9 Prosody

This chapter deals with prosodic features of speech of the kind commonly described as tone, stress and intonation. After a general introduction (9.1) the chapter explains the chief phonetic correlates, notably pitch, duration and loudness (9.2).

Prosody may be systematized in various ways in language (9.3). Many of the world's languages can be described as tone languages (9.4); a less common type of systematization can be found in so-called pitch-accent languages (9.5).

The latter part of the chapter focuses mainly on English and discusses

- the phenomenon of lexical stress (9.6)
- the extent to which stress patterns are governed by rules (9.7)
- the description of intonation (9.8)
- the outline of a particular model of English intonation (9.9).

9.1 Introduction

Features of spoken language which are not easily identified as discrete segments are variously referred to as PROSODIC FEATURES, NONSEGMENTAL FEATURES or SUPRASEGMENTALS. The terms imply a difference between segmental sounds (traditionally consonants and vowels) which are commonly thought of as entities, and features such as pitch, rhythm and tempo which are likely to be perceived as features extending over longer stretches of speech. The distinction is reinforced by many writing systems (including that of English) which have an alphabet of consonant and vowel symbols but no comparable indication of prosody, other than through the use of punctuation marks and devices such as italicization. The distinction is by no means clear-cut, however, and the term 'prosody' has no single definition. 'Prosody' or 'prosodic features' are often used synonymously with 'suprasegmental features'. These features *can* be just as discrete as consonants and vowels; and, as we have noted in looking at the details of acoustics and articulation, consonants and vowels are not always identifiable outside the context of speech in which they appear. Moreover, the implication that *supra*segmentals are somehow superimposed on a basic message

of consonants and vowels is decidedly misleading, given that patterns of pitch, loudness and tempo are an integral part of speech production and often a fully meaningful contribution to the message itself. After all, no one utters stretches of English consonants and vowels in an absolutely even-measured monotone – or if they do, the result is perceived as highly marked speech, perhaps as a comic affectation of extreme boredom or as an imitation of a robot. Pitch or timing patterns tend, much more than consonants and vowels, to be directly related to higher levels of linguistic organization, such as the structuring of information, so according to this view of prosody, one cannot always readily separate these features from other long-term settings and adjustments, such as voice quality and rate of articulation. On the other hand, suprasegmental effects are often intertwined with the production of individual consonant and vowel segements, which leads us to the second definition of prosody as a set of higher-level organizational structures that account for variations in pitch, loudness, duration, spectral tilt, segment reduction and their associated articulatory parameters. This is a more abstract definition of prosody and is very much related to developments in prosodic phonology (see section 11.17 below). Beckman (1986, 1996), Shattuck-Hufnagel and Turk (1996) and Gussenhoven (2004) discuss these two interpretations of prosody in some detail.

We need to take into account here a continuum of functions and effects, ranging from the nonlinguistic or extralinguistic at one end, through the paralinguistic, to the essentially linguistic. At the nonlinguistic end, for example, are features of voice quality that reflect the nature of the speaker's larynx and vocal tract; at the linguistic end are features such as lexical stress and tone, postlexical prominence or accentuation, and intonation, which are functional within specific linguistic systems and often vary widely in their systematization from language to language. But note that the term PARALINGUISTIC points to a grey area in between the two reasonably uncontroversial extremes: it is not at all easy, for instance, to determine whether a particular style of speech delivery is an unconscious habit, perhaps related to the speaker's anatomy or physiology, or a deliberate – and therefore communicative – attempt to project a certain personality. An obvious example is the effect of nervousness. We are all familiar with certain features of speech that tell us that the speaker is nervous, often despite the speaker's best efforts to disguise the nervousness; but it is also possible for a speaker to adopt some of these features deliberately, whether as a long-term style of speech, or to gain sympathy on a particular occasion. Readers will probably be able to think of various comparable examples, noting their ambiguous status – the affected stammer, the characteristic languid drawl or the constant nervous giggle, for instance.

Features at the least linguistic end may be thought of as a substratum of underlying properties of the signal, including the phonation quality determined by the anatomy and tensions of the larynx, the pitch range determined by management of the larynx, and long-term articulatory settings such as tongue root posture and articulation rate, among others. Some of these factors may be a function of anatomy, others may be acquired as habitual characteristics. While they are generally not considered to form part of the functional system of language – since they do not reflect genuine options exercised by the speaker

– they may still be communicative to the extent that they enable us to identify a particular speaker or type or class of speaker. (See Laver 1980, pp. 3–7, for comments on the way in which features such as nasality and phonation quality may characterize social or regional groups.)

Physical illness such as respiratory tract infection may of course have a profound and pervasive effect on speech, and emotions such as fear or anger may alter pitch range, phonation quality, loudness and articulation rate. Laughter, sighing and sobbing may all cause complete interruption to the normal processes of articulation. These phenomena are in one sense common to all human beings and outside language, but they are by no means beyond functional control. English speakers often use a deliberate laugh or sigh as a meaningful comment or reaction, and, in general, the conventions governing such behaviour as laughing and sighing vary sharply among societies and social settings.

More closely related to linguistic functioning is the speaker's use of such features as overall voice quality, pitch range, pitch movement and articulation rate to indicate a general attitude. Again conventions vary widely, but it is probably safe to say that most speakers in most languages have ways of signalling authoritativeness or submissiveness, seriousness or lightheartedness, excitement or calmness, even though these states or attitudes will certainly not be identical across cultures.

Strategies such as tempo may also be used to demarcate stretches of speech within discourse. Crystal, for example, refers to the way in which accelerated tempo may serve to indicate an embedded phrase or clause in English (1969, pp. 152–3). Note the underlined sections in the following examples:

It's one of those reply-immediately-in-five-lines-or-less memos!
Is the How to Write English course on this year?

In examples such as these, a clause constituting a unitary description or title may be spoken distinctly faster than the rest of the utterance, as a signal of its shifted status. This mechanism may be coupled with pauses at the boundaries of the unit. The functions of tempo and pause here are clearly linguistic, as much part of the grammar as any other device signalling relations among clauses.

At the most linguistic end of the continuum are systems such as stress, intonation and tone. Typically, phonetic features of pitch, loudness and duration (and possibly spectral tilt) are relevant here, contributing to the organization of discourse and even to lexical distinctions. Since languages vary in their systematic exploitation of these features – and are often categorized accordingly, as 'tone languages' or 'intonation languages' and so on – we shall return to these systems in more detail below (sections 9.3–9.8).

While the continuum described above indicates the range of information carried by the speech signal, it leaves an incomplete picture of what is really happening, for all kinds of information may be encoded simultaneously in the speech signal, and listeners abstract what they judge relevant to the communication context. Note, for example, that listeners may be perceiving emphasis on certain words or phrases, and responding to the speaker's organization of the message, at the same time that they are forming a judgement about the

speaker's regional background and emotional state. Communication is also inherently interactive. There is after all some reason for speaking, and usually a listener or audience, even if at some remove via a telephone or other such communication system. Even in public speaking, the audience is a collective listener whose reactions (real, imagined or anticipated) are likely to influence the speaker. Thus both speakers and listeners operate in a context of (largely) shared assumptions about the significance of prosody, ranging from expectations about the effects of tiredness and nervousness and sore throats, and conventions about appropriate levels of loudness and speeds of delivery, to knowledge of systematic ways of structuring discourse.

In summary, a simple taxonomy of phonological features, isolated from contextual function, is not enough to account for linguistic prosody, and much of what is written about suprasegmental phonology provides only a point of departure for analysis. Here there is some justification for a distinction between segmental and suprasegmental phonology, in that the phonetic resources underlying segmental distinctions can be more easily and directly related to indisputably linguistic organization and are more amenable to taxonomic treatment.

Against this background, it is not surprising that much of the traditional literature on the analysis and description of suprasegmentals has tended to concentrate on generalized abstractions (for example, about typical intonation melodies or the functions of tones) rather than on the complex and highly variable phonetic detail.

Since the late 1980s, however, there has been great interest in describing the phonetic realization of abstract prosodic structure and its interaction with the realization of consonant and vowel segments, particularly within the crossover discipline of laboratory phonology (section 11.18 below, and see e.g. Beckman et al. 1992, Beckman and Cohen 2000, Tabain and Perrier 2005, Keating in press). Research in information technology has also stimulated much closer scrutiny of phonetic aspects of the suprasegmental structure of the speech signal. A notable example is the strong interest in the development of intonation models (e.g. 't Hart 1979, Pierrehumbert 1981, Taylor 2000) and durational rule models (e.g. Klatt 1979, Campbell and Isard 1991, Local and Ogden 1998) for use in text-to-speech systems. Van Santen et al. (1998) is a useful set of papers covering major developments in these areas in speech synthesis research.

Laver (1980) deals with voice quality, taking a broad view of what is involved and commenting helpfully on the problem of deciding what is or is not part of language. Kreiman et al. (2005) and Kreiman (1997) also outline the ways in which voice quality contributes to many aspects of speech and language behaviour. Detailed discussion of the ways in which phonetic resources are used for 'affective' and 'attitudinal' functions, with some spectrographic analysis, can be found in Crystal and Quirk (1964). Crystal (1969) offers a thorough taxonomy of English prosodic features in the context of a wide-ranging survey of the traditional literature, including a useful review of the linguistic status of prosodic and paralinguistic features (pp. 179–93). Pierrehumbert (1980), Ladd (1996) and Gussenhoven (2004) present an overview of modern phonological approaches to the study of intonation. Wichmann (2000) gives a detailed account of the relationship between intonation and discourse, and Hirschberg

(2002) is a useful summary of the contributions of different aspects of prosody to spoken communication, with particular reference to applications like spoken dialogue systems. For a general overview of prosodic features and the issues raised in this section, see Cruttenden (1997, esp. chs 1 and 6), Gussenhoven (2004, esp. chs 1 and 2) and Jun (2005a, esp. ch. 16).

9.2 The phonetic basis of suprasegmentals

The principal phonetic correlates of the more linguistic aspects of prosody and prosodic structure are traditionally thought to be the dynamic patterns of pitch, duration and loudness, although vowel quality and possibly spectral tilt are also phonetic correlates of stress in some languages, such as Dutch (Sluijter 1995). The three suprasegmental parameters, pitch, loudness and duration, are both overlaid on, and influenced by, the less dynamic substratum of voice quality as determined by the state of the vocal tract. These dimensions of the speech signal, interacting with each other and with the segmental structure, are fundamental to our perception of emotion, attitude and other such information conveyed in speech.

VOICE QUALITY and VOCAL TRACT STATE are treated by Laver (1980), who uses the concept of 'articulatory settings' of the vocal tract (pp. 12ff.). These long-term settings are the underlying articulatory positions or postures upon which all the dynamics of articulation – both segmental and suprasegmental – are superimposed. The settings have articulatory – and hence acoustic – consequences which pervade the whole stream of speech.

Very importantly, Laver notes that the time domain of these settings may vary. A setting contributing to personal voice quality may be for all practical purposes a permanent feature, while another setting may be controlled contrastively. Thus a speaker may talk with habitually rounded and slightly protruded lips, which will influence the overall frequency range of formant patterns and be judged part of his personal characteristics. On the other hand, laryngeal tension settings may be changed to produce a voice quality associated with a particular attitude: many speakers of English, for example, may use vocal creak to indicate boredom or dismissiveness.

Laver's system has two basic divisions: supralaryngeal and phonatory settings. Supralaryngeal settings describe the vocal tract state longitudinally (larynx height and labial protrusion) and latitudinally (labial, lingual, faucal, pharyngeal and mandibular settings); they also include velo-pharyngeal settings, affecting the coupling of the nasal tract and perceptions of nasality. Phonatory settings describe phonation types relative to normal or modal phonation, and allow for compound phonation types.

Acoustically, supralaryngeal vocal tract settings are reflected primarily in formant distribution. Thus, a raised larynx may shorten the vocal tract and raise formant frequencies, although not always in a simple linear relationship to larynx height, The speech production model of Lindblom and Sundberg

(1971) provides a theoretical basis for estimating some of the acoustic correlates of the supralaryngeal settings defined by Laver. The acoustic properties of phonatory settings are rather more difficult to establish independently, since phonation provides the excitation source for the vocal tract and is therefore always modified by the current vocal tract filter function. Special measurement techniques do exist for cancelling out the effects of vocal tract resonance, such as the reflectionless tube (Sondhi 1975) and computer-based antiresonance filtering (section 7.11 above). The overall effects of changes in phonatory setting can be seen in the spectral slope of voiced speech spectra, and in the degree of periodicity of phonation, as revealed in speech spectrograms. A general measure of long-term vocal tract setting differences can also be obtained from long-term spectrum measurements of the kind described in section 7.19 above. These measurements will show average changes in the energy distribution of the speech spectrum caused by both phonation and vocal tract settings. (See chapter 2 for the phonetic background to these settings, and chapter 7 for the relevant acoustic information.)

PITCH is widely regarded, at least in English, as an important cue to intonational prominence, often referred to as sentence stress or phrasal stress. This kind of prominence is also called ACCENT (see section 9.3). In other words, when a word is perceived as accented relative to surrounding words in English, it is usually pitch height or a change of pitch associated with a lexically stressed syllable in that particular word that is mainly responsible. The stressed syllable sounds more prominent because it is also associated with an intonational PITCH ACCENT (e.g. Bolinger 1958). Earlier studies (e.g. Fry 1958) claimed that pitch, and to a lesser degree loudness, length and vowel quality, were the major phonetic correlates of stress. Cutler (2005) provides a good summary of earlier experimental analyses of stress, including one by Lieberman (1960) where he claimed that no single cue (i.e. pitch, loudness or length) can be correlated with stress (in English); rather there is often a trade-off between the different acoustic correlates. However, many of these early studies did not take into account that citation words under investigation were produced as a full intonational constituent, and so the stressed syllables in many of these studies were also accented (see Bolinger 1958, Beckman 1986, Ladd 1996, Pierrehumbert 2000 and Gussenhoven 2004 for a discussion of this). It is important not to confuse different kinds of prosodic prominence in languages like English. We will outline the differences between lexical stress and postlexical intonational prominence in sections 9.3 and 9.8.

