

CHAPTER 1

Intonation

Pitch, loudness and length are among the most salient of the properties of speech perceived by the listener. Linguists sometimes refer to these features as suprasegmental, suggesting that they are somehow ‘above’ the string of consonants and vowels. This connotation is misleading: rather, in speech, the string of consonants and vowels is overlaid onto a base of phonation. Phonation, or voicing, is generated by an airstream from the lungs passing through the larynx, which results in the perception of fluctuations in pitch and loudness distributed over chunks of varying durations. The term *prosody*, and the related adjective *prosodic*, commonly refer in a broad sense to features of pitch, loudness and duration in speech, encompassing their use in individual words and their component syllables, as well as the use of these features over longer stretches of speech, i.e. phrases, complete utterances, conversational turns. These longer stretches are the main focus of this book. Over such longer stretches, the meaningful patterning of pitch, along with related patterning of loudness and length features, is commonly referred to as *intonation* – hence the title of the book. While some authors restrict the use of this term to the patterning of pitch alone, we use it as a shorthand to refer to communicatively relevant systems, operating over stretches of speech consisting minimally of a single word but usually longer, which are realized primarily through features of pitch, also often of loudness and length, and sometimes of voice quality and articulation.

From the very beginning, long before words made up of consonants and vowels can be identified, infants produce the prosodic features that are used for intonation. Moreover, adult carers respond to them as signalling various kinds of meaning. Infants, in their turn, appear to react to the prosodic components of adults’ speech addressed to them; indeed, adults systematically exaggerate some of these prosodic features when addressing infants and young children (see Chapter 6). These observations appear to be true of all cultures that have been studied so far, although the degree of modification varies across cultures. It appears then, that from birth, prosodic features form a set of resources that the child can exploit for communicative purposes, even though in different mature adult languages, prosodic features turn out to be organized in a wide variety of ways. Some of the differences will be described later.

Some children’s intonation develops in ways that are unusual for their linguistic environment. The basis for postulating an impairment in intonation is likely to be the listener’s auditory impression that the child’s use of prosodic features is in some way different from that of the speech community and cannot be attributed to other causes, e.g. being a

non-native speaker whose intonation in the second language is affected by the mother tongue. Beyond that, the identification, description and explanation of the impairment are influenced by the approach to analysis taken by the investigator, who may be a speech and language pathology practitioner or a researcher, for example. The basic tools of description are phonetic and linguistic. These complementary approaches will now be explained.

The phonetic approach

Typical and atypical prosodic development can be explored by using a range of methodologies that are based on auditory-perceptual and instrumental techniques. Instrumental techniques enable features of the speech signal to be recorded and measured reliably. Some of these techniques measure speech production directly. For example, using electrodes attached to the neck close to the larynx, the laryngograph (or electroglottograph) monitors the vibration of the vocal folds directly as they produce voicing (Abberton & Fourcin, 1997). While this can produce an accurate signal, and is not subject to interference from other noises (unlike a microphone recording), the need to wear a neckband attached by a wire to a computer militates against the recording of natural conversational speech, not least with young children.

For such reasons, the most common type of instrumental analysis used in intonation research is acoustic. An audio recording of the speakers is made via microphones. This can be done using digital audio or video tape recorders, a solid state digital audio recorder, or direct onto a PC or laptop. Conventional audio or video cassette recorders produce an analogue recording, which can subsequently be digitized using special software. The resulting digital audio files can then be analysed using one of the many computer speech analysis packages available. A much-used and freely downloadable program for acoustic analysis is Praat (Boersma & Weenink, 2014), which can be used to display and measure prosodic parameters.

However, acoustic analysis has its limitations too. The first, already mentioned, is that as soon as recording moves outside the artificial environment of the speech laboratory into children's everyday talking environments, for example, home and school, then there will be sources of noise that may be difficult to control – traffic, clattering of play bricks, and so on. It can be difficult and time-consuming to remove such noises from the recorded signal. One particularly important and tricky source of noise is voices of speakers involved in background conversations or even in the same conversation: in a review of the occurrence of overlapping talk, Kurtić, Brown, and Wells (2013) report that, even in relatively formal meetings, for up to 10% of the time, two or more speakers will be talking at the same time and that, in spontaneous adult conversation, up to 45% of all changes of turn contain overlap between the speakers. Overlap is thus a natural part of human spoken interaction; it can only be suppressed by using an artificial speech elicitation task such as a monologue or reading aloud. Although speech research is making progress in developing techniques for separating out the voices of different talkers in overlap, at present, the only satisfactory procedure is to record talkers on separate channels, with each talker using a close-fitting microphone attached to a headset. While this has been done successfully when recording adults in meetings, it is too invasive for research with young children. A commonly adopted response to this issue in child language research has been to omit instances of overlapping talk from analysis. However, this is not a satisfactory solution when analysing intonation because, as will be seen later, the occurrence of overlap is an important source of evidence in revealing how intonation works.

