles and Peters 2004; Kügler 2004), or of listeners' ability to use two dialectal systems of intonation (Cruttenden 2007), or distinguish various dialects based on intonation alone (Bezooijen and Gooskens 1999; van Leyden and van Heuven 2006).

Comparative prosody

Just as *adstratum*, or *substratum* influences have often been hypothesized as the cause of PVI variation, many Latin American dialects have phrasal or sentential "tunes," which can be markers—or even stereotypes—of the region they come from (Sosa 1999). Often local accents are assumed to derive from a substratum Indian language; unfortunately, in most cases, no evidence remains to support the theory. Similarly, while a number of studies have assumed African American English intonation might be usable as an ethnic marker, no tangible results have come from such studies (Thomas and Carter 2006).

Studies of language contact prosody are on firmer ground, e.g., Portuguese–German bilinguals (Birkner 2004), Latin Americans (Simonet 2008) English–Spanish bilinguals in Gibraltar (Levey 2008), Catalan–Spanish contact (Simonet ms), Quechua–Spanish bilinguals in Peru (O'Rourke 2008), Turks in Germany (Queen 1996), and North Africans in France (Fagyal 2005).

Intonation in interaction: high-rising terminals (hrt) or "uptalk"

A number of studies have documented pragmatic considerations for dialectal intonational variants: since the 1980s sociophoneticians have considered the pragmatic and regional variation in rising terminals (often referred to as "uptalk," or "HRT" for high rising terminals[2]) used in statements. These high rising terminals have been interpreted as a feminine-indexed display of insecurity (Lakoff 1975), of solidarity (McLemore 1991, referred to in Liberman 2010), or as a regional tendency (Guy et al. 1986; Britain 1992; Warren 2005a, b; Arvaniti and Garding 2008). If HRT occurs in the speech community you are studying, you have a variable that can be easily/reliably coded (perceptually or visually) and analyzed. Even in this case, further research is needed to connect detailed phonetic analysis of prosody with careful documentation of ethnographic observations and pragmatic functions.

From impressions to coding

This section will emphasize acoustic techniques for the analysis of F0. There is increasing evidence that even for highly trained listeners one minute of speech takes at least 20 minutes for one listener to code perceptually, and as we have seen, the results are even then not very reliable (Syrdal et al. 2001). Thus, while acoustic analysis is not trouble free, we agree with Wightman (2002) in suggesting a combination of acoustic-phonetic and auditory analyses of pitch movements. We assume that the reader is acquainted with at least one of the speech analysis programs with a usable "pitch tracker"; this section will discuss the analysis of variation in F0. Since the programs used for this prosodic analysis are frequently upgraded, our discussion of which programs are more likely to provide accurate measurements can be found on the book website under "Analysis of pitchtracks." At this time, most of the programs available are equivalent, but the reader interested in more detail on the type of acoustic analyses performed should study the plethora alternatives for speech analysis in acoustics or the speech and hearing sciences.

As discussed, acoustic correlates of lexically distinctive stress in American English are generally manifested as increased amplitude and duration on the stressed vowel (Klatt 1976) and often, particularly if the word is under focal stress, there are also specific F0 variations that can be coded with MAE-ToBI symbols or other ways (see Shattuck-Hufnagel, Brugos, and Veilleux 2007), and which are conventionally assumed to be the primary marker of stress. Sentential and conversational prosody are superposed on this pattern. Yaeger-Dror et al. (2010) found that for even a minimally distinguished set of choices for conversational English, ToBI codes had to be supplemented. For remedial intonation they found that while the pragmatic choice—of whether to focus on a disagreement—varied in different regions of the country, there was no significant regional difference in choice of prominence contour (x*) in American English. The prosody-pragmatics interface has also been explored in greater detail outside laboratory phonology. With recent improvements and greater access to acoustic software, many new research topics have been explored on the prosody of voicing suggestion and complaints, and strategies of evasion in parliamentary discourse (Wichmann and Blakemore 2006).

A potential trap for a student coder arises when the contour seen on the screen does not match the actual F0. An ideal program could track fundamental frequency accurately whether the speaker is a man [F0 between 80 and 200 Hz], a woman [between 130 and 300 Hz], or a child [200 Hz and above], even without the researcher setting the parameters; but the programs available actually need you to set the parameters for pitch-tracking range, since analytical software can often double or halve the fundamental frequency of a speaker's production. Listen carefully **before** looking at the pitch track, and if what you hear is not reflected in what you see, change the settings used for the analysis to reflect the speaker's range. In addition, if you are using a system which requires you to bundle information on the fundamental frequency [F0] with information on amplitude and duration, it is fairly easy to determine if F0, prominence (amplitude) and duration are being used in tandem (as is most common); if not, this should be noted. Intonation contours deployed over major phrases convey information about the speaker's orientation toward the meaning of utterances, and toward interlocutors.