Pitch is the perceived correlate of fundamental frequency. It is commonly measured on the mel scale, or less frequently on the ERB (equal rectangular bandwidth) scale, since changes of perceived pitch are proportional to, but not the same as, changes of frequency (section 7.9 above). Fundamental frequency (F_0) – the number of times per second that the vocal folds complete a cycle of vibration – is controlled by the muscular forces determining vocal fold settings and tensions in the larynx, and by the aerodynamic forces of the respiratory system which drive the larynx and provide the source of energy for the phonation itself (sections 6.4 and 7.11 above). It has been argued by Lieberman (1967) that aerodynamic forces, specifically subglottal pressure (Psg), are primarily

responsible for pitch control and that laryngeal adjustments are a secondary or alternative form of control. He maintains that Psg patterns have an archetypal shape in utterances, and that Psg variations are superimposed on them. Ohala (1970) refutes Lieberman's evidence and (1978) reviews the 'larynx versus lungs' controversy in general. It seems that for the majority of languages, laryngeal adjustments are primarily responsible for pitch control. In particular, the cricothyroid muscle is always active during pitch raising by its direct tensioning of the vocal folds. Vertical movement of the larynx, controlled by its extrinsic strap muscles, correlates well with corresponding rises and falls in pitch. In general, pitch raising is better understood than pitch lowering, which appears to involve relaxation of the cricothyroid muscles, and contraction of the infrahyoidal strap muscles (Erikson et al. 1983). In the lowest portion of the pitch range, these mechanisms seem to be supplemented by other muscles such as the lateral cricoarytenoid and the thyroarytenoid and vocalis, which shorten, slacken and thicken the folds. Furthermore, Sagart et al. (1986) have found that the sternohyoid muscle also contributes to pitch lowering.

Although less significant than laryngeal muscle action, Psg does show a positive correlation with pitch movement. Data for English suggest that Psg is responsible for about 5 to 10 per cent of the total range of pitch change in normal speech. But it is not clear how far this generalization extends to all languages: in at least some dialects of Chinese, for instance, it does seem that Psg provides the primary form of pitch control (Rose 1982). Detailed discussion of pitch regulatory mechanisms can be found in Sawashima (1974), Ohala (1978), Hollien (1983) and Gussenhoven (2004), and there are relevant data in Ladefoged (1967).

Our ability to discriminate pitch has been investigated in various studies, many of them focusing on the threshold of minimal perceivable difference, or DIFFERENCE LIMEN (DL). A change in pitch of as little as 0.3–0.5 per cent may be perceivable, at least in vowels synthesized to simulate a male voice (Flanagan 1972). Studies by 't Hart (1981) and Harris and Umeda (1987) show that the DL may be substantially higher in running speech, the actual value depending on the average fundamental frequency, the speaker and the complexity of the speech signal concerned. Rietfeld and Gussenhoven (1985) also report data suggesting, surprisingly, that perceptual judgements of the magnitude of prominence tend to match frequency values rather than a pitch scale.

DURATION as a property of sounds or units cannot be separated from the larger context of time and timing in speech production. The duration of individual speech segments varies enormously, depending on both segment type and the surrounding phonetic context. A vowel, for example, may last 300 ms or longer, while the release of a voiced stop may be only about 20 ms. Duration is also constrained by biomechanical factors: part of the reason why the vowel in English *bat*, for example, tends to be relatively long is that the jaw has to move further than in words like *bit* or *bet*.

In the context of prosodic distinctions, overall syllable duration is more important than segment duration, and relative duration more important than absolute duration. Vowel duration is obviously the most significant component of syllable duration, but maintenance of appropriate durational relationships

within the whole structure of the syllable is very important if segmental relationships and distinctions are to be preserved.

Overall syllable duration is influenced by many contextual factors. These include the rate of articulation, the placement of prominence or stress, the position of the syllable within a word or other larger unit, and the structure of those larger units themselves. Although syllable duration is quite elastic – and the actual duration is an important contribution to the perceived prominence of the syllable – not all components of duration are equally elastic. Many researchers note that vowels are more compressible than consonants (Klatt 1976, Crystal and House 1988). Classic acoustic studies of vowel target reduction and undershoot by Lindblom (1963), Stevens and House (1963) and Stevens et al. (1966) have shown that as syllable length is reduced, consonant transitions tend to be preserved at the expense of vowel target length (although not absolutely so). Consonant durations vary with the number of consonants in the syllable, and are also influenced by overall syllable duration. Campbell and Isard (1991) and Greenberg et al. (2003) review the general properties of durational structure in syllables. The way in which the temporal components of the syllable can be varied differentially is shown in the phonological rules set out by Allen et al. (1987), where the variability is expressed quantitatively for a text-to-speech system. Campbell and Isard (1991) show how a syllable-based timing template can be modelled for speech synthesis of English.

The way in which each language exploits durational relationships within the syllable for phonological purposes will also influence its internal temporal structure. In English, for example, vowel length is substantially increased when the vowel is followed by consonant voicing, and the length of the vowel becomes a significant perceptual cue to the voicing contrast (see, for example, Lisker 1978, p. 134). The effect is also clearly seen in the data from a fricative consonant study by Clark and Palethorpe (1986), and in stop consonant studies by de Jong (1995). There are other examples from earlier experimental studies reviewed by Lehiste (1970).

It is important to note that the way in which rhythmic structure and stress placement are integrated in a given language will also influence duration patterns. In languages such as English and Dutch, stressed syllables are generally much longer than unstressed (see, for example, Lieberman 1960, Gay 1978, Beckman 1986, Summers 1987, Sluijter 1995). In languages traditionally classified as tone languages, such as Thai, there may be an interaction between tone and vowel duration in that vowels carrying rising tones are generally longer than vowels carrying falling tones, all things being equal (e.g. Gandour 1978). Finally, we should not forget silence: pauses are an important ingredient of our total communicative resources. Cruttenden (1997, pp. 30–2) gives a useful overview of pause types and pause function in English, noting the role of pauses in signalling structural boundaries as well as what are usually called 'hesitation phenomena'. Allen et al. (1987) have proposed rules for pause durations in English in which the length of pause increases with the size of the syntactic or informational units which the pauses demarcate. Zellner (1994) also presents a schema for silent pause estimation in connected speech for application in speech synthesis systems.

LOUDNESS is the perceptual correlate of intensity, which is usually expressed as magnitude of sound pressure variation in the speech signal (section 7.6 above). Intensity is primarily controlled by subglottal pressure (Ladefoged 1967, Lehiste 1970, Ohala 1970) but is also influenced by the natural sonority of the segments or sequences of segments in the relevant syllables. For example, the vowel of the English CVC syllable *shack* is more sonorous relative to its neighbouring consonants than the vowel in, say, *wool*. Stressed syllables often have greater overall acoustic intensity than more weakly stressed ones, particularly in English, where unstressed vowels are generally reduced to schwa or are realized as weak lax vowels, although loudness seems to be the least salient and least consistent of the three parameters of pitch, duration and loudness – at least for linguistic purposes such as signalling prosodic prominence. (See Kochanski et al. 2005 for an alternative view, and section 9.3 below.)

The segmental and suprasegmental dimensions of the speech signal do not function independently of each other. In particular, there are important interactions between the segmental structure and its accompanying pitch pattern. Several studies have been devoted to the effects of voiceless and voiced consonants on the pitch of adjacent vowels (see Hombert 1978 and Silverman 1984, 1990, for a review of evidence). It seems, for instance, that voiceless pulmonic egressive stops often, though not universally, result in a higher pitch on the following vowel. A major reason for the interest in such phenomena is that they explain the origins of tonal distinctions in some languages: a distinction between, say, syllable-initial voiced and voiceless stops is lost (by historical change) but a tonal distinction on the following vowel, originally conditioned by the preceding consonants, is preserved. Thus a secondary cue supplants the original primary one in the process of sound change. Hombert (1978, pp. 78–9) points to a number of south-east Asian languages, including Chinese and Vietnamese, in which such changes are reported. Figure 9.2.1 shows examples from Hombert (1978) of the conditioning of fundamental frequency by a preceding stop in English and French. The data have been normalized for comparison.

The reasons for this conditioning of pitch are not fully understood. One theory is that the larynx is often lower in voiced stops, to enlarge pharyngeal volume and maintain sufficient transglottal pressure to continue phonation during the occlusion; this lowering results in lower pitch. Conversely, the larynx remains higher for voiceless stops. Nevertheless, while there is a tendency for the pitch to be lowered during the occlusion phase of voiced stops, the evidence suggests that it is only voiceless stops that have a significant effect on the pitch in the initial part of the following vowel. An alternative theory proposes that vocal fold tension during a (voiceless) stop consonant may influence pitch at the onset of phonation in the vowel (e.g. Löfqvist et al. 1989). A major problem for this explanation is that voiceless stops do not seem to have the same influence on the pitch of a *preceding* vowel. Nor do studies of muscular activity indicate that muscular tensions in the larynx are significantly correlated with stop voicing. Overall there is no really satisfactory explanation for the pitch perturbation effects of prevocalic stops, especially in the light of the fact that postvocalic stops appear to have weaker and less consistent effects.

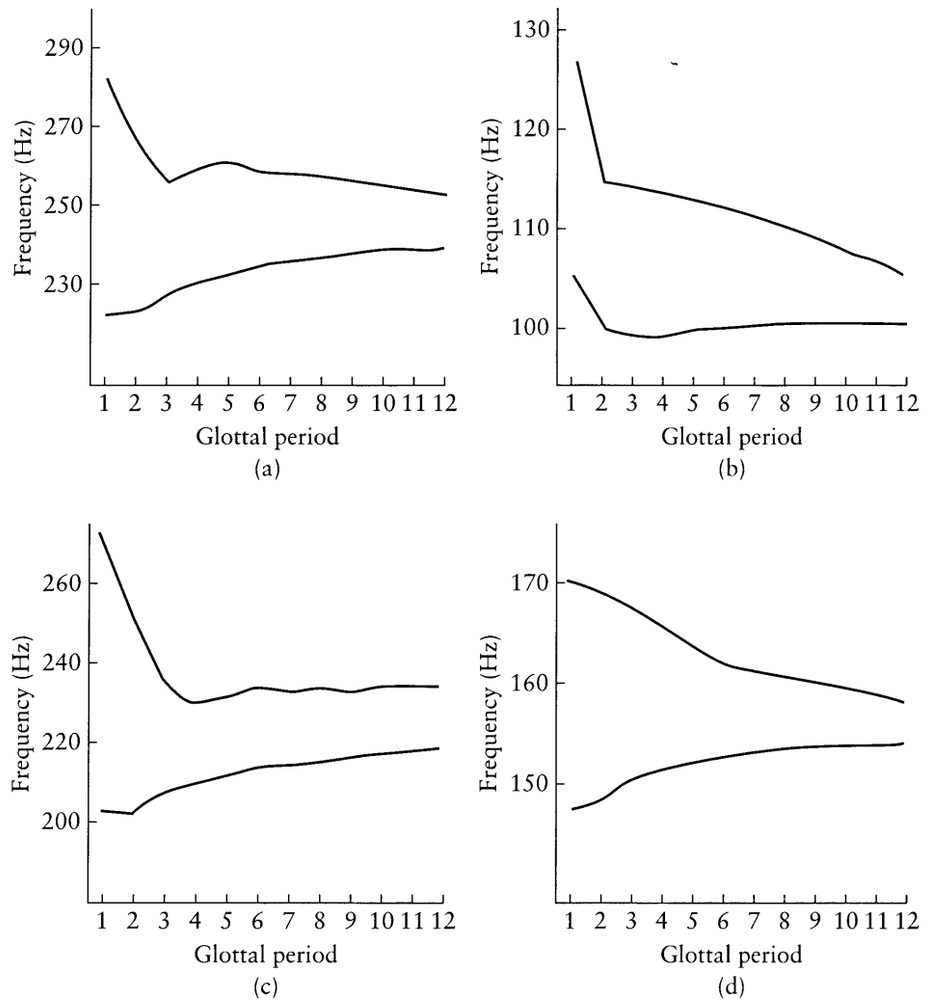

Figure 9.2.1 Effects of stop consonants on F_0. In each case the upper curve shows average F_0 values after *p t k*, the lower after *b d g*: (a) American English female speaker; (b) American English male speaker; (c) French female speaker; (d) French male speaker
Adapted from: Hombert 1978, p. 88.

Prenasalized stops and breathy voiced stops lower the pitch of following vowels more than plain voiced stops do. In the case of breathy stops, this is thought to be due to the lower intrinsic laryngeal muscle tensions used in breathy phonation. Implosives lower pitch less than plain voiced stops, possibly partly because glottal airflow is rapid as the larynx is lowered during implosion. It may also be that the muscular tensions required to close the glottis during the implosion counteract other factors tending to lower pitch in the following vowel. Generalizations are again dangerous, and Pinkerton (1986) shows that the conventional wisdom about the articulatory nature of various kinds of 'glottalized'

stops (including implosives) is not always supported by careful instrumental investigation.

Postvocalic stops have little effect on tone, but a glottal stop raises the pitch of the preceding vowel, and a rising tone may in time replace the syllable-final glottal stop. By contrast, postvocalic /h/ causes pitch lowering – presumably because of anticipatory relaxation of the laryngeal muscles – and may give rise to a falling tone. Ohala (1978) and Hombert (1978) provide details of consonantal pitch perturbation; see also section 9.4 below.

Vowels themselves tend to have an 'intrinsic' pitch which correlates with vowel height: high vowels have high pitch and low vowels have low pitch. According to Lehiste (1970) and Ohala (1978) the difference may be as great as 20–5 Hz, but Ladd and Silverman (1984) suggest that intrinsic pitch effects are not as strong in running speech as those observable in test words in citation sentences (see Silverman 1990 for a useful summary). There are several hypotheses about the causes of this effect, two of which will be noted here. The first is that the narrow constriction of high vowels causes an 'acoustic loading' of the vocal folds, which means that F_0 tends to be pulled towards the F_1 of the vowel. In high vowels, F_1 is quite low and in many cases within the speaker's pitch range. The second hypothesis is that there is 'tongue-pull', in other words that the mechanical coupling between the tongue root musculature and the larynx influences the height of the larynx and its phonatory adjustments. On this hypothesis, tongue raising will cause larynx raising. According to Lindblom and Sundberg's evidence (1971), if the mandible is fixed, there is more extrinsic tongue muscle contraction in high vowels, and the pitch difference between low and high vowels is thus enhanced. Silverman (1984) reviews the evidence for and against this hypothesis and others, and concludes that the evidence is not adequate to support any single explanation based on acoustic or physiological factors. He argues that although these factors may contribute to intrinsic vowel pitch, comparable effects may be, in part at least, phonologically motivated aspects of the speech production process and may be demanded by the perceptual expectation of the listener.