ACTIVITY 1.1

Aim: To explore the extent to which you can track different voices speaking in overlap.

This activity involves listening to a fragment from a recording of spontaneous conversation between English students. Listen to Media File 1.1. If possible, use headphones. You will hear two female adults, Daisy and Beth (these are not their real names), speaking at the same time. Daisy speaks first. Using normal orthography, write down what each speaker says.

To check how you did, refer to the Key to Activity 1.1 at the end of the chapter, where you will find a transcript of the recording, presented as Extract (1.1).

For these reasons, the auditory perceptual analysis and transcription of intonation remain important research and clinical tools. Although not as reliable as instrumental analysis for identifying the individual phonetic parameters of speech which has been recorded under ideal conditions, the human ear is able to track with some accuracy different voices speaking in noise, including in overlap. In the following activity you can decide how far you agree with that statement.

If you managed to transcribe some of the speech accurately from the mixed file, you may (after congratulating yourself!) like to consider why the task you have just done is extremely difficult for a computer-based speech recognition system. Thus, one of the skills that you possess is the ability to at least sort out the contributing voices when two or more speakers are talking simultaneously, even if it remains hard to work out the words being spoken. Most children appear to develop this skill, and it is likely that one of the aspects of speech that we learn to attend to in order to manage this is intonation. You can listen to the same extract again but this time, listen to each speaker separately. Daisy's contribution is presented as Extract (1.2), Beth's as Extract (1.3):

(1.2) used to hang out in Stevenage when I was younger summer yeah

(1.3) that'd be good we should do that over like summer

Unlike acoustic analysis software, the human listener does not habitually attend to separate phonetic parameters but to speech as an integrated whole. A skilled listener is able to take into consideration a range of different parameters that may be relevant to the realization of an intonation pattern on a particular occasion, for example, variations in loudness, duration, voice quality (such as creak or whisper), and even the way in which word-final consonants are released. Such features will only be evident in the results of acoustic analysis if the analyst is on the lookout for them and so has chosen the relevant instrumental settings for the analysis process. For these reasons, as when analysing other aspects of speech, a combination of instrumental and perceptual methods is ideal (Howard & Heselwood, 2011).

The perception of pitch is mainly determined by fundamental frequency (F_0), which is a property of the acoustic signal. In the case of speech, F_0 relates to the number of periods of vocal fold vibration per second. It is most commonly measured in Herz (cycles per second) and forms an important part of the investigation of intonation. Figure 1.1 depicts an acoustic analysis of Beth's contribution to the conversational fragment from Activity 1.1. The dark line depicts changes in F_0 in the course of the utterance, the F_0 scale in Herz being given on the left-hand vertical axis of the graph. The most noticeable features are the big jump-up in F_0 , of about 80 Hz, from "should" to "do" and the subsequent fall of approximately 100 Hz from "do" through "that".

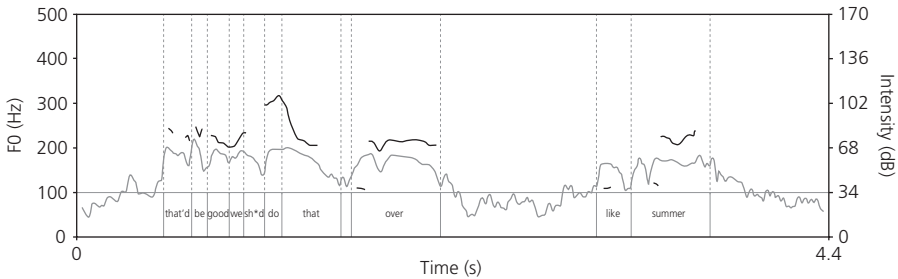


Figure 1.1 Acoustic analysis of Beth's turn from Activity 1.1 (Extract 1.3).

There is a disadvantage to using the linear Herz scale, particularly when dealing with children's intonation: the problem is that the relationship between F0 and the perception of pitch is not linear. This means that a greater Herz interval is needed when high in the pitch range, in order to produce a perceived pitch step, compared to the same pitch step produced low in the pitch range. Thus, a semitone interval produced by an adult male in the lower end of his usual pitch range may be around 5 Hz, while a semitone interval produced by a young child in the upper part of his usual pitch range may be around 20 Hz. In order to facilitate comparison between child and adult intonation throughout this book, F0 measurements are therefore usually expressed in semitones, which is a nonlinear scale, rather than in Herz. In Figure 1.1, the fall on "do that" is approximately seven semitones.