As discussed earlier, your preferred software can also be programmed to give a readout of maximum, minimum, and mean F0, as well as an estimated reference F0 (the frequency toward which a talker's F0 tends in the absence of tonal accents), and range. Such data should be checked carefully by visual inspection of the waveform:

- Make sure a given sample being analyzed has only one speaker talking.
- Correct doubled or halved measurements.
- Contrast radical falsetto displacement on a single word, and continuous variation across the speaker's range.
- Make sure your analyzer isn't averaging in zeros for devoiced segments, falsely lowering the "mean."

If analysis were not automated, F0 estimates for each 15 ms frame would be done by hand: As with the automatic alignment, LPC, or other automated programs, checking that the "pitch tracker" program did its job is much less time consuming than doing all the work yourself!

Given that telephones generally cut off the range below 20 kHz, researchers often question the ability of a pitch tracker to accurately recognize the fundamentals of men with low F0s, when the initial recording is telephone data; this problem might not be important if you have carefully chosen your own corpus, and each speaker has his or her own [ideal] microphone; however, as discussed in an earlier section, if you are interested in discourse (and therefore in intonation), it is often difficult to acquire a large "natural" conversational corpus. Telephone corpora have decided advantages:

1. Each speaker has his/her own microphone.
2. What you see is what you've got—that is, there is ample evidence that in conversation speakers supplement their speech with gaze, and other nonverbal signs (Goodwin 1981; Massaro 1998; Beskow, Grandstrom, and House 2006; Massaro et al. 2008). However, in telephone conversations, all communication is verbal, and available to the researcher.
3. LDC has several "parallel corpora," where the same speaker is talking by phone to different interlocutors (mixer), or where speakers from a given social group, or language community, are speaking to friends, permitting you to compare the communicative behavior of different groups in the same social setting (call home, call friend). Over the years, these LDC corpora have been used for several studies, and despite the fact that the speakers are using telephone microphones, pitch trackers have been used effectively. The only caveat here is that you not analyze the minor differences between speakers' "basic pitch": it is now known that different phone types (land, wireless, cell) alter the signal; while variation over the course of the interaction can be gauged reliably, absolute pitch should not be compared from one phone type to another even for a single speaker (Harrison 2008).

It is possible to transcribe the speech directly into a tier of the program (as you saw in chapter 4). Multiple coding tiers permit automated analysis of multiple-coder reliability. As mentioned in the transcription chapter, it also simplifies the search for outliers to determine whether they are caused by coder error, or provide evidence for a previously unnoticed pattern. Following the advice in this chapter will hopefully facilitate future research on the social aspects of phonetics of prosody. Given evidence that prosodic variation entails fine-tuned prosodic interactive patterns, as well as accommodation to the speech of others, it is also our conviction that, as acoustic trackers become more sophisticated, future analyses will incorporate interdisciplinary collaborations as well.

Exercises

1. A. Go to the webpage http://ocw.mit.edu/OcwWeb/Electrical-Engineering-and-Computer-Science/6-911January--IAP--2006/CourseHome/index.htm; download Exercises 2.1 and 2.9, and do the assignment once just by ear, and once with the pitch tracker turned on. Does your analysis without benefit of the pitch track support one

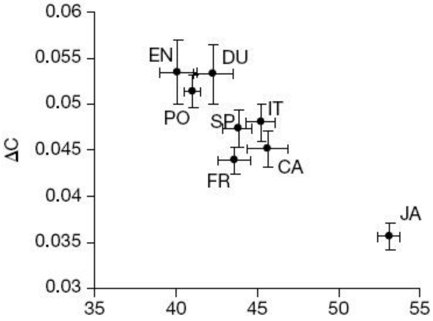

Figure 10.2 This figure provides information from Ramus, Nespor, and Mehler (1999: 273) demonstrating evidence from nPVI studies showing that languages cluster in different rhythm types.

specific analysis, or is there more than one analysis available to you? How different is your analysis when you have the acoustic evidence before you?

B. Go to the book website. Download "sound bytes" 1–4. Provides analyses both by ear and by eye for these. What conclusions can you draw from your coding of the four segments?

2. Consider Figure 10.2, extracted from the Ramus, Nespor, and Mehler (1999: 273). Answer the following questions: (a) What measure should be placed on the *x* axis and why? (b) What rhythm classes can you distinguish, and what phonetic/phonological phenomena are the two measures trying to capture? (c) What would be the expected rhythmic characteristics of a language, if any, placed on the bottom left of this figure?
3. What wide-spread intonation contour does the pitch track on Figure 10.3 exemplify and why? The words are: "up the very top on the left-hand side" (from Warren 2005a: 211). (The sound file is available on the book website.)
4. Study the following pitch track of an excerpt downloaded from the Linguistic Data Consortium's "Language log" website (http://languagelog.ldc.upenn.edu/nll/?p=568). Why is this contour a good example of what is customarily referred to as "HRT" in American English? (The sound file for this pitch track can be found on the book website.)

Notes

[1] Much less the equally meaningful and oft-studied "filled pauses."

[2] See however, Liberman's (2010: http://itre.cis.upenn.edu/~myl/languagelog/archives/002967.html) discussion of the terminology.