PITCH PATTERNS are essentially either steady, rising or falling, and it is changing pitch that has the greater perceptual salience. Evidence reviewed by Ohala (1978) suggests that falling pitch is more common in language than rising pitch, and that falling pitch uses a wider range of F_0 movement. It also seems that speakers can produce falling pitch more readily than rising pitch, and can achieve downward pitch movements more rapidly than upward movements. On the basis of this evidence, Ohala very tentatively hypothesizes that falling pitch is more salient perceptually and is more likely to be accomplished within a single syllable (1978, p. 31). But this is debatable, and significant pitch movement is not necessarily constrained within specific syllables of polysyllabic words, even when its main function is to mark major prominence on a single syllable.

DECLINATION is the term for what appears to be an almost universal tendency in language, namely a moderate progressive fall in pitch from the beginning to the end of any sequence of speech of appreciable length (Vaissière 1983). The term DOWNDRIFT is sometimes used with the same meaning, for example by Hyman (1975, pp. 225ff.), who distinguishes between this 'automatic' process

of lowering and the tonal phenomenon of DOWNSTEP. But there is potential confusion between 'downdrift' and 'downstep' in some authors, and we will reserve the term 'declination' for the phonetic pattern of F_0 behaviour.

Declination can generally be observed over identifiable units of the intonation system, often corresponding to clauses or clause complexes. There are of course constraints and exceptions, for example where the speaker selects a rising pitch pattern to signal that the utterance is a query. Declination occurs in both tonal and nontonal languages, and although listeners are not usually conscious of the effect, Breckenridge (1977), Pierrehumbert (1979) and Lieberman and Pierrehumbert (1984) have shown, for English at least, that listeners do compensate for its presence in judging pitch height.

There has been considerable debate about the status and causes of declination. Some researchers have argued that it is essentially an involuntary or automatic process, probably due to interaction between the larynx and the respiratory system. (This physiological explanation does not of course deny that declination can be deliberately suppressed or overridden for functional purposes.) Others have suggested that it is essentially the observable consequence of a phonological lowering of pitch on successive accented syllables. Ohala (1978) and Vaissière (1983) review the explanations that have been put forward, focusing on those related to speech production mechanisms. Gussenhoven (2004) discusses downtrends in general, and how they may operate differently from language to language. Ladd (1984, 1996) discusses declination in some detail: he makes the case that declination need not be a distinct component of pitch patterning, and that declination effects might be included in phonological behaviour rather than in quasi-intrinsic phonetic behaviour. In other words, the phonological rules of the language would include the generation of pitch declination (if required), and it would be wrong to assume that declination was an underlying pattern on which phonological pitch was superimposed. Figure 9.2.2 illustrates declination.

It has also been suggested that declination effects are observed mainly in formal reading aloud, oriented to prose sentences, and that these effects are much less noticeable in the patterns of informal speech. Certainly, declination effects can be suspended (Hyman 1975, pp. 227–8; Cruttenden 1997, pp. 162–4), and an essentially phonological explanation has strong appeal; but the debate is not resolved (see Pierrehumbert 2000 for a summary of the issues).

9.3 The systemic organization of prosody

Understandably, many terms used in describing prosody take on a particular meaning within particular languages: just as terms such as 'noun' and 'verb' or 'consonant' and 'vowel' cannot be expected to have identical reference across different languages, so also the terminology of stress and pitch needs to be carefully interpreted in the context of its use. In this section we review some of the more common terms and the uses to which they are put, as background

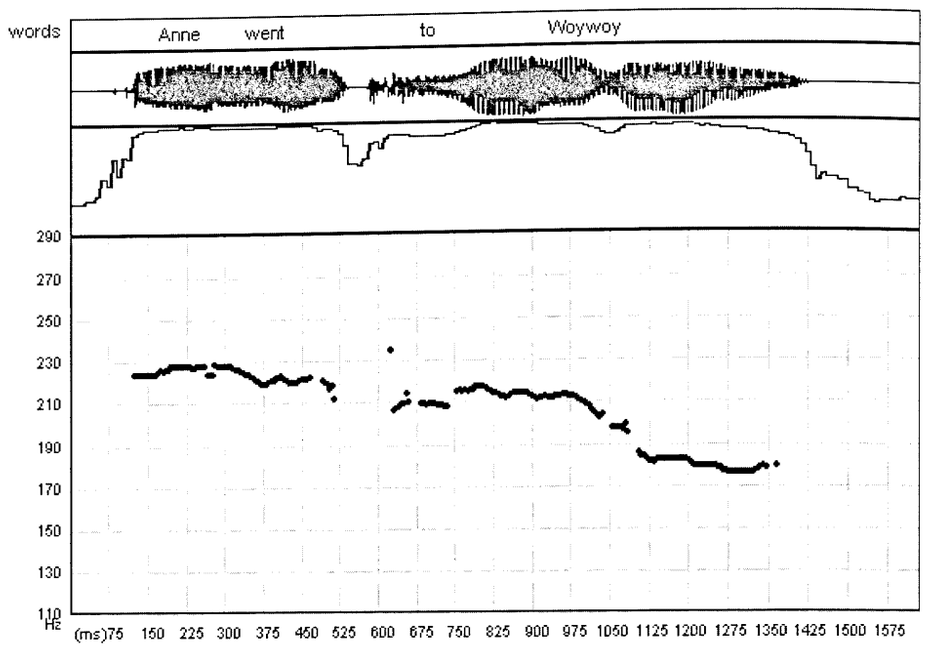

Figure 9.2.2 Declination of F_0 in overall sentence contour. (a) Time domain waveform; (b) intensity contour; (c) F_0 contour for the utterance *Anne went to Woywoy*

to subsequent sections dealing with some of the types of prosodic system that have been classically recognized; namely word prosody typology and rhythm or timing typology.

The traditional typological approach to word prosody separates languages on the basis of whether they have lexical tone, lexical stress or lexical pitch (Trubetzkoy 1939, Hyman 1978a). Languages such as Mandarin Chinese or Thai, in which differences of pitch serve to distinguish word meanings, are often called TONE LANGUAGES (section 9.4 below). Pitch is the chief phonetic correlate of tone, and in languages like Mandarin Chinese, distinctive pitch patterns are associated with syllables or lexical items. In other words, pitch contrasts serve to contrast word meanings even when the rest of the phonetic content of the word remains unchanged. Thus in Mandarin Chinese, what might seem to English speakers to be a single lexical item /ma/ is in fact four different words, depending on the associated tone (McCawley 1978, p. 120):

ma^1 (with high level pitch)	'mother'
ma^2 (with high rising pitch)	'hemp'
ma^3 (with low, or falling then rising, pitch)	'horse'
ma^4 (with falling pitch)	'scold'.

We can also say, for example, that the English word 'no' may be uttered either with falling tone (in which case it is likely to count as a definite refusal

or denial) or with rising tone (as a query, checking whether a denial or refusal is indeed intended). More precisely, we can identify a number of distinct pitches or pitch patterns in a language such as English, which some consider to be 'the tones' of English (e.g. Halliday 1967, Crystal 1969). Here the term takes on a systemic value, since we recognize only a finite number of discrete tones – such as high, low, falling, rising – which are functional in the language. In this sense, tone is not synonymous with pitch, since a tone in a linguistic system will be realized in such a way that it contrasts with other tones in the system while varying according to context. In the case of English, tones are part of what is usually called intonation (section 9.8). This is why languages like English are also called intonation languages in the older prosody typologies. This can be confusing because in fact tone languages have intonation too, as do all languages. We will return to this question below.

The term STRESS similarly has wider and narrower senses in linguistic description. Some writers (e.g. Trager and Smith 1951) have particularly related stress to loudness; one might then be able to distinguish the stress of a syllable (its perceived relative loudness, reflecting force of articulation) from its tone (the perceived pitch, whether relative height or a movement such as falling or rising, reflecting fundamental frequency). More commonly, stress is a conventional label for the overall prominence of certain syllables relative to others within a prosodic structure in a linguistic system. In this sense, stress does not correlate simply with loudness, but usually represents the effect of factors such as pitch, loudness and duration often in differing combinations depending on whether it is word or phrasal stress. A more restrictive definition of stress in languages like English suggests that duration and vowel quality largely contribute to a syllable's overall prominence at the word level (see section 9.6 below). It is in this sense that we say that the English words *over*, *supper*, *China* and *broken* are all stressed on the first syllable, while words such as *ahead*, *before*, *suppose* and *career* are stressed on the second. Where words have their own stress pattern or potential in this way, the stress is often called WORD-STRESS or LEXICAL STRESS.

Patterns of stress are highly important in a language such as English: this is not so much because the patterns are significant in distinguishing one word from another, although there are instances where this is true in English (e.g. *content* meaning 'pleased, satisfied' versus *content* 'that which is contained'); rather it is because the rhythm of spoken English is to a very large extent determined by strong beats falling on the stressed syllables of words. Thus a typical spoken utterance of English will consist of a number of rhythmic units, each of which is dominated by the beat of the stressed syllable. In verse, the wording is characteristically and deliberately organized to yield a regular rhythm, and the units of this rhythm are commonly called 'feet'; but the term FOOT is no less applicable to ordinary spoken English, even though the feet may not be consciously constructed. In a normal reading of, for example,

Wanda's joining the parade,

the rhythm is determined by the stress patterns of the words 'Wanda' and 'joining' (stressed on the first syllable) and 'parade' (stressed on the second syllable).

The words 'the' and 'is' are normally unstressed, so much so that 'is' can be pronounced without any vowel at all and written as a single consonant (s) tacked on to the preceding word. The resulting rhythm of the utterance can be informally conveyed as

WAN-da's JOIN-ing-the-pa RADE.

This kind of rhythm puts a characteristic stamp on the nature of spoken English. Normal tempo is such that unstressed syllables are greatly reduced ('swallowed' as some critics would have it) and simply form a tail of varying length in each foot. In the above example the three unstressed syllables following JOIN are likely to be the most rapidly articulated, while the final stressed syllable RADE (which happens to have no unstressed syllables following it) may be given extra length. This kind of foot is often referred to as the 'Abercrombian' foot (e.g. Abercrombie 1967), and consists of a stressed syllable and any following unstressed syllables, ignoring word boundaries. This is different from the prevailing phonological definitions of a foot as primarily disyllabic, and constrained by word boundaries (e.g. Hayes 1995; see section 9.7 below).

Traditionally it was also thought that each foot, whether a single syllable or several, will tend to take – very roughly – the same amount of time. This ISOCHRONY (equal timing) based on stress is often commented on – although the experimental evidence for it is not at all satisfactory (see Dauer 1983; Cruttenden 1997, pp. 20–2; Ramus et al. 1999; and section 8.9 above). It is related to the frequent description of English as a STRESS-TIMED language (Pike 1945, Abercrombie 1967, Halliday 1985a). By contrast many of the world's languages have been categorized as SYLLABLE-TIMED in traditional treatments of language rhythm (e.g. Pike 1945, Abercrombie 1967), although there is little empirical evidence to suggest that languages are either perfectly 'syllable-timed' or 'stress-timed' (e.g. Roach 1982, Dauer 1983). In other words, there is no real proof that stress feet are perfectly 'isochronous' (i.e. of equal duration) in stress-timed languages. Nor is there any experimental evidence to show that syllables are of equal or near-equal duration in syllable-timed languages. A third category, MORA-TIMING, is used traditionally to describe the characteristic rhythm of Japanese. A mora is a subsyllabic prosodic constituent or 'timing unit' that generally consists of a vowel, or vowel plus following consonant. Once again, there is little or no experimental evidence that morae are of equal or near-equal duration in spoken Japanese. However, the terms are still used widely and there is some consensus that the categories are still valid for informal descriptions of rhythm classes. Dauer (1983) suggests that languages may sound more syllable-timed if they have simple syllable structure and if they do not tend to reduce or centralize vowels strongly in unstressed syllables. Ramus et al. (1999) also claim that the more complex syllable types a language has, the more likely it is to sound stress-timed. This could also be due to the fact that in many 'stress-timed' languages (e.g. English), stressed syllables are usually complex or heavy syllables.

Readers may like to test their sense of spoken rhythm by articulating (or getting another to articulate) an English utterance while tapping out the rhythm of the major beats. The chances are that each tap will coincide with a stressed syllable in the utterance and that the taps will appear to occur at reasonably regular intervals. If the beats occur at more or less regular intervals, regardless of the number of syllables in each foot, the result is likely to sound like reasonably normal English. Deliberate variations in this pattern – for example, making each syllable a full beat, or adopting a simple alternating rhythm in which odd-numbered syllables are stressed and even-numbered syllables unstressed – should demonstrate that other rhythmic patterns are quite feasible but foreign to normal English. At the very least, the difficulty of this exercise should underline the way in which timing and rhythm are essential to the nature of spoken language.

The term ACCENT is sometimes used loosely to mean stress, referring either to prominence in a general way or more specifically to the emphasis placed on certain syllables. In traditional word prosody typologies, the term PITCH ACCENT, like tone, has a particular use in describing certain languages which are, in a sense, limited tone languages because of their use of lexically contrastive pitch. If, for example, a language has restricted tonal options such that two-syllable words are either high–low (high tone on the first syllable, low on the second) or low–high, then we might well simply say that words carry a (high-pitched) accent which falls either on the first or on the second syllable. In general functional terms, this is tantamount to saying that words are stressed on either the first or second syllable. Thus although most languages described as 'pitch-accent languages' are rather more complicated than this simple example, it is debatable how far they constitute a definite type, distinct from both 'tone languages' and 'stress languages'. We will discuss this further in section 9.5 below.

Where accent refers to relative prominence within longer utterances, this generally means intonational prominence. English is noteworthy for the way in which a stressed syllable (already prominent within the normal rhythmic pattern determined by word-stresses) can be further accentuated relative to other stressed syllables by virtue of its position in an intonational contour. For example, in the English utterance 'Hector started running' each word has its own (lexical) stress pattern, in this case on the first syllable in all three instances. Each of these stressed syllables can serve as the location of an intonational prominence or a pitch accent. This kind of accent is functionally different from lexical pitch accent as described above, in that it is not lexically *contrastive*, but serves as a pragmatic device to highlight a particular word in an utterance.

In the example below, normally, the last of these three words will be the most prominent, and will bear the NUCLEAR ACCENT. But it is also possible to put the nuclear accent on the first or second word, usually by ensuring that the major pitch target or movement of the intonational contour occurs on the stressed syllable in that word. Suppose – with considerable oversimplification – that a falling pitch is placed somewhere in this utterance and that the rest of the utterance is relatively level. If the fall (marked below) occurs over the first syllable of 'Hector' and the pitch of the rest of the utterance is relatively unchanging and lower than the starting point, English speakers will perceive emphasis on the first syllable:

\ – – – – –
HECTor started running.