Just as the perception of pitch is determined to a large extent by the fundamental frequency of the signal, so the perception of loudness is largely determined by the intensity of the signal. This can be quantified in a number of ways for collections of utterances, such as mean, median or range of intensity, measured in decibels. When studying the intonation of an individual utterance, it is the loudness of a syllable relative to the rest of that utterance that is most relevant, since key elements of intonation are linked to rhythmically prominent syllables and loudness is one factor in perceived rhythmic prominence.

Intensity measurements outside the laboratory are not very reliable, since reliability depends on a constant distance between mouth and microphone. While this can be achieved by using a close-talking microphone mounted on a headset, as was the case when recording the conversation between Beth and Daisy used in Activity 1.1, this is rarely feasible when recording children in a naturalistic setting. Nevertheless, over the course of a short utterance, the speaker's distance from the microphone is unlikely to vary greatly, so the intensity trace, in combination with careful listening, can be used to help identify the words or syllables that are relatively loud. In Figure 1.1, variation in intensity within the utterance is depicted by the pale grey line beneath the dark F0 line. For example, the intensity on the stressed syllable of the word "over" is about 5 dB lower than on the accented syllables ("that", "good", "do") of the preceding stretch; it drops approximately 3 dB more to the final two words "like summer".

The third main aspect of prosody relates to the temporal domain of speech, and includes the duration of whole utterances, of words, syllables, consonants and vowels; the duration and location of pauses; articulation rate (often measured by number of syllables per second, excluding pauses), including the speeding up and slowing down of rate. Where a clear signal can be obtained for acoustic analysis, it is relatively straightforward to measure most of these temporal parameters using a speech analysis package

like Praat. In Figure 1.1, for example, there is a silent interval of around 0.8 seconds between “over” and “like summer”. There are also variations in the articulation rate (Dankovičová, 1997). The phrase “that’d be good we should do that” is produced in just one second, i.e. a rate of 8 syllables per second. By contrast, even if we ignore the pause between “over” and “like”, the five-syllable phrase “over like summer” has a rate of approximately 4 syllables per second, which is half the rate of the first phrase. At the bottom of Figure 1.1, spacing reflects the relative length of each word.

The prosodic transcription

Measurement of aspects of the speech signal using acoustic analysis is a valuable aid to the accurate description of prosodic features. It is essential where the aim is to make statistical generalizations from large amounts of data, for example, about the habitual fundamental frequency range or the average articulation rate of an individual speaker or of a group. However, auditory perceptual judgement is also an invaluable tool. Most of the examples discussed in this book are taken from more or less naturalistic interactions involving children and adults, where overlapping talk is frequent and where the recordings were made under less-than-ideal conditions. Prosodic features were first transcribed impressionistically, as were segmental aspects of the talk in some cases, particularly where the child has immature or atypical speech. Key portions of the recordings were later subjected to acoustic analysis, using Praat. However, because of the quality of the original recordings and characteristics of the children’s vocal productions, it was not always possible to obtain a reliable acoustic record. The transcripts represent the convergence of results from both these procedures, with the aim of keeping them readable. The International Phonetic Alphabet (IPA), and the conventions used in Conversation Analysis, provide ways of notating these prosodic features in a transcription. Pitch height and pitch movements are presented iconically between staves, the staves representing the upper and lower limits of the speaker’s habitual pitch range. Loudness, pause, duration and tempo features are notated using symbols and diacritics of the IPA, including extensions. Examples of this approach to transcription will be found later in the chapter. A full list of transcription conventions can be found in Appendix 1.

The linguistic approach

While the phonetic approach is concerned with the accurate description of prosodic parameters in speech, using auditory perceptual and instrumental techniques, the linguistic approach to intonation shows how these features communicate meaning. Prosodic features serve to realize linguistic systems such as tone (in tone languages), stress and intonation. From this perspective a prosodic impairment impacts on the linguistic system in question, with the result that the speaker’s meaning, in its broadest sense, may be obscured. Identification of a linguistic impairment of intonation is therefore dependent on the analyst having a description of the intonation system of the language. The intonation of English has probably received more scholarly attention than that of any other language, though systematic descriptions are now available for a large number of different languages (Hirst & Di Cristo, 1998; Jun, 2005).

Various approaches to the systematic description of intonation have been adopted. The ToBI notation, derived from autosegmental-metrical theory, has been widely

adopted in phonetic and speech technology research (Jun, 2005; Ladd, 2008). However, it has so far been little applied to clinical analysis or to studies of children's intonation development, where versions of more traditional analyses have tended to be used, following the lead of David Crystal (Crystal, 1987; Snow & Balog, 2002). As a necessary foundation for the rest of the book, the next part of this chapter provides a descriptive framework for English intonation, based on this more traditional approach.