This pattern invites the hearer to attend specially to the first word. Instead of 'nuclear accent', terms such as SENTENCE STRESS and CONTRASTIVE STRESS are sometimes used, the latter because the functional meaning is often one of contrasting the accented word with alternatives, for example where the speaker is contradicting the addressee ('you say that Rupert started running but I assert that HECTOR started running') or going against the likely assumption ('you would not have expected it but it was HECTOR who started running'). It is of course also possible to put the 'accent' on the second word, in which case we have – again in a simplified version – something like:

– – \ – – –
Hector STARTed running.

Here the function may be to correct the impression that Hector did not run at all or to indicate that he only started to run but quickly decided to walk instead.

Given that this phenomenon of 'sentence stress' or 'accent' in English is (often) a matter of the location of an intonational event rather than some heightening or intensification of the degree of lexical stress, it is preferable to place it within the wider context of the English INTONATION system. We deal with 'word-stress' and 'intonation', with special reference to English, in separate sections below (9.6–9.9).

In general, the terms reviewed here should be approached with caution. The terms 'stress' and 'accent' in particular are notoriously ambiguous, and it would be misleading to suggest that there are standard definitions. Certainly in the description of English, the phenomena to which the terms usually refer are best understood within an analysis of intonation, but it remains important, given the enormous attention that has been devoted to the prosody of English, not to assume that what is true of English is necessarily true of other languages. For an overview of terms see Cutler and Ladd (1983, pp. 140–6), Ladd (1996) and Cruttenden (1997). See also Beckman (1986), Ladd (2001), Gussenhoven (2004) and Jun (2005a: ch. 16), for a good discussion of stress accent and other accent typologies.

9.4 Tone languages

Many of the world's languages are traditionally recognized as 'tone languages'. The precise definition of a tone language is controversial but it is common among linguists to stress lexical relevance: in a tone language, tone is 'a feature of the lexicon, being described in terms of prescribed pitches for syllables

or sequences of pitches for morphemes or words' (Cruttenden 1997, pp. 8–9); or, more informally, pitch 'distinguishes the meanings of words' (Pike 1948, p. 3). This is in contrast to a language such as English, where pitch is certainly functional and where one can equally speak of distinctive tunes or melodies, but where these cannot be directly associated with lexical meaning. Speakers of tone languages can be expected to regard tone as a significant part of a syllable (or morpheme or word). Most of the world's languages are in fact tonal in this sense, including major East Asian languages such as Chinese, Vietnamese, Burmese and Thai, as well as a substantial proportion of the languages of Africa, the Americas and Papua New Guinea. Pike (1948) remains a classic introduction to the nature of tone languages and strategies of analysis and description. Otherwise most information is available in papers dealing with particular languages or with general problems of theory and description; a particularly useful collection was published as Fromkin (1978).

Pike is responsible for a distinction between REGISTER (or LEVEL-PITCH) tone systems and CONTOUR (or GLIDING-PITCH) systems. In a register system there are distinctive pitch levels, often two or three and rarely more than four, although Gussenhoven (2004) cites studies by Connell (2000) and Wedekind (1983) who posit a four-way contrast for Mambila, and a five-way contrast for Benčnon respectively. These levels will of course be relative to each other rather than absolute values, so that a high tone, for example, will be perceived as high relative to any adjacent mid- or low-tone syllable. In fact, it may not be possible to distinguish a two-syllable word with two high tones, uttered in isolation, from a two-syllable word with two mid or two low tones. On the other hand, in a contour system it is the pitch movement or glide that is characteristic: the contrast will be among patterns such as falling, rising and 'dipping' (fall–rise) rather than among relative heights or levels.

For Pike, it was important to differentiate these two types of tone language because of the method of analysis. While it may be possible to identify the distinctive tones of a contour tone language merely by listening to them, a register system will generally require points of reference against which the relative levels can be judged. Thus single syllables bearing high-, mid- or low-level tone will not be clearly identifiable unless adjacent to a 'marker' level. Hence Pike emphasizes the importance of tonal frames that provide a fixed context. Suppose, for example, that we have already established in a tone language that a certain prefix, meaning 'my', always carries low tone. If we then use this prefix as a frame, getting a native speaker of the language to utter various phrases such as 'my house', 'my garden', 'my vegetables' and so on, we can judge the tone of each noun relative to the low-tone prefix. If the initial syllable of the noun is at (more or less) the same pitch level as the prefix, we can identify it as low tone; if it is noticeably higher, it must be mid or high. The use of another frame, say a prefix bearing mid tone, will enable us to sort out the mids from the highs, and to check the lows, which should now be identifiably lower than the preceding mid tone. The procedure is clearly laborious, since one must identify suitable frames to begin with and then run extensive lists of items through them, but it offers a principled way of basing a tonal description on reasonably solid evidence.

The two types of tone language are nevertheless not quite as distinct as this account may suggest. In the first place, register systems rarely if ever consist of a few perfectly level and consistent tones. The effects of declination (section 9.2 above) may be such that what is in theory a sequence of identical tones may actually fall in pitch, and there may be quite specific assimilatory processes whereby a mid tone is realized as a rising tone between a low and a high, or a high tone is realized as a high fall if before a low, and so on. In fact the literature on tone languages suggests that interactions among tones are typical rather than unusual, and Pike himself devotes major attention to what he calls PERTURBATIONS of tone or tone SANDHI. He describes such phenomena in detail for two languages of southern Mexico, namely Mixteco and Mazateco (1948, chs 7 and 8).

Moreover, it often seems to be the case in register tone languages that tonal options on individual syllables are constrained by word patterns. For example, Leben (1978, pp. 186ff.) suggests that there are five basic word patterns in Mende, a language of Sierra Leone. The five are (1) high; (2) low; (3) high–low; (4) low–high; (5) low–high–low. But these five patterns may be distributed over words of varying length, so that a monosyllabic word carrying pattern (3) actually has a falling pitch, while a three-syllable word with the same pattern will have high–low–low. Examples of pattern (4) on words of different length are (Leben 1978, p. 186):

mbu ('rice')	(monosyllable with rising pitch)
fande ('cotton')	(first syllable low, second high)
ndavula ('sling')	(first syllable low, others high).

Thus pitch glides or contours are by no means excluded from register tone languages: what is significant is that these glides can be analysed as realizations of (sequences of) level tones.

On the other hand, contour systems frequently if not always include level or near-level tones. One of the four tones of Burmese, for example, is described as low level. Indeed, although the pitch is probably always the dominant cue, other factors, such as duration and abruptness, are relevant. The four tones, as described by Tun (1982, p. 80) are (1) low level; (2) high rising–falling; (3) high falling; (4) high falling, with abrupt ending. No fewer than three of the five tones of Thai are traditionally labelled high, mid and low. Gandour (1978, p. 42) lists the tones as (1) mid; (2) low; (3) falling; (4) high; (5) rising. As with register tone languages, it is the system within which these tones function that is significant, rather than a simple categorization of stable or gliding tones. Gandour (1978, pp. 43ff.) refers to experimental evidence suggesting that Thai listeners readily distinguish all five tones in isolation, without any frame of reference of the kind that seems necessary in the analysis of a register system.

In fact, the crucial difference between the two kinds of tone system may be that in contour systems tone is a property of syllables and in register systems tone is a property of larger units such as words. Hombert (1986, pp. 180ff.) reports a word-game experiment in which speakers of tone languages were asked to transpose parts of words (either vowels or syllables). Thus if the game

were applied to English, participants would be asked either to reverse the vowels of, say, *fifteen* (yielding presumably, *feef-tin*) or to swap the syllables (yielding *teen-fif*). Speakers of three west African languages (Bakwiri, Dschang and Kru) and four East Asian languages (Mandarin Chinese, Cantonese, Taiwanese and Thai) were asked to participate in the experiment. In the traditional classification, the African languages would be regarded as register tone languages, and the Asian as contour systems. Although Hombert points out that the results are not quite straightforward, it does seem that speakers of the four Asian languages tended to carry the tone with a transposed syllable, whereas the African participants moved the segmental component but left the tone behind, so to speak. Cruttenden (1997, p. 9) also comments that many African languages have 'characteristic tone', in which the tone is sensitive to word structure and affixation, as opposed to the more narrowly 'lexical tone' of languages such as Chinese.

Apart from research of this kind exploring the nature and diversity of tonal systems, considerable attention has been paid in recent years to the way in which tone patterns can be explained by rules. This is not just a matter of formulating rules to explain assimilatory adjustments and perturbations of sequences of tones, but also a more fundamental question of how tone is mapped on to segmental structures. Leben (1978), for example, uses data such as the Mende words given earlier in this section to support the notion that tone is a separate prosodic component of phonological representation. Certainly where a language, like Mende, has patterns that distribute themselves over words of varying structure, there is an obvious case for treating tone as something independent of, but associated with, segmental structure (Leben 1978, pp. 177–80). Schuh (1978, esp. pp. 251–2) relates the elaboration of tone rules to the question of typology, again making a distinction between the African and Asian types. Explorations of this kind, prompted by analysis of tone, are in turn related to more general issues in phonology which have been taken up in 'autosegmental' and 'metrical' phonology (sections 11.12 and 11.13 below).

A further perspective on the description of tone comes from the investigation of its historical development, including its origin or TONOGENESIS. It is clear for many languages that tone has arisen where pitch differences, originally conditioned by consonants, have become distinctive when the consonants have been changed or lost. In Vietnamese, for example, rising tones seem to be a consequence of lost glottal stops: a final glottal stop must originally have conditioned a rise in the pitch of the preceding vowel, and when final glottal stops were dropped, the rising pitch became a distinctive tone (Hombert 1978, pp. 92–3). It is not surprising that tones interact not only with each other but also with their segmental context (section 9.2 above); Hyman (1978b) provides a summary of ways in which tonal changes may be motivated, and Ohala (1978) and Hombert (1978) are useful reviews of evidence.

There is no standard way in which tones are marked, either in conventional orthographies or in linguists' representations. In traditional Chinese orthography, tones are implicit in the characters and there is no particular symbol or diacritic to indicate each tone; on the other hand, many of the world's tone languages, in Africa and the Americas, have relatively modern spelling systems

devised by missionaries or linguists, in which tone, if indicated at all, is usually marked by some kind of diacritic within an alphabetic writing system. Pike (1948, pp. 36–9) notes various ways of using accents in practical orthographies. Linguists themselves sometimes resort to pictorial representations of tone, based either on a plot of fundamental frequency or on an impressionistic trace of the perceived pitch. This is particularly helpful in displaying contour tones, which may differ in the duration and slope of a pitch movement and not just in direction of movement. Such displays are of course cumbersome as a regular notation, but a system devised by Chao (1930) for Chinese is an interesting compromise between pictorial accuracy and alphabetic convenience. In this system, an iconic shape representing the tone is attached to a vertical marker line at the right of each symbol. Examples are

mid level tone ˧
rising tone ˧˥
falling tone ˥˩

The system is quite often used (e.g. McCawley 1978, p. 120), and the International Phonetic Alphabet now includes a large number of tone symbols.

Several other notational strategies are common among linguists. Firstly, simple diacritics, notably accent marks, may be used. While the shape of the accent can usefully indicate pitch movement (e.g. acute for a rising tone), it is often convenient to be even more conventional and to use acute for high tone and grave for low. Various other semi-arbitrary conventions are often adopted, such as the use of a bar above or below a vowel to indicate mid- or low-level tone. A second strategy is simply to number the tones and mark each syllable with its number, e.g. [ma^1] or [ma^2]. Pike uses this notation for Mazateco (1948, ch. 8): there are four contrasting level tones, which he numbers 1–4 from highest to lowest. In the same work, Pike uses accents for the three tones of Mixteco (acute for high tone, bar or macron for mid, and grave for low; 1948, ch. 7). Thirdly, contour tone languages may be transcribed using a system developed by Chao (1930), where at least two numbers from 1 to 5 are used to describe a tone's trajectory. In this system, 1 is the lowest pitch level and 5 is the highest pitch level. Fourthly, tones may be represented by letters, e.g. H for high, or L for low. This notation has become popular in recent work in which tone is assumed to constitute a separate layer or component mapped on to segmental structure, as in

H L
[b a m a] (high tone followed by low) or

HL
[b a :] (falling tone).

Mandarin (Putonghua) tones can therefore be transcribed in several ways (after McCawley 1978, p. 120; Peng et al. 2005, p. 235):

ma^1 (with high level pitch)	55	H	'mother'
ma^2 (with high rising pitch)	35	LH	'hemp'
ma^3 (with low, or falling then rising pitch)	21(4)	L	'horse'
ma^4 (with falling pitch)	51	HL	'scold'.

Two brief and readable accounts of tone in particular languages can be found in Fudge (1973a): an extract from Kratochvil (1968) describes the tones of Chinese with details of the nature of the four tones and their relationship to stress, and Smith (1968) deals with tone in the west African language Ewe. The reader may find both accounts informative about how tone functions in language and illustrative of methods of description and notation. Comrie (1987) also includes accounts of several tone languages, with concise notes about the tonal system, notably: Hausa, Yoruba and other west African languages (chs 35 and 49, esp. pp. 707, 711, 974, 977); Thai (ch. 38, esp. pp. 761–3); Vietnamese (ch. 39, esp. p. 783); Chinese (ch. 41, esp. pp. 814–16); Burmese (ch. 42, esp. p. 842). Yip (2002) is a good textbook introduction to the phonetics and phonology of tone systems across a wide range of languages.

9.5 Pitch-accent languages

Several of the world's languages are said to have PITCH ACCENT: these include Japanese, Norwegian, Swedish and Serbo-Croatian. As already noted (section 9.3) they are in a sense on the fringes of fully fledged tone systems. Pike refers to such languages as 'word-pitch' systems and describes them as 'utilizing pitch in the differentiation of the meaning of various lexical items, but with the placement of the pitch limited to certain types of syllables or to specific places in the word' (1948, p. 14).

In Japanese, the location of a pitch pattern or pitch accent on a particular syllable can differentiate the meaning of words. Words also have the option of not having a pitch accent; that is, Japanese contrasts accented and unaccented words. This is illustrated in the following examples (from McCawley 1978, p. 113). Falling pitch is marked by \, showing that the preceding syllable is accented, and that the following syllable (if any) drops to a lower level. (A word or phrase may also be unaccented, in which case there is no fall; and the pitch of any syllable preceding the accented one is predictable, since a word-initial syllable is low and any other syllables before the accent are high.)

ka\ki ga ('oyster')	H L L (first syllable accented)
kaki\ ga ('fence')	L H L (second syllable accented)
kaki ga ('persimmon')	L H H (unaccented).

Recent treatments of Japanese tonal structure reformulate McCawley's analysis (e.g. Pierrehumbert and Beckman 1988, Venditti 2005) and suggest there is only one composite lexical pitch accent shape in Japanese, H*+L (the HL sequences

in the first two words above), which is anchored to an accented mora. The rest of the tones (e.g. the LHH sequence of the third accented word shown above) are provided by the postlexical intonation system (see section 9.8 below). The key question becomes one of where the 'accent' is located, and not one of opposition of tone type such as high versus low, or rise versus fall. A second key question is whether a word is accented at all.

In Swedish, there are two tones or accents, and about 500 pairs of words are distinguished by this tonal contrast; thus there are only two options, but these may be considered to constitute a simple tonal system. Swedish also has lexical stress, and the location of contrastive pitch accents is constrained by the location of stress in Swedish words. The word *anden* 'the duck', for instance, has falling tone on the first syllable, whereas *anden* 'the spirit' has a double-peaked pattern with a fall on the second syllable as well as the first (Gandour 1978, pp. 53–4). Bruce (2005) remodels the two basic word accents (accent I and accent II) as H(igh) + L(ow) tone sequences or pitch turning points that can be distinguished by where they are aligned relative to a stressed syllable. In accent I cases the H tone is aligned to the pre-stress syllable, and the low occurs on the stress syllable, but in accent II the H is aligned to the stressed syllable and the low to the post-stress syllable (Bruce 1977). Swedish has more in common with other stress languages, like English, than it does with Japanese, for example. The precise realization of these accentual patterns varies both according to the context in which the words appear and according to the dialect of Swedish, but in many dialects, the H+L tonal gesture is timed earlier for accent I and later for accent II, all things being equal. The system is highly constrained – it does not apply, for example, to monosyllabic words, which always carry the fall as the normal accent. See Bruce (2005) for a full discussion of these issues.

According to traditional analyses, Serbo-Croatian (Browne and McCawley 1973; Gandour 1978, pp. 49–53) is usually described as having four tones or 'accents', although, like Swedish, Serbo-Croatian is also a stress language. The four pitch accents of Serbo-Croatian are described as (1) short rising; (2) short falling; (3) long rising; (4) long falling. But vowel length is distinctive in unaccented as well as accented syllables, so that the tonal contrast is essentially one of fall versus rise, intersecting with an additional opposition of length. A recent alternative analysis is presented by Godjevac (2005), who reduces the number of contrastive pitch accents to two: falling (HL) and rising (LH). There are again limitations on the exercise of the options – the falling accent is restricted to initial syllables (including monosyllabic words), while the rising accent is restricted to nonfinal syllables in Serbo-Croatian. A system of this kind is open to more than one analysis (as demonstrated by Browne and McCawley, and Godjevac), but one can again speak of a limited tonal system.

The brief analyses of Serbo-Croatian and Swedish highlight some important problems with traditional word prosody typology. McCawley's discussion of Japanese and other languages (1978) leads him to a rather different conclusion from that adopted by Pike (1945), for example. He points out that any language will exhibit a combination of characteristics, including

1 whether tones or accents are integral to lexical items and if so whether this is a matter of tone type (as in Chinese) or accentual pattern (as in Japanese);
2 what effect the rules of the language have, for example by limiting the options of the system;
3 what units are relevant in the operation of the system, for example whether tones or accents are carried by syllables or words.

This leads McCawley to reject simple classifications of the kind that typify languages as 'tone languages' or 'pitch-accent languages'.

Another related view is expressed by Beckman (in press), who combines aspects of word prosody typology with intonational typology. She observes that all languages have tones or tunes (whether provided by the lexical system, as in a tone language, or by the pragmatic system, as in 'intonation' languages like English). A second important factor is how these tones or tunes are aligned to the text (i.e. tune–text alignment). Some may be aligned with a stressed or rhymically prominent syllable, or to some other syllable or prosodic unit, such as the mora, which is not rhythmically strong (as in Japanese), or they may be aligned to the edges of larger prosodic constituents such as accentual phrases and intonational phrases, as in languages like Korean. Ladd (2001) and Jun (2005b) give a good overview of related issues, making the link between word prosody and intonational typology.

Prosody, like other global systems, comprises subsystems – such as choice of tone type and placement of tone (or accent) – which can certainly be compared and may show similarities, even among otherwise dissimilar languages. In that light, while simple categorization of linguistic types always runs the risk of superficiality, specific phenomena and functional mechanisms are worth study. Hyman (1975, ch. 6), Cruttenden (1997, ch. 1) and Gussenhoven (2004, chs 2 and 3) are useful in this regard, both as overviews and as pointers to more detailed literature on word prosody.

9.6 Stress in English

The phenomenon of lexical stress in English has received considerable attention and is probably best described as a word pattern or potential. Halliday (1970) speaks of 'word accent' as the potential salience of certain syllables within certain words; Gimson includes a detailed description of the 'accentual patterns' of English words within his more general treatment of English pronunciation (1980, esp. ch. 9); and Fudge introduces his book-length treatment of English word-stress (1984) by referring to the way in which one syllable in a given word is picked out or singled out.

Some authors (e.g. Trager and Smith 1951) particularly associate (lexical) stress with loudness. In their treatment of English stress, Chomsky and Halle (1968, pp. viii, ix, 15) say they are concerned with 'stress contours', not with

pitch, although they do not explicitly claim that stress is purely a matter of loudness (cf. Crystal 1969, pp. 113–20, 156–61). If the word 'sugar' is uttered on its own – say in reply to the question 'What's in this container?' – the first syllable of the word is likely to have higher pitch than the second as well as being (relatively) loud and long. Our perception is in fact likely to be more responsive to the pitch pattern than to the other factors. All of the factors are of course relative, and integrated within the intonation system. Hence, for example, if the speaker opts for rising pitch, to signal a query ('Sugar? Is that what you said?'), the second syllable will be higher than the first, but the *change* of pitch, in the context of a rising pattern, coupled with the relative loudness and duration of the first syllable, will normally be perceived as prominence on the first syllable. Indeed, in longer utterances it is often the point at which the pitch level changes substantially that signals prominence, rather than the level itself. Moreover the integrated nature of the system is such that loudness (or duration) may become a primary cue for stress and prominence where pitch has been pre-empted for some other function (Crystal 1969, p. 120). This view is further supported in work by Kochanski et al. (2005), who found that loudness is a more reliable cue to prominence than F_0 in their investigation of a large corpus of multiple varieties of British English.

Despite the persistence of the terms 'word-stress', 'lexical stress' and 'prominence', the patterning of spoken English is not based on words – or at least not on words in a grammatical or orthographic sense. Phrases such as 'the table' or 'a party' or 'leave it' will normally have the pattern of single words, with only one prominent syllable. In fact, there is normally no difference in spoken English between single words such as 'array' or 'arise' and two-word combinations such as 'a ray' or 'a rise'. Some writers therefore redefine the word for phonological purposes, as a PHONOLOGICAL WORD (e.g. Nespor and Vogel 1986, 11.17), or use some other term such as STRESS GROUP (Fudge 1984, p. 1) or FOOT (section 9.3 above; Halliday 1970, p. 1, and 1985a, pp. 271–3). This is not to be confused with the use of the term 'stress foot' in metrical phonology (e.g. Liberman and Prince 1977).

The corollary of this informal use of the term 'stress group' or the more formal concept of the phonological word is that certain English words (grammatical or orthographic words) are characteristically unaccented in connected speech, and can be grouped together with a neighbouring word that bears a main stress or accentual prominence. We must distinguish those monosyllabic words that normally are stressed or accented in connected speech from those that are not. The latter are a small minority but are words of very high frequency, including articles and prepositions such as 'the', 'a', 'at' and 'to', pronounced virtually as prefixes to the following word, and pronouns such as 'he', 'him' and 'them', pronounced as suffixes of the preceding word, as well as diverse other items such as 'and', 'than' and 'that'. A full list is given by Gimson (1980, pp. 261–3). English intonation does allow the option of stressing or accenting these words, but the stress or accent is then meaningful, in contrast with the normal or unmarked pronunciation. Compare a normal reading of

Joe was angry (two phonological words)

with a reading in which 'was' is stressed

Joe WAS angry (three phonological words).

The second reading signals that the speaker is contradicting a previous statement or implying that Joe was angry but no longer is. Many of these normally unaccented words have a reduced segmental shape as well: for example 'he' has no initial [h] in 'did he?' (= 'diddy') or 'was he?' (= 'wozzy'). Some speakers of English have unnecessary misgivings about such reduced pronunciations, and, especially in formal situations, produce fully accented versions where the normal unaccented version would be more communicative. (Note for instance the effect of an overcareful reading of 'numbers one to four': a speaker who stresses the word 'to' runs the risk of causing confusion with 'two'.)

Gimson (1980) comments that the stress pattern of English words is free, in the sense that there is no simple rule that lexical stress always falls on a particular syllable of the word (say the last or the penultimate). But there is a large measure of predictability about English lexical stress (section 9.7 below) and Gimson himself comments that lexical stress is usually fixed, in the sense that the stress falls (almost always) on the same syllable of any given word (1980, p. 221). Gimson illustrates the variety of patterns in some detail (pp. 226–30), including those instances where the position of the stress is grammatically distinctive, such as in the nouns *conduct* and *rebel* as opposed to the corresponding verbs *to conduct* and *to rebel* (p. 233). Fudge likewise notes that, subject to certain exceptions, 'the place of word-stress within the word remains constant' (1984, p. 3); he also gives a comprehensive list of those words that do have distinctive stress (mostly noun–verb pairs, pp. 189ff.).

The exceptions to which these authors refer are cases where the stress pattern of a word may vary according to context, and where other aspects of English prosody may be said to override the 'normal' word-stress pattern. An example is the word 'afternoon', which usually has major lexical stress on the last syllable (e.g. in 'in the afternoon') but has the stress on the first syllable in phrases such as 'afternoon tea' or 'afternoon sun'. This is often what is referred to as the 'rhythm rule' (e.g. Hayes 1989). Other words that vary in similar fashion are 'fifteen' (compare 'at three-fifteen' and 'fifteen teachers') and, at least in a conservative variety of English, 'princess' (compare 'a princess' and 'Princess Margaret'). But the word 'princess' also demonstrates that lexical stress patterns may vary among individuals and groups, for many speakers of English consistently stress 'princess' on the first syllable, regardless of context. Indeed, there is considerable 'instability' of lexical patterning in English (Gimson 1980, pp. 230–2). A common tendency, for example, is for speakers to stress the second syllable of certain longer words that were traditionally stressed on the first syllable:

INtegral	or	inTEGral
COMMunal	or	coMMUNal
FORmidable	or	forMIDable
CONtroversy	or	conTROVersy.

Pronunciations in the first column may generally be considered conservative, those on the right increasingly the norm. The change may reflect a preference for a pattern in which the major stress is surrounded by unstressed syllables, rather than an initial stress followed by two or three unstressed syllables. It is probably safe to say that most younger speakers of English would regard 'FORmidable' as an awkward pronunciation. Examples such as these are not necessarily unstable within the speech of an individual – although some speakers, knowing the alternatives, may be hesitant about their pronunciation – but are another reminder that English phonology is not a single system, uniform across all groups and regions.

The normal pattern of a word may be systematically overridden by the placement of the nuclear accent (what some writers call 'sentence stress', here dealt with as part of the intonation system in section 9.8 below). The characteristic pitches of English utterances usually fall on syllables that are potentially stressed by virtue of word-stress patterns. Thus when the placement of the nuclear accent is varied in the following sentence, it is the (lexically) stressed syllable of the relevant word or foot that is selected:

Joanne	wanted	Louise	to join	the paRADE
Joanne	wanted	Louise	to JOIN	the parade
Joanne	wanted	LouISE	to join	the parade
Joanne	WANTed	Louise	to join	the parade
JoANNE	wanted	Louise	to join	the parade.

But in certain cases, a syllable which does not normally receive lexical stress may be selected, for example when the syllable is specifically contrasted, as in

ThirTEEN girls and thirTY boys
compare: THIRteen girls and FOURteen boys

or

I said 'MYology' not 'BIology'
compare: I need a book about myOLogy.

So far we have spoken in terms of patterns in which one syllable is stressed, relative to the other unstressed syllables of the foot. But some writers recognize intermediate degrees of stress in English. The following words, for instance, all seem to have the major stress on the first syllable; but some speakers pronounce the words on the right with a second syllable that seems to bear some degree of stress. The transcriptions represent a typical Australian pronunciation:

collar /ˈkɒlə/	follow /ˈfɒloʊ/
lacquered /ˈlækəd/	placard /ˈplækad/
conquered /ˈkɒŋkəd/	concord /ˈkɒŋkɔd/.

Likewise many words of three or more syllables may be perceived as having a syllable which is intermediate between stressed and unstressed. Compare words with initial stress and two unstressed syllables (on the left below) and those with major stress on the initial syllable and minor stress on the final (on the right). The transcriptions again represent Australian pronunciation:

numerous /ˈnjʉmərəs/	universe /ˈjʉnəˌvɜs/
quantity /ˈkwɒntəti/	pedigree /ˈpɛdəˌgri/
delicate /ˈdɛləkət/	indicate /ˈɪndəˌkeɪt/

On the basis of pronunciations such as these, some linguists recognize degrees of word stress, in particular PRIMARY and SECONDARY stress. Thus the word 'universe' can be said to have primary stress on the first syllable, no stress on the second syllable, and secondary stress on the final syllable. Formally, this amounts to a three-level system (in which zero or unstressed is the lowest level).

Not all speakers of English will agree that these examples demonstrate an intermediate level of stress. Some American speakers, for instance, may pronounce 'conquered' and 'concord' identically (and with syllabic /r/ in the second syllable rather than a vowel); and many speakers, whether from North America or not, may judge 'quantity' and 'pedigree' to have exactly the same stress pattern. Certainly it is true that the distinction between a syllable with secondary stress and an unstressed syllable almost always hinges on the occurrence of schwa, the so-called indeterminate vowel [ə]. In English, this vowel can be considered to signal minimal or zero stress. A syllable containing any other vowel quality, but not given prominence by the normal devices of English stress-marking, will then count as having secondary stress. With the exception of 'pedigree' (which is in any case disputable), all of the examples of secondary stress given above are open to this interpretation. There is consensus that in English the content of syllables can determine whether a syllable is stressed or unstressed. A 'full vowel' or vowel with full vowel quality is stressed, whereas a vowel that is schwa or schwa-like will be deemed weak and unstressed. English is often described as 'quantity' sensitive in this way.