English intonation: a brief introduction

The term *intonation* is generally used to refer to the linguistic patterning of pitch height and movement, together with loudness and duration, to realize meanings that are additional to the meanings conveyed by words and grammar. When pitch is used in lexical and/or morphological systems, for example, to distinguish word meanings, we talk of a *tone language*. Well-known examples include Thai and Mandarin Chinese, which will be discussed in Chapter 6 in the context of young children's need to work out whether or not the language they are having to learn is a tone language. The linguistic use of pitch in almost all varieties of English is intonational only. This description applies to Southern Standard British English. While many accents of English, including General American, have systems that are very similar, others are very different, for example, in Northern Ireland or among many speakers of Afro-Caribbean ethnicity. Dialect variation is discussed at various points in the course of the book, in relation to children's development.

Traditional approaches to English intonation generally identify basic intonation patterns from which the speaker can select. These patterns will be referred to as (nuclear) Tones. In some current research on intonation, it is usual to talk about pitch accents rather than Tones. Ladd (2008) discusses equivalences between the two descriptions. However, in this account we will follow the tradition of using descriptive labels and iconic diacritics, as in Table 1.1.

As illustrated in Table 1.1, the diacritic representing the Tone is placed in front of the Tonic syllable (described below). The number of meaningfully different Tones in English intonation continues to be the subject of debate (Cruttenden, 1997; Gussenhoven, 2004; Ladd, 2008). Our own approach to this issue is presented in Chapter 4. When the word 'Tone' is used with an initial upper case letter, as here, it signifies that the word is being used with a specific meaning as defined in this book. The convention also applies to other technical terms that are defined in this chapter.

The unit of intonation at which the Tone system operates is now usually referred to as the Intonation Phrase (IP), although tone unit or tone group used to be common alternatives. Its boundaries are marked by double vertical lines ||, the IPA symbol for a major intonation boundary.

To illustrate the transcription of English intonation, we return to Daisy's turn from the fragment of conversation used in Activity 1.1. An acoustic trace is presented in Figure 1.2.

Table 1.1 Diacritic notation for Tones in English intonation, as used in reading transcriptions in this book.

| Fall | Rise | Fall-rise | Rise-fall |
|------|------|-----------|-----------|
| ˘YES | ˆYES | ˘ˆYES | ˆ˘YES |

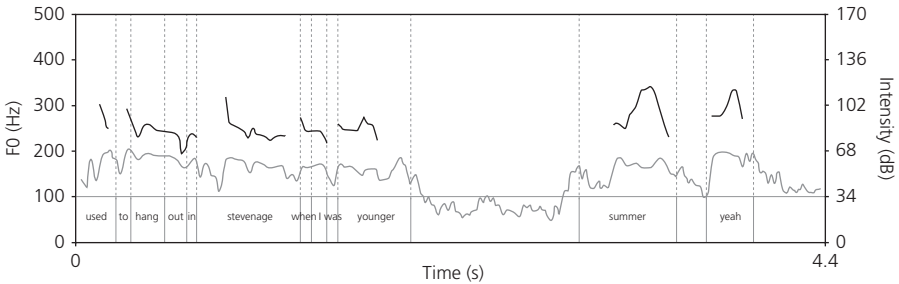


Figure 1.2 Acoustic analysis of Daisy's turn from Activity 1.1 (Extract 1.2).

The Intonation Phrase (IP) is the fundamental structural unit of intonation, and indeed may be viewed as the basic unit of speech production. In Extract (1.2.1) there are three Intonation Phrases:

(1.2.1) || used to hang out in Stevenage when I was younger || summer
|| yeah ||

IP boundaries are realized by the following features, in varying combinations (cf. Cruttenden, 1997):

- 1 occurrence of the Tonic, which carries the Tone (see below). The Tonic is not always located at the end of the IP, but it must have occurred for the IP to be heard as complete.
- 2 lengthening of final syllable(s) of the IP;
- 3 pause following the IP;
- 4 jump in pitch from the end of one IP to the start of the next IP.

Feature (3), a silence of around one second, is evident following “younger”, suggesting a boundary there. Feature (1), the Tonic, is the one criterion for an IP that is obligatory. The Tonic is recognized primarily by the Tone, a noticeable pitch movement on the stressed syllable of a word which makes that syllable stand out from what precedes and follows it. In addition, the Tonic is often relatively long and loud compared to surrounding words. In the first IP, the word “Stevenage” is a good candidate for a Tonic, as it has a fall of five semitones on the first (stressed) syllable; no other syllable has a comparably large pitch movement until we get to “summer” (rise-fall over five semitones) and “yeah” (rise-fall spanning three semitones). “Stevenage” is relatively loud compared to the preceding and following words and also its first syllable, measuring 150 milliseconds, is 50% longer than any other syllable in the IP until “younger”, which itself illustrates feature (2). On this basis, we have added Tone diacritics in (1.2.2), which indicate not only the pitch direction but also the fact that the following word is the Tonic:

(1.2.2) || used to hang out in ^ˆStevenage when I was younger || ^ˆsummer
|| ^ˆyeah ||

In the transcription system used here, a distinction is made between an ordinary Tonic and a Supertonic, notated by \uparrow placed before the Tone diacritic. The Supertonic has extra prominence, often from a combination of a particularly big pitch movement, noticeable loudness and extra lengthening, compared to an ordinary Tonic; while the surrounding words may be correspondingly quiet and fast, with very little pitch

movement. In many publications, the presence of a Supertonic is indicated by underlining, italics or capitals, inviting the reader to use the features just listed if reading aloud. Here is an example from the novel *Atonement* by Ian McEwan. Bryony, aged 13, is coaching her young cousin to read the lines of a play she has written:

(1.4) “It’s a question”, Bryony cut in. “Don’t you see? It goes up at the end.”
 “What do you mean?”
 “There. You just did it. You start low and end high. It’s a *question*.”

(McEwan, 2001: 33–34)

The italics on the final word of the final sentence contrast with the earlier presentation of the same words at the beginning of the extract. They can be transcribed as in (1.4.1):

(1.4.1) || 'its a `question || versus || 'its a ↑`question ||

The functions of the Supertonic will be explained in Chapter 3, while Bryony’s misapprehension about the intonation of questions will be discussed in Chapter 4.

The IP can be subdivided into constituent parts. Minimally, it must have a Tonic. Any syllables that follow the Tonic are known as the Tail. The pitch of the Tail generally continues the direction of the pitch of the Tone, thus if the Tone is a rise, then the syllables of the Tail will continue to step up. Conversely, if the Tone is a fall, the syllables of the Tail will continue to fall until reaching the base of the speaker’s normal pitch range. In the first IP of Extract (1.2), the Tail consists of “when I was younger”.

The first prominent syllable of the IP is called the Onset. It is useful to make a distinction between a (neutral) Onset and a High Onset because this contrast has a specific role to play in the organization of talk, as will be explained in Chapter 2. In our transcription system, a High Onset is marked ↑, whereas a neutral onset is simply marked with the IPA stress symbol ' (see next section: Stress and Rhythm, for further explanation). Any words from the Onset, up to but not including the Tonic, form the Head. Any syllables preceding the Onset are known as the Pre-head. This gives the following structure:

|| (Pre-head) ' (Onset + rest of Head) ` Tonic (Tail) ||

Elements in brackets are optional. This means that a short utterance of a single monosyllabic word can be intonationally complete, that is, it can constitute a complete IP, if it carries a Tone, since it thereby becomes a Tonic, which is the minimal element for an IP. The final two IPs in (1.2.3) are examples of single-word IPs. The top row of (1.2.3) presents a systematic reading transcription of Daisy’s turn; beneath it is an indication of how it maps onto the constituent parts of the IPs:

(1.2.3)

| | | | | | |
|-------|-----------|------------|------------|---------|-------|
| used | to hang | 'Stevenage | when I was | ^summer | ^yeah |
| | out in | | 'younger | | |
| Onset | rest of | Tonic | Tail..... | Tonic | Tonic |
| | Head..... | | | | |

IP boundaries frequently indicate a potential place for a change of speaker, referred to as a turn transition relevance place (TRP). For this reason, intonation is important for the regulation of turn-exchange. This is explained in detail in Chapter 2. Now complete Activity 1.2 before reading further.

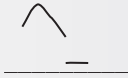
ACTIVITY 1.2

Aim: To discover how some of the concepts introduced in this section can be used when analysing talk involving a young child.

Study Extract (1.5), a short interchange between Robin (CA 1;9), and his mother. Then answer the questions that follow. Robin and his mother are engaged in play with a jigsaw puzzle board, into which pieces are to be fitted. At this point, his mother is quizzing him about one particular piece, depicting a train. Mother's turns are transcribed using the system for adult English intonation described under 'The linguistic approach'. Robin's turns are transcribed in the way described under 'The phonetic approach', since at this point in his development we cannot assume that he is operating with the mature adult intonation system.

(1.5)

1 M: ||↑ 'whats ↑`this bit 'called though ||



2 R: ɕi:j akəl'

{f}

3 M: || it ˘isnt || its 'called a ↑`funnel ||



4 R: f a f a

{f}

5 M: || thats ˘right || n ↑`what comes 'out of the ˘funnel ||



6 R: ʔə mɛ ʊ kʰ

- 1 Where is the Tonic located in line 1? Why do you think that Robin's mother has placed it there?
 - 2 In line 3, there are two IPs. Why do you think the boundary between them occurs where it does?
 - 3 In line 3, the two IPs have different Tones. Why do you think Robin's mother has chosen these Tones?
 - 4 It is sometimes said that questions in English have rising intonation. In Extract (1.4), is there evidence to support this claim?
 - 5 Turn exchange proceeds smoothly through the extract: Robin's mother does not start talking before Robin has finished. Why do you think there are no overlaps?
- Compare your answer to the key to Activity 1.2 at the end of this chapter.