Many writers recognize even more than three levels of stress (e.g. Trager and Smith 1951, Chomsky and Halle 1968, Liberman and Prince 1977) although the instances that seem to require four or more degrees of stress are, as we shall see (section 9.7), complex structures such as compounds and phrases.

In this context it is not surprising that the notation of lexical stress in English is far from standardized. The use of numbers above the relevant syllables is common in North American publications and is an attractively easy way of indicating several levels of stress. Within the tradition of the IPA, and especially in British descriptions, the marks [ˈ] (primary stress) and [ˌ] (secondary stress) are widely used. For convenience, an accent above or after the stressed vowel (as used in many dictionaries) or capitalization of the stressed syllable are simple and handy devices, but they do not lend themselves to systematic display of different levels of stress. For accuracy, indication of pitch, loudness and duration can be combined in a stylized pictorial display, often known as

'tadpole' notation. Illustrating these possibilities we have (assuming secondary stress on the third syllable of 'pedigree')

1 3 2				
pedigree	'pedi¸gree	pédigree	PEDigree	pedigree

9.7 Stress assignment

The extent to which the placement of lexical stress in English can be explained by rule remains a controversial issue. In many languages the patterns of lexical stress seem to be governed by relatively simple principles. Thus it may be possible to predict the occurrence of stress from phonological structure, as in the statement that words are always stressed on the initial syllable in Finnish (Comrie 1987, p. 598) or on the final syllable in Turkish (Comrie 1987, p. 628) or on the penultimate in Polish (Comrie 1987, p. 354). Sometimes stress rules are not strictly a matter of phonological structure, but are sensitive to grammatical structure as well. For example, both Farsi (Persian) and Turkish are said to have stress on the last syllable of words, but in both languages certain suffixes do not count as part of the word for this purpose (Comrie 1987, pp. 529, 628). For such languages the stress rule may be better phrased as assigning major stress to the last syllable of a (grammatically defined) stem or root. In Polish also, certain words break the rule of penultimate stress, but these exceptions are predictable from their grammatical form. Indeed many languages of which it may be said that stress is regular or predictable are actually subject to several rules. In Italian, for example, stress commonly falls on the second syllable of three-syllable words:

aMIco ('friend') voLAre ('to fly');

but note also

CApito ('I turn up') capiTO ('s/he turned up').

These forms reflect other rules, such as one that puts stress on the final syllable of verbs meaning 's/he . . . ed'. This of course undermines the simple distinction sometimes drawn between languages in which the position of stress is predictable by rule and those in which its position is contrastive or meaningful. In Italian the stress on the second syllable of *amico* is predictable, but the final stress on *capito* is meaningful. Several principles of this kind can be observed in English: for example, the position of lexical stress serves to distinguish noun from verb in words such as *conduct*, *insert* and *reject*, where the rule is that stress is on the first syllable of the noun and on the second syllable of the verb; while stress can also be said to fall on the root, regardless of prefixes and suffixes, in examples such as

FRIEND	DECK	BOARD	RISE
beFRIEND	beDECK	aBOARD	aRISE
FRIENDly	DECKing	BOARDer	aRISen.

An adequate account of English lexical stress must in fact recognize three relevant factors in relation to each word: firstly its origin (e.g. whether it is of Greek or Latin origin); secondly its phonological structure (e.g. whether it contains certain kinds of vowel and consonant combinations); and thirdly its grammatical organization (e.g. whether it is a compound noun, or a root plus suffix, and so on).

The fact that English words of Latin origin tend to follow the Latin rules of stress has long been noted (Chomsky and Halle 1968, p. 59, n. 3). Since words of Greek origin also show some 'non-English' stress characteristics, Kingdon (1958) made a tripartite division of English vocabulary into words of Greek, Romanic and English origin, as part of an explanation of English word-stress. Interacting with this distinction are structural considerations, notably the effects of what Chomsky and Halle call 'strong' and 'weak' clusters (1968, p. 29). A weak cluster is a sequence consisting of a short vowel followed by at most one consonant; a strong cluster consists either of a short vowel followed by at least two consonants or of a long vowel or diphthong followed by any number of consonants. Now this structural difference is relevant in a stress rule (more or less reflecting the stress patterns of Latin) which applies to words such as the following, which end in a weak cluster and have stress on the penultimate syllable:

deVElop deLIver inHERit inHIbit EDit;

whereas those that end in a strong cluster have stress on the final syllable:

eLOPE comPLETE reVEAL aLLOW exIST.

But morphological factors are also relevant – in particular, certain suffixes have their own effect on the stress pattern. Thus the suffix *-ance* or *-ence*, although it ends in a strong cluster, does not attract the stress when added to the above words, i.e.

deLIVer	deLIVerance
inHERit	inHERitance
aLLOW	aLLOWance
exIST	exISTence.

On the other hand, the suffix *-ion* requires stress on the preceding syllable, which in some cases causes a shift of stress:

inHIBit	inhiBItion
EDit	eDItion
DEDicate	dediCAtion

Furthermore, if word-stress rules are intended to cover the patterning of compounds and phrases, they must account for the English tendency to stress the first element of a compound but the final element of a phrase. Note examples of contrast such as

a BLACKbird a black BIRD
a BLACKboard a black BOARD
a BLACKberry a black BERRy.

The significance of this distinction is actually quite subtle, and not always reflected in the spelling (as one word or two). Thus many speakers treat the following as compounds, with stress on the first word:

COFFee table
BIRTHday party
BIRD'S nest
CHURCH Street
the WHITE House (the presidential residence in the USA);

but not the following:

garden SHED
leather JACKet (but note the fish: LEATHerjacket)
Church ROAD
the white HOUSE (a house which happens to be white).

Notice that word stress is preserved within the larger context of a compound or phrase: when *berry* or *jacket* is stressed as the second element of a phrase, the stress falls on the first syllable of the word because that is where it normally falls in these words. Hence we have structure within structure, which can be displayed by bracketing, for example:

loganberry = [[logan] [berry]].

Within the innermost brackets (surrounding each word) there is a lexical stress assignment that determines which syllable of the word is most salient; at the higher level of the outer brackets (taking the two words together as a unit) there is a further stress assignment that determines which of the two lexical stresses will be heightened – the first if the unit is a compound, the second if it is a phrase.

A formalized version of this stress assignment procedure is set out in some detail by Chomsky and Halle (1968, pp. 15–27; see also section 5.6 above). They argue that it follows a CYCLIC principle, such that the same rules (say a compound rule heightening initial stresses and a phrase rule heightening final stresses) may be repeated at each level of constituency, working up from the innermost bracketing to the outer. Thus they take examples such as the following:

1 blackboard eraser (i.e. the thing for cleaning a blackboard)
= [[[black] [board]] [eraser]].
Innermost brackets enclose the single words; the next level up is the bracketing of [black board] as a compound; and the outermost brackets enclose all three words as a compound. Hence the total structure is of a compound noun, the first part of which is itself a compound of an adjective and a noun.
2 black board-eraser (i.e. a board-eraser which is black)
= [[black] [[board] [eraser]]].
Innermost brackets enclose the single words; the next level up is the bracketing of [board eraser] as a compound; and the outermost brackets enclose all three words as a phrase. Hence the total structure is of a phrase, consisting of an adjective preceding a compound noun.

Now Chomsky and Halle assume several degrees of stress in English. Conveniently, they number them, taking 1 as the maximum degree, and they adopt the convention that any rule that assigns stress actually lowers the stress on all other syllables. This is as if we start with the assumption that every syllable is (potentially) numbered 1: we then assign stress to, for example, the first word of the compound *blackboard* by lowering the second by one degree, yielding the stress pattern 1 2.

A simplified summary of what happens to example (1) under this scheme is as follows:

1 (i) Input [[[black] [board]] [eraser]].
(ii) Lexical stress assignment (within innermost brackets); for simplicity we ignore unstressed syllables of words:
1 1 1
[[black board] eraser].
(iii) At the next level, only the compound rule applies (within the now innermost bracketing [black board]), heightening the first word of the compound:
1 2 1
[black board eraser].
(iv) At the next and highest level, the compound rule again applies, now strengthening the first word of the entire structure (i.e. lowering all others by one degree):
1 3 2
black board eraser.

The usual reading and perception of this compound should indeed be with greatest stress on the first word and least on the second.

The second example will, in response to its different grammatical structure, acquire a different stress pattern from the same procedural routine.

2 (i) Input [[black] [[board] [eraser]]].
(ii) Lexical stress assignment (within innermost brackets);
1 1 1
[black [board eraser]].

(iii) At the next level, only the compound rule applies, heightening the first word within the now innermost bracketing [board eraser]:
1 1 2
[black board eraser].

(iv) At the next and highest level, the phrase rule now applies, strengthening the last major stress of the entire structure (note that this is 'board', as 'eraser' has been weakened to a lower level on the previous cycle):
2 1 3
black board eraser.

This should again accord with our usual reading and perception of the phrase.

This generative routine is often referred to in the older phonological literature as the STRESS CYCLE or PHONOLOGICAL CYCLE, and a useful simple account (under the latter name) can be found in Schane (1973, pp. 100–4). Schane's analysis of the phrase *Spanish American history teacher* (following Chomsky and Halle) presupposes five levels of stress and several alternative interpretations of the phrase. With the major stress on *history*, for example, we have

2 5 5 4 5 5 1 5 5 3 5
Spanish American history teacher

Schane's point is that this reading significantly reflects the structure, yielding the meaning 'a teacher of American history who is of Spanish nationality'. Alternative readings, namely 'a history teacher who is Spanish American' and 'a teacher of Spanish American history', will have different stress patterns reflecting the different interpretations.

It remains an open question how far this is truly a stress system, independent of intonation, and how far Chomsky and Halle can justify their assertion that they deal with 'stress contours', not pitch (1968, pp. viii, ix, 15). It is doubtful whether English speakers control the stress pattern of such phrases independently of the wider context of tone choice and placement within larger units of language. Certainly if 'word-stress patterns' are taken to be relevant within relatively small domains, it is unnecessary to recognize more than three levels of stress at most (including unstressed). Given the role of the vowel [ə], a simple two-way opposition of stressed and unstressed may be descriptively adequate.

Liberman and Prince (1977) recast many elements of the stress rules of *The sound pattern of English* (SPE) developed by Chomsky and Halle, but initiated a new tradition in generative phonology wherein stress is no longer analysed as a gradient phonological feature [*n* stress], and is best treated as a structural position in a metrical constituent, i.e. a foot. This approach in phonology is termed 'metrical phonology' (see section 11.13 below). In English, a word can consist of one strong foot that is rhymically stronger than adjacent feet, if the word is polysyllabic. The primary stressed syllable is the rhythmically most prominent syllable in the word, and occupies a structural position in the foot that is usually called the HEAD. Syllables that occupy the heads of surrounding feet in the word are secondary stressed syllables. A crucial structural element of English stress feet in this tradition is that they are binary (i.e. consist of two syllables) and are left-headed (i.e. the rhythmically strong syllable is followed by a rhythmically weak syllable; see section 11.13 below). Stress feet are also not formed across word boundaries, and are dominated by a larger constituent, the phonological word (section 11.17 below), unlike the Abercrombian foot, described in section 9.3 above. For a brief illustration of the formalisms adopted in metrical phonology, see section 11.13 below, and for detailed examinations of English stress assignment within the earlier stages of the metrical tradition, see Hogg and McCully (1987) and Goldsmith (1989, esp. ch. 4). Gussenhoven and Jacobs (2005, ch. 13) is highly recommended, along with Kenstowicz (1994, ch. 10) and Hayes (1995), for later versions of the framework.

9.8 Intonation in English

The importance of English intonation, both as an area of difficulty for the foreign learner and as a challenge to theory and description, has been acknowledged in a number of classic studies. Among the works prompted by the needs of learners are Pike's outline of American English intonation (1945) and treatments of British intonation by O'Connor and Arnold (1973) and Halliday (1970). Pike (1945, pp. 3–18) includes a survey of work prior to his own, and Crystal (1969) is a detailed account of English, which spans a wide range of prosodic features and pays thorough attention to relevant work both inside and outside linguistics. More general accounts of intonation are Lieberman (1967), Bolinger (1972) – which is a collection of papers that includes extracts from works mentioned above as well as treatments of languages other than English – and Cruttenden (1997). In recent years, several researchers have turned their attention to the role of intonation in discourse: this perspective is reflected in, for example, Brazil et al. (1980) Brown et al. (1980) and Wichmann (2000). Dialectal variation in intonation, particularly in the British Isles, has also been explored in recent years (e.g. Grabe et al. 2000). There is a significant body of work on English intonation that falls within the 'autosegmental-metrical' tradition, tying together elements of metrical phonology, briefly mentioned above, and autosegmental phonology, described in section 11.12 below (Pierrehumbert 1980, Gussenhoven 1984, Beckman and Pierrehumbert 1986, Ladd 1996, Beckman et al. 2005). A brief overview of the main features of one of these models for English will be presented in section 9.9.

Intonation is often described, somewhat impressionistically, as a matter of 'musical features' or speech 'tunes or melodies' (O'Connor and Arnold 1973, p. 1). While this may be a useful nontechnical pointer, it is sometimes linked with a conception of intonation as something superimposed upon the intrinsic meaning of words themselves, conveying the speaker's attitude rather than any fundamental meaning (Pike 1945, p. 21; O'Connor and Arnold 1973, p. 2). It is true that the intonational features of utterances – including such aspects as overall pitch setting – and features such as overall tempo signal what may

loosely be summarized as 'attitudinal' factors, such as the speaker's anger or tiredness. It would nevertheless be an injustice to English intonation to suggest that it does no more than provide an overlay of feelings or emotions. It is in fact a crucial part of the English language, carrying important semantic and discourse and/or pragmatic functions. These functions may be 'attitudinal' in the sense that they express, for instance, definiteness or tentativeness, but these meanings are no more superimposed or extrinsic than other functional options such as whether to ask a question or make a statement, or whether to qualify a statement by including the word 'probably' or 'possibly'.

If we narrow the concept of intonation to exclude both basic rhythm (as determined by lexical stress patterns; section 9.6 above) and overall settings (such as faster or slower rate of utterance and higher or lower pitch range), there remain three ingredients that are central to English intonation: TUNE, or pitch pattern, TUNE–TEXT ALIGNMENT (e.g. how the tune is aligned with stressed syllables in words to make the words accented), and INTONATIONAL PHRASING. The first of these is a matter of the pitch patterns available in the system; the second and third can be taken together as aspects of the structural organization of utterances into units within which prominences are positioned. In contour-based models of intonation (e.g. Halliday 1970, Crystal 1969) the placing of the nucleus or nuclear accent in the tone group is referred to as TONICITY.

The fundamental tune choice of English is between rise and fall. The selection is highly functional and in the normal case the tune begins on the last primary stress of an utterance, the nuclear accent. Thus the following (with the fall marked \ preceding the relevant syllable) are complete or definite:

She lent him her \ CAR
Would you leave the \ ROOM
Do be \ QUIET.

Notice that although the wording of these structures is quite different, the final falling tune is significant in determining the interpretation. In particular, the second example is ostensibly a question but with falling tune is likely to count as an authoritative demand. But our rather vague assertion that these utterances are complete or definite becomes more meaningful when we consider the opposition between fall and rise. If the utterances have a rising tune (/), they will be interpreted as open-ended or indefinite, usually inviting response or reaction.

She lent him her / CAR
Would you leave the / ROOM
Do be / QUIET.

The first utterance is now likely to convey a surprised query ('did she really?'), and the second will be tentative or polite, as if the speaker is hesitant or unsure of the right to make the request, or at least willing to qualify that right. The third utterance will also sound tentative – despite the wording, which on the face of it, is pretty blunt. Indeed, it is precisely the kind of utterance that school

teachers are well advised to avoid with an unruly class: it has the appearance of authority but the rising tone will surely signal hesitancy or uncertainty.

Of course, it is somewhat artificial to isolate this simple choice between rise and fall from all the other options at a speaker's disposal, for we normally combine resources if we can. Thus, to achieve a polite request, we are unlikely to rely only on a rising tune but may add wording such as 'please' or 'would you mind . . .', and so on. Nevertheless, even the examples given here should be enough to suggest the inadequacy of comments to the effect that questions always have rising pitch and statements falling pitch. The system is both simpler than that – in that the fundamental opposition is between what Halliday calls the 'certainty' or 'polarity known' of the falling tone, and the 'uncertainty' or 'polarity unknown' of the rising tone (1970, p. 23; 1985a, p. 281) – and more subtle, in that this fundamental choice is combined with all the other options of wording that yield different interpretations of certainty and uncertainty.

The tune options are not limited to simple rise and fall. Nor are they limited in terms of degree. For example, low rise tunes contrast with high rises. The former is generally associated with some kind of 'continuation' function whereas the latter is often associated with questioning intonation in many varieties of English. Complex tunes are also possible. For example, tunes may be combined in a falling then rising pattern (fall–rise tune) in which the rise so to speak cancels or qualifies the definiteness of the fall (Halliday 1985a, pp. 281–3). Compare

She doesn't lend her car to \ ANYone (definite statement)
She doesn't lend her car to / ANYone? (querying the statement)
She doesn't lend her car to ∨ ANYone (qualified statement).

The implication of the fall–rise is that she doesn't lend her car to everyone ('not just ANYone') but may lend it rarely and exceptionally, say only to very close friends. In this sense, the definiteness of the fall is maintained but the open-endedness of the rise is added. English intonation also makes use of a rise–fall tune. If the final fall is subsituted with a rise–fall tune, this serves to emphasize the 'definiteness' of the proposition, by indicating that absolutely NO ONE is allowed near her car. Again, although these meanings might conceivably be described as performing some kind of pragmatic or discourse function, there is nothing vague or idiosyncratic about systemic distinctions which English speakers clearly use and recognize and which convey precise information about whether someone will or will not lend her car.

Tune choices are expressed within a highly organized structure. In the first place, the fundamental rhythm of spoken English is determined by the alternation between rhythmically prominent stressed syllables and weak unstressed syllables (section 9.6 above), and pitch accents are normally realized on the rhythmically most prominent syllable in a word, i.e. the primary stress – indeed the occurrence of the tone is part of what signals that the lexical stress pattern is maintained. But since the foot may contain unstressed syllables following the stress, the tune may spread over these unstressed syllables. Hence a fall, for

example, still marked below as before the relevant syllable, may actually be realized by successively lower pitch on each syllable of the foot. Compare:

Take the \ CAR
Take the \ CAMera
Take the \ CARamel

where in the last example the fall may be realized as three descending pitches over three syllables.

Secondly, the overall tune itself characterizes an intonational constituent or an intonational PHRASE or TONE GROUP. The number of nuclear accents in an utterance and its division into intonational constituents thus go hand in hand. A simple and common instance in English is where a descriptive word or phrase, in apposition, forms a separate intonational phrase echoing the one before. The boundary between the two groups (here marked ‖) is likely to be represented by a comma in written English:

He has two \ BROTHers ‖ in \ BRISbane.
(He has two brothers, who live in Brisbane.)

You mean his / FRIEND ‖ the / ARchitect?
(You mean his friend, who happens to be an architect?)

Contact the \ MANager ‖ who deals with com\PLAINTS.
(Contact the manager – he deals with complaints.)

If these utterances are spoken as single intonational phrases, the second element will no longer be in apposition as a kind of addition or afterthought but will be interpreted as a restrictive specification:

He has two brothers in \ BRISbane.
(He has two brothers in Brisbane – and possibly other brothers elsewhere.)

You mean his friend the / ARchitect?
(You mean the architect friend? – He may have other friends who are not architects.)

Contact the manager who deals with com\PLAINTS.
(Contact the manager who deals with complaints – not any of the other managers.)

Compare also the following, with two intonational phrases

I didn't \ TELephone ‖ because I was \ ANGry

and the single intonational phrase

I didn't telephone because I was ∨ ANGry . . .

In the first case the two phrase-final melodies, each ending a group, serve to divide the utterance so that it makes a statement (the speaker did not telephone) and gives the reason for this (the speaker was angry). In the second case, the single intonational phrase brings the reason within the scope of the negation, so that it is the reason that is denied, not the telephoning. This interpretation is reinforced by the fall–rise which signals a qualification – as Halliday puts it, 'there's a but about it' (1985a, p. 282). Hence we take this utterance to mean something like 'I telephoned, not because I was angry, but. . . .'

At the same time, the structuring of English intonation allows flexible placement of the tune itself. While the nuclear pitch accent on the final lexical stress can be taken as the normal or unmarked case, the nuclear accent can actually be placed on virtually any syllable (sometimes called 'sentence stress'; section 9.6 above). If it is not on the primary stressed syllable of the final word, the nuclear accent can serve a 'contrastive' function, or can change the 'focus' of the utterance, e.g.

He has \ TWO friends in London	(not just one)
He has two / BROTHers in Toronto?	(not sisters?)
He doesn't live \ IN Auckland	(but nearby).

Analysts often talk about 'broad focus' versus 'narrow' focus in this respect. English has the option of placing a 'focal' accent on any lexical item to bring it into focus.

Structural organization goes beyond the fundamentals noted here, and there is considerable complexity within intonational constituency in English. Lexical stresses may still be maintained and may also be associated with intonational targets or pitch accents – these are usually referred to as PRE-NUCLEAR accents. Tunes may be realized with, say, a rise commencing on an early accented word, with a final fall commencing on the nuclear syllable. Likewise, a fall can be separated from a rise. In other treatments of English intonation (e.g. Halliday 1970) these are called COMPOUND tones. Halliday also claims that in the second kind of compound tune, i.e. the fall–rise, the falling part of the tune has more 'prominence' than the final rise. Both tunes are illustrated below:

Do you want a / SNACK or a \ MEAL?

Do you want a \SNACK or a /MEAL?

The simple kind of notation adopted here, with tunes shown by conventionalized devices, and intonational phrases separated by boundary markers, is adequate to show the basic options. It does not reveal the details of how a pitch fall may be distributed over several syllables, or of what is happening in the rest of the intonational phrase, nor does it cope well with background variables, such as a general raising or widening of the pitch range over certain stretches of discourse. For such purposes, more intricate notations may be used, mirroring more closely the actual contours but at the risk of obscuring the systematic choices underlying them. Pike (1945) provides extended passages of English marked with a line notation, more or less as follows

Can you see me?

Pike couples this with a numbering system, in which the numbers 1–4 indicate relative pitch height. Numbering is common in older American publications, not without some confusion between levels of stress and levels of pitch. British authors, such as O'Connor and Arnold (1973) and Gimson (1980), have generally preferred the 'tadpole' notation (more correctly 'interlinear tonetic') for relatively detailed transcription. Crystal (1969), however, offers a notation which is based on simple stylized symbols (such as / and \ for tones) but also includes pitch range markings (such as arrows to indicate raising or lowering) and even allows for musical-style signatures (such as 'forte' and 'crescendo') at the beginning of an utterance. In section 9.9, we outline one set of annotation conventions that are widely used in intonational analyses of English today.

9.9 Tones and break indices

An increasingly important approach to intonational phonology analyses or decomposes the characteristic tunes of English into separate tone targets which are aligned in particular ways to elements of an utterance (recall the term 'tune–text alignment', introduced in section 9.5 above). This is the approach adopted within the autosegmental-metrical framework, a term adapted by Ladd (1996) to refer to these kinds of intonational models. In fact, many of the conventions used in transcription of tone languages (see section 9.4 above) are employed by these models to transcribe English intonation, especially the use of just two pitch levels to describe the tonal patterns of a phrase. There are competing analyses of English intonational phonology within this framework (e.g. Gussenhoven 1984, Grabe 2001), but only key elements of one annotation system will be described here, namely TONES AND BREAK INDICES (ToBI) (Beckman and Ayers-Elam 1994, Beckman et al. 2005). This is a simplified version of a model of American English intonation originally developed by Pierrehumbert (1980), and subsequently revised by Beckman and Pierrehumbert (1986). The tones component refers to the tone targets that are used to annotate an intonational tune, and the break indices annotate the degree of prosodic juncture between units of different types (i.e. words, intonational phrases). A brief outline of both components is given below. For a detailed description of the evolution of this model, see Beckman et al. (2005), and for a highly detailed overview of ToBI annotation criteria with numerous speech examples, see Beckman and Ayers-Elam (1994).

Some of the main terminology associated with this intonation model has already been introduced in sections 9.3 and 9.7 above. The characteristic intonational melodies of English are broken down into H(igh) and L(ow) tones that either associate with rhythmically stressed syllables and perform an accentual function, or function as edge tones that mark the right edge of two intonational constituents, i.e. the intermediate (intonational) phrase and the higher-level intonational phrase. The intonational phrase is more or less synonymous with tone group or tone unit (see Crystal 1969, O'Connor and Arnold 1973),

and the intermediate phrase is sometimes called the phonological phrase or minor prosodic phrase in other treatments of intonation (see section 11.17 below).

An important feature of this model is that no distinction is made between a nuclear pitch accent and a pre-nuclear pitch accent, unlike in the British models (e.g. Crystal 1969). The tones associated with all accents are selected from a single inventory of pitch-accent shapes. 'Nuclearity' is indicated by position in the phrase (i.e. the nuclear accent is the last pitch accent in an intermediate phrase) and is signalled by a following phrase tone and/or a boundary tone. The phrase tone has multiple functions in a sense, because it serves to describe what the pitch is doing immediately after a nuclear accented syllable, as well as denoting the right edge of the intermediate phrase. The boundary tone marks the absolute right edge of the higher-level unit, the intonational phrase. This decomposition of tone 'functions' makes this model quite different from approaches that are based on dynamic descriptions of contours (e.g. Crystal 1969, Halliday 1970, O'Connor and Arnold 1973). These models posit different inventories of 'tones' for different structural positions in a tone group, and there is no obvious division into the 'accent' component and the 'boundary' component of an intonational tune in the annotation conventions. We could of course argue that these functions are implied in the contour-based models in that a falling nucleus, for example, implies a pitch fall that starts in and around the nuclear accented syllable and terminates at the right edge of the tone group. Another important difference is that ToBI analyses are always done in conjunction with the acoustic waveform and F_0 contour. The latter is considered to be the 'obligatory phonetic representation' of the tonal component of a ToBI analysis of English (Beckman et al. 2005, p. 37). But auditory impressions are also important in intonational analysis.

There are five major kinds of pitch accent that are included in the ToBI tones inventory. Recall that in English (unlike Japanese, for example), it is a pragmatic choice on the part of the speaker to accent a word, and to use a particular pitch-accent shape. In the following transcription conventions, the diacritic '*' after the L or H symbol indicates that this is the tone aligned to the stressed syllable. The '+' indicates a composite tone (section 9.7 above), although the tone marked with the * is the one that is anchored or aligned to the stressed syllable.

The main pitch accents of English can be summarized as below (after Beckman and Ayers-Elam 1994, Beckman et al. 2005; see also Cruttenden 1997, pp. 59–61).

1 H*: a high-pitch target realized in the relatively high part of a speaker's pitch range (see section 9.2 above). It is usually associated with a turning point or high peak in an F_0 contour. According to the ToBI transcription conventions, this accent type also subsumes the H*+L accent that was originally part of the Pierrehumbert (1980) pitch-accent inventory. The simple H* pitch-accent is probably the most common pitch-accent type observed in General American English, and in other English varieties like Australian English.

2 L*: a low-pitch target realized relatively low in a speaker's range that corresponds to a trough or valley in an F_0 contour.
3 L+H*: the 'rising' accent. The H* target is aligned to the stressed syllable but there is a clear rise through the first part of the syllable to a relatively high F_0 peak. This accent often marks out a word as particularly contrastive or associated with narrow focus (Hirschberg 2002). It often sounds 'assertive'.
4 L*+H: the 'scooped' accent. This is another kind of rising accent. The L* target is aligned to the stressed syllable and the rise may continue into the following syllable, where the F_0 peak may be observed. The difference between L+H* and L*+H is one of phonetic alignment of the starred tone in English. When the rise starts very early in the stressed syllable, listeners are more likely to hear the accent as an L+H*, whereas the later the alignment of the F_0 peak, the more likely it is that an L*+H accent will be heard (Pierrehumbert and Steele 1989). This accent too is distinctive and often conveys a sense of 'incredulity' or 'uncertainty' (Hirschberg 2002).
5 H+!H*: this pitch accent indicates a lowered high accent. The accent-bearing stressed syllable is preceded by a sharp drop in F_0 from a very high point at the beginning of the intonational phrase. The 'H+' in this sequence indicates this stretch of preceding high pitch that starts at the left edge of the intonational constituent, and drops down to the !H* accent. The '!' indicates that the H* accent is downstepped (section 9.4 above) and is realized with a lower F_0 level relative to the preceding unaccented material. According to Hirschberg (2002), this pitch accent suggests that there is a sense of 'familiarity' with the accented item.