Stress and rhythm

Smaller units may be identified within an IP. This can be appreciated by reading aloud the phrases in Extracts (1.6) and (1.7):

(1.6) CHOCOLATE-CAKE AND HONEY

(1.7) CHOCOLATE, CAKE AND HONEY

Syntactically, (1.6) consists of a compound noun (cake made with chocolate) followed by another noun (honey). You are likely to have read it with an intonation structure such as in (1.6.1):

(1.6.1) || 'chocolate cake and `honey ||

Syntactically, in (1.7) there are three separate nouns, referring to three separate items of food. You may have read it with an intonation structure as in (1.7.1):

(1.7.1) || 'chocolate 'cake and `honey ||

There is a separate peak of prominence on “cake” that was not present in the compound noun in (1.6.1). The difference can be described in terms of number of units within the Head. In (1.6.1), the Head consists of a single unit, “chocolate-cake and”. In (1.7.1), the Head consists of two units: “chocolate”, then “cake and”. In order to provide a helpful characterization of these smaller units, it is necessary to consider the role of stress and rhythm, and their relationship to intonation.

Word stress

In English and many other languages, part of the stored phonological representation of a word of two syllables or more is its (word) stress pattern. This pattern is an abstract property of the word, deriving from the opposition between stressed and unstressed syllables. Most words have one stressed syllable, the remaining being unstressed. The IPA symbol for (primary) stress is the superscript vertical stroke ^ˈ positioned in front of the stressed syllable; unstressed syllables are not marked. In English, stress is fixed for each word, but the position of stress varies on different words. Because English has words of different numbers of syllables, ranging from one to five or more, when these are combined with different stress locations, the number of possible, and indeed actual, patterns is considerable. Some examples are presented in Table 1.2.

The stress pattern of a word is very salient: along with number of syllables, it is one of the features most likely to be retained, for example, when we make a slip of the

Table 1.2 Examples of English lexical stress patterns.

| | Stress on 1st | Stress on 2nd | Stress on 3rd |
|-------------|---------------|----------------|----------------|
| 1 syllable | 'CHEESE | — | — |
| 2 syllables | 'POSTMAN | GUI 'TAR | — |
| 3 syllables | 'TELEPHONE | BA 'NANA | KANGA 'ROO |
| 4 syllables | 'CATERPILLAR | KA 'LEIDOSCOPE | PANO 'RAMA |
| 5 syllables | — | EX 'TERMINATOR | UNBE 'LIEVABLE |

ACTIVITY 1.3

Aim: To investigate informally the effects on pronunciation of increasing speech rate.

For this activity, ideally you will have access to an audio recording device, so that you can record and listen to your performance. If that is not available, just observe your performance in real time.

Repeat the following phrase five times, starting slow and careful, getting faster on each repetition, ending at your maximum speed:

(1.8) THE TELEPHONE IS LIKE A BANANA.

What happens to your pronunciation as you speed up?

Check your observations with the Key to Activity 1.3 at the end of the chapter.

tongue. To some extent, the position of the stress is predictable from the segmental content of the word (Roach, 2009) but overall the position of word stress is far less predictable in English than in many other languages, for example, Polish, where the stress is almost always on the penultimate syllable, or Czech, where it is generally on the first syllable (Laver, 1994).

This abstract pattern of word stress is realized phonetically when a speaker produces the word in an utterance. More strongly stressed syllables tend to be relatively loud. There is also a natural tendency for a syllable with strong stress to be longer in duration and higher in pitch, though these tendencies can be overridden in particular languages or dialects, for example, for speakers of the West Midlands accent of English, around Birmingham, the stressed syllable is characteristically lower in pitch than adjacent unstressed syllables. Thus loudness, pitch prominence and relative length are the main auditory correlates of stress. As we have seen, these are also the main auditory correlates of the Tonic.

Lexically unstressed syllables are realized phonetically as less prominent than stressed syllables. They are likely to be quieter, shorter in duration and less obtrusive in pitch: they are unlikely to carry much dynamic pitch change and are usually not far away, in the speaker's pitch range, from the preceding and/or following syllables. In addition, in English, the vowel in an unstressed syllable frequently has a more central quality, and is often schwa [ə]. While stressed syllables generally remain salient, unstressed syllables or their vowel nuclei may not always be realized phonetically in actual speech, particularly when the tempo of speech is increased or a more casual style is used.

Rhythm

In speech, phonetically prominent and non-prominent syllables are organized into patterns, which constitute the rhythm of the utterance. The phonetic features that convey rhythmic prominence include pitch, loudness, duration and vowel quality. These features were mentioned in the context of word stress in the previous section. It is important to make a distinction between, on the one hand, word stress as an abstract property of a word, that is, part of our phonological representation for that word, that we draw on for comprehension as well as production, and, on the other hand, the phonetic realization of stress within a particular utterance, as a peak of prosodic prominence that contributes to the rhythm of the whole utterance. It is the latter that provides the basis for the

intonation structure of the IP. Rhythmic patterns are thus a property of actual utterances rather than of individual words. The inherent stress pattern of a word may relate in various ways to the rhythm of an utterance in which that word occurs.

Rhythmic patterns vary widely across different languages and are often one of the most striking features of a foreign accent. There is also an increasing recognition of the importance of rhythm in young children's perception and acquisition of their first language (see Chapter 6).

The Foot

The unit of rhythm is the Foot. A Foot consists of one strong syllable, which may be followed by weak syllables. Foot boundaries can be marked by the IPA symbol for stress, a vertical stroke ' , which precedes the strong syllable beginning the Foot. Since the Tonic syllable, marked with a Tone diacritic, is always rhythmically strong, there is no need to mark it separately with a stress diacritic: it is implied in the Tone diacritic.

An IP can consist of a single Foot, e.g. a single word, of one or more syllables, as in (1.9) and (1.10):

(1.9) || `this ||

(1.10) || `funnel ||

The Foot that contains the Tonic can be preceded by one or more Feet, forming the Head. In (1.9.1), there is one Foot, "whats", preceding the Tonic, while in (1.10.1) there are two Feet before the Tonic: "what comes" and "out of the":

(1.9.1) || 'whats `this ||

(1.10.1) || 'what comes 'out of the `funnel||

The Foot beginning with the Tonic may be followed by one or more Feet, which form the Tail. In (1.9.2), "called though" is a Foot that forms the Tail of the IP:

(1.9.2) || 'whats `this bit 'called though ||

It is possible to have a silent beat, where there is no audible strong syllable at the beginning of the Foot but nevertheless both speaker and listener perceive a beat. The speaker may signal it by a gesture or nod. In this transcription system, the silent beat is symbolized by an underlined space, preceded by the Foot boundary mark. Such a Foot would constitute a Pre-head, as in (1.11):

(1.11) || ' its 'called a `funnel ||

Rhythm and grammar

In general, the rhythmically strong syllables in an utterance are more likely to belong to lexical items (noun, lexical verb, adjective or adverb) rather than function words (pronoun, preposition, auxiliary verb, article, etc.). In (1.11), CALLED is a verb and FUNNEL is a noun. In (1.12), TEDDY and NECK are both nouns:

(1.12) || whats 'teddy got round his `neck ||

FUNNEL and TEDDY are both disyllabic, the first syllable having the lexical stress. This then becomes the strong syllable of the Foot. Where the lexical item has more than one syllable, the strong syllable is (almost always) the syllable that is marked for word stress in the word's phonological representation. CALLED and NECK are monosyllabic and so their phonological representations are not specified for word stress. However, in these utterances, they are realized as the strong syllable of their Foot.

Although it is most often the lexical words in an utterance that appear in rhythmically strong positions, there are plenty of occasions when function words occur in a strong position too. In such cases, the function word often has a contrastive implication, in which case it is likely to be realized with a Supertonic. An example, already discussed, is "this" in (1.9.3), reproduced from Extract (1.9.2) in Activity 1.2.

(1.9.3) ||↑ 'whats ↑`this bit 'called though ||

In Activity 1.3, we considered different realizations of the sentence THE TELEPHONE IS LIKE A BANANA. The uncontracted form of *is*, namely [ɪz], is most likely to be found when *is* forms the Tonic, as shown in (1.8.5), for instance, when the speaker is correcting the previous speaker who had denied the resemblance between telephone and banana:

(1.8.5) || the 'telephone ↑`is like a ba'nana ||

Intonation and meaning

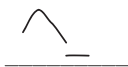
The aim of these sections has been to present a simple and usable notation system for English intonation, as spoken by mature speakers. The exposition has covered intonation structure, through its relation to the rhythmic structure of utterances, which includes links with the abstract patterns of word stress. From this presentation, it may be apparent that the main role of the prosodic features of speech (notably pitch, loudness and length) is to integrate two types of meaning.