In fact any of the H(igh) pitch accents (i.e. H*, L+H* and L*+H) can be downstepped. Strong pitch range downtrends that are sometimes observed throughout an intonational constituent are often largely due to the presence of downstep, rather than declination (section 9.2 above). Downstep is always indicated by placing the '!' diacritic in front of the H tone of the accent (e.g. !H* L+!H* L*+!H). Cruttenden (1997) also refers to these as stepping accents. The effect is similar to the step-like pattern that is observed in many register tone languages (section 9.4 above).

Downstep is generally described as a phonological effect that is triggered by a preceding 'bitonal' or composite accent (i.e. either L+H*, L*+H or the underlying H*(+L) pitch accents). A downstepped accent never commences a downstepped accent series unless it is an H+!H* accent. However, recent corpus-based analyses of connected speech suggest that a preceding bitonal does not *always* trigger downstep of the following accents (Pierrehumbert 2000). Moreover, the function of downstep is somewhat debatable. Some claim it is akin to deaccenting a word associated with the downstepped pitch accent. There is continuing debate about this and many other aspects of these models (see Ladd 1996, ch. 3; Gussenhoven 2004).

H(igh) and L(ow) pitch levels are also used to transcribe phrase accents and boundary tones. Boundary tones are transcribed with a '%', and mark the absolute right edge of an intonational phrase, whereas phrase accents are transcribed with a '-' and indicate the edge of an intermediate phrase. To mark the edges of intonational phrases (the largest constituent), both a phrase accent and a boundary tone are annotated. Implicit in the two-tone transcription of intonational phrase boundaries is that these boundaries also coincide with the right edge of an intermediate phrase. In other words, an intonational phrase can also consist of one or more intermediate phrases. The difference between an intermediate and an intonational phrase boundary is often described in terms of the level of DISJUNCTURE. Compared to intermediate phrase boundaries, intonational phrase boundaries generally coincide with more pronounced lengthening of final syllables as well as having the extra boundary tone. In sum, the phrase accents H- and L- correspond to relatively mid/high or low F_0 values that indicate the right edge of intermediate phrase boundaries respectively, and they also describe what the pitch is doing after the nuclear accented syllable (i.e. whether it is staying level, rising or falling).

Different combinations of phrase accents and boundary tones give rise to the characteristic intonational boundary types of English. The main combinations are L-L% for a low boundary, corresponding to a low point in a speaker's pitch range (e.g. the endpoint of a final fall), and L-H% for a low rising boundary, which starts from a low point in the speaker's range (corresponding to the L-) and rises to mid level or higher (the H% boundary). One of the other major boundary configurations is H-H%, the high rising boundary, which starts at a relatively high pitch level (corresponding to the H-) and continues to rise into the highest part of a speaker's pitch range. This illustrates another major feature of the H- phrase tone. In this model of intonation, it has the power to UPSTEP or raise the actual pitch realization of any following boundary tone. So in intonational boundaries annotated as H-H%, pitch continues to rise after the H-phrase tone to an even higher pitch level at the edge of the intonational phrase. The upstep rule is also important for another boundary configuration, H-L%. This configuration of tones is used to account for plateau-like boundary tunes. The H-phrase accent upsteps the following L% boundary, resulting in a mid-to-high-level sustained tune. This is admittedly a very abstract representation of this type of tune, and the reader is directed to Ladd (1996) for a good discussion of this.

As well as triggering upstep of a following boundary tone, an H-phrase tone can be downstepped when it follows one of the bitonal accents (see above). This means that we can add !H-, !H-H% and !H-L% to our inventory of phrasal and boundary tone options. A final important pitch-range-related ToBI convention is the 'HiF_0' label. This is usually assigned to the highest F_0 level associated with a high-tone pitch accent (including the composite accents, but excluding segmental perturbation effects described in section 9.2 above) in an intermediate phrase. It allows the analyst to track pitch-range variation across a series of phrases in connected speech.

Ladd (1996, p. 82) gives a good summary of how traditional contour descriptions of English nuclear tones can be translated into tone-based transcriptions. Figures 9.9.1 and 9.9.2 show six versions of the name 'Mary', transcribed according to some of the ToBI principles outlined above. The examples here are from General Australian English. The ToBI inventory, proposed for

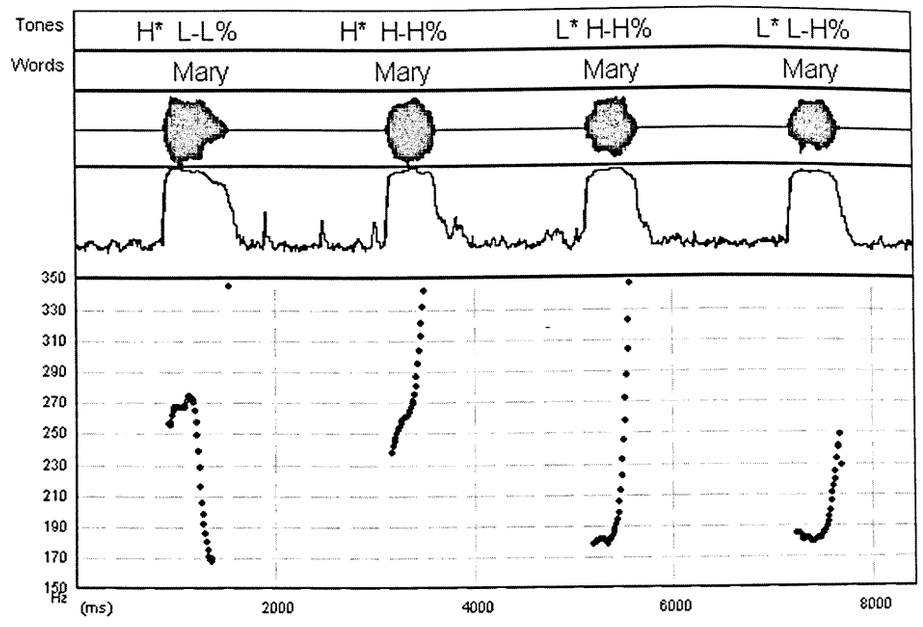

Figure 9.9.1 Major tunes in English

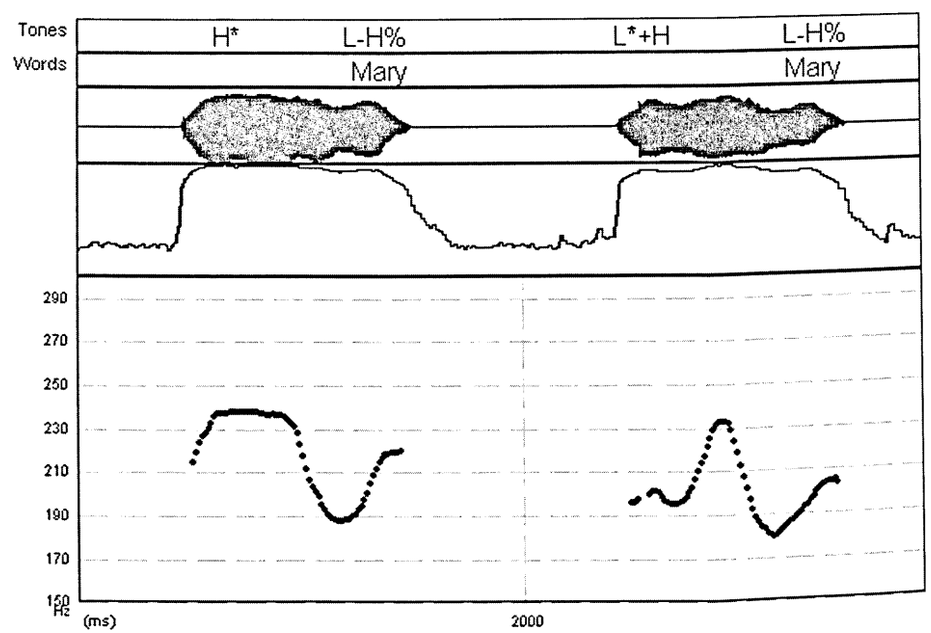

Figure 9.9.2 Complex rises in English

Mainstream American English, has been used to model this variety of English (Fletcher and Harrington 2001).

In figure 9.9.1, the first F_0 contour shows a 'falling' tune, which is transcribed H* L-L%. The H* is aligned with the first stressed syllable of Mary, while the L-phrase tone accounts for the pitch pattern between the pitch accent and the final low boundary tone L%. In fact all of the earlier illustrations using the '\' can be transcribed as an H* aligned with the stressed syllable of the accented word, followed by L-L%, which accounts for the rest of the tune. The three remaining F_0 contours in figure 9.9.1 show different kinds of simple rises. The second F_0 contour is a high rise, starting with a H* pitch accent and rising further because of the H- and H% tones. The third F_0 contour shows another high rise, this time with a low pitch onset, coinciding with an L* pitch accent. As discussed earlier, high rising tunes are usually interpreted as questions in many varieties of English, but in some varieties (for example Australian English and New Zealand English), high rises are often used by speakers to terminate declarative statements (Fletcher and Harrington 2001, Warren 2005). The fourth F_0 contour in figure 9.9.1 shows a low rising tune, commencing with a low F_0 target in the speaker's range, corresponding to an L* pitch accent, but the rise now terminates at a pitch level somewhat lower than the previous contour. This particular tonal combination (L-H%) is often called a continuation rise.

Figure 9.9.2 shows two kinds of complex tune: a fall–rise and a rise–fall–rise. In both cases the final rise component is indicated by the combination of the L-phrase tone and the final H% boundary tone, which is realized over the second syllable of 'Mary'. The fall of the fall–rise is shown by the early falling F_0 pattern, from the F_0 peak of the H* pitch accent on the stressed syllable of 'Mary' to the low F_0 trough of the L-phrase tone located around the beginning of the second syllable. In the rise–fall–rise, the scooped accent L*+H is realized as the initial low F_0 trough, located around the initial syllable of 'Mary', followed first by a sharp rise in F_0 late in the stressed syllable, then a fall to a relatively low F_0 level corresponding to the L-phrase tone. The final part of the tune is realized in a similar fashion to that of the fall–rise, with a rise in F_0 to mid level on the final syllable.

The two different accent types in figure 9.9.2, in combination with the same LH final boundary configuration, give rise to quite different pragmatic effects. For example, the second tune generally expresses 'uncertainty' or 'uncertainty about a scale' or 'incredulity' (after Hirschberg and Ward 1992), whereas the first tune is often given interpretations like 'limited agreement' (e.g. Roach 1991) or may be used in 'polite' questions in Standard British English (in fact Halliday 1970 calls this tune the 'questioning' fall–rise). For a more advanced treatment within a discourse framework of the different components of many of the tunes described so far, see Pierrehumbert and Hirschberg (1990).

The break indices conventions provided by the ToBI schema also allow the analyst to transcribe major elements of prosodic constituency (see section 11.17 below). Briefly, the break index tier of ToBI has five main values from 0 to 4. A break index value of 0 is used where there is a clear phonetic indicator of

a clitic group, for example the affricate in many pronunciations of 'did you' as 'didja'. A value of 1 is used to indicate the right boundary of a word that is phrase-medial. An index of 2 is used when it is not obvious if there is a phrase boundary or not. Index 3 marks an intermediate phrase boundary. A full intonational phrase is marked as 4. As mentioned above, there is a final boundary tone after the phrase accent, as well as phrase-final lengthening and often a silent pause. There may also be glottalization at the onset of a level-4 consituent that is not present at the start of a level-3 constituent (see Shattuck-Hufnagel and Turk 1996 for a survey of these additional boundary affects). The degree of disjuncture is perceived to be greater between a 4 and following phonetic material than between a unit marked as a 3 and following material.

Elements of the intonational model underlying the ToBI conventions have been questioned in recent years. For example, the distinction between the L+H* and H* pitch accents is often difficult to discern in spontaneous speech, or when speakers use a particularly narrow pitch range. The status of the phrase accent has been the subject of much debate, as mentioned above. Not all autosegmental-metrical analyses of intonation posit two levels of intonational phrasing (e.g. Gussenhoven 2004), and other inventories of pitch accents have been posited for different varieties of English (e.g. Grabe et al. 2000). However, Grice et al. (2000) suggest a provocative analysis of the function of the phrase accent in the context of intonational structure in English and other languages. The reader is directed to Ladd (1996, pp. 88–9) and Gussenhoven (2004, chs 14 and 15) for a discussion of the relevant issues.

Cruttenden (1997, ch. 3) and Ladefoged (2006, pp. 124–8) also provide brief outlines of the main features of ToBI. Jun (2005a) presents a collection of papers that show how ToBI-style transcription systems can be developed for a range of different kinds of languages, including 'tone' languages and 'pitch-accent' languages as well as 'stress' languages. Ladd (1996) and Gussenhoven (2004) also provide an advanced overview of issues in intonational phonology. The interpretation of intonation contours in English is notoriously slippery and context-bound, so it is no surprise that intonation has been referred to as a 'half-tamed savage' (Gussenhoven 2004, p. 57, after Bolinger 1978).

Exercises

1 Why is the distinction between segmental and suprasegmental phonetic phenomena not clear?
2 Why is it an oversimplification to equate stress in language with loudness?
3 Which speech production mechanisms contribute to English prosody?
4 What are register tone languages and contour tone languages?
5 What is meant by the terms 'syllable-timed' and 'stress-timed'?
6 Describe the two uses of the term 'pitch accent'.
7 How extensive is vowel reduction in your own pronunciation of English? In as natural a pronunciation as possible, what vowel do you say in the following:

   a. the first syllable of *confuse*, *correction*, *delicious*, *obtain*, *parade*, *production*?
   b. the second syllable of *antelope*, *asterisk*, *crocodile*, *daffodil*, *parachute*?
   c. the final syllable of *carpet*, *bannister*, *bursar*, *daffodil*, *hostel*, *item*, *parrot*, *synod*?

8 List as many words as you can which are stressed differently by different people – for example, some people say *PREFerable*, some say *preFERable*. Can you relate the differences to differences in the speakers' age, sex, status or regional origin?
9 What ToBI conventions are used to differentiate intermediate and intonational phrases in English intonation?