First, an utterance is made up of words, each of which has a stress pattern that is a major cue to its accurate recognition by listeners. This abstract pattern therefore has to be realized in the utterance in such a way as to make it identifiable. Clearly, one aspect of phonological learning for young children is to learn the correct stress pattern and how to realize it. The importance and salience of this aspect of a word's phonology are implicit in Extract (1.5), the exchange between Robin and his mother studied in Activity 1.2 and reproduced here:

Robin: circle

(1.5)

1 M: ||↑ 'whats ↑`this bit 'called though ||



2 R: ɔi:j akəl^v

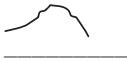
{ f }

3 M: || it ˘isnt || its 'called a ↑`funnel ||



4 R: f a f a
{ f }

5 M: || thats ˘right || n ↑'what comes 'out of the `funnel ||



6 R: ?ə mɛ ʊ kʰ

In line 3, Robin's mother explicitly corrects his attempt at labelling the item in line 2. In line 4, Robin duly has another attempt. This is accepted by his mother in line 5, even though it is clearly still quite a long way from her own pronunciation of FUNNEL. It is interesting to note the features that his and her pronunciations share: they both have two syllables, the same word-initial consonant, and in both, it is the first syllable that is prominent in terms of pitch and loudness. By contrast, if we compare Robin's word in line 2 with his mother's pronunciation of FUNNEL, we can note that Robin's word has three syllables, a different initial consonant, and prominence on the second syllable. On the other hand, the final syllable is very similar to the final syllable of FUNNEL as produced by his mother – a similarity that is lost when he produces his revised attempt in line 4. This example suggests that not all phonological features are equal, when speakers are concerned to ensure the recognizability of a word, and that conveying the word's stress pattern may be particularly important.

The second type of meaning conveyed by prosodic features is meaning that is associated with intonation. Some aspects of intonational meaning, such as the ability to mark the end of the turn and to highlight important parts of the utterance, have been briefly touched on in this chapter. The most important task for the young child with regard to intonation is to learn to understand and convey its meanings. The next three chapters therefore address the three main types of meaning that English intonation serves: to organize the orderly exchange of speaking turns (Chapter 2); to highlight what is important or new in a turn that the listener should attend to (Chapter 3); and to initiate a new action in the talk, or to display alignment with an action that is underway (Chapter 4). These three areas of meaning are particularly relevant to carer–child interaction where, as we have seen, the meanings that young children try to convey through words are often obscured as a result of limited linguistic knowledge and ability. For this reason, we will continue to draw on the interactions between Robin and his mother when explicating these three areas of meaning and the role of intonation in communicating them.

Key to Activity 1.1

A transcript of this extract of overlapping talk is presented as Extract (1.1). Start and end of overlap are marked with []. Silences are indicated in tenths of a second, that is, (0.8) signifies 800 milliseconds:

(1.1)
 Daisy used to ha[ng out in Stevenage when I was younger]
 Beth [that'd be good we should do that o]ver
 (0.8)
 Beth [like summer]
 Daisy [summer](.)yeah

Key to Activity 1.2

- 1 There is a Supertonic on “this”, which suggests that they have just been talking about a different bit of the train piece, and the mother is now highlighting a new bit. This function of Supertonic placement is discussed further in Chapter 3.
- 2 The IP boundary occurs after the end of one clause, “it isn’t”, and before the start of the next clause. IP boundaries most often occur at the end of a major grammatical unit such as a clause. Research has shown that while there is no deterministic relation between IP boundary and the boundaries of grammatical units, IPs tend not to be greater than a sentence and most often coincide with a clause (Crystal, 1969).
- 3 The Tone in the first IP is a fall-rise, a Tone that often projects that the speaker is going to continue talking, as here. The Tone in the second IP is a fall, which often projects a TRP, that is, the end of the current speaker’s turn, as here. The relationship between IP boundaries and turn construction is considered in Chapter 2.
- 4 There are two syntactic questions, in line 1 and line 5. The Tone is a fall on each occasion. This does not support the notion that in English, questions are routinely spoken with a rising Tone.
- 5 Robin seems to display ability to mark the end of his turn, by a (rise-)fall pitch movement (lines 2, 4). For this reason he can be credited with the ability to produce an IP. As noted under (3) above, falling Tones (including rise-falls) in Southern British English often mark a TRP. Robin’s mother displays an orientation to Robin’s IP boundary and falling Tones as marking a TRP, i.e. the end of his turn, by starting her own turn (lines 3, 5).

Key to Activity 1.3

- 1 It is likely that one of your first changes was to use the contracted form of *is*, i.e. [z] rather than [ɪz]. In most positions other than the very beginning or end of the IP, contraction of this copula/auxiliary verb form is usual, even at a relatively slow speech rate. Here this would create a coda cluster [nz]; though in connected speech

it is very likely that the alveolar nasal itself will no longer be articulated, being marked just by nasality on the preceding vowel: [fəʊz]

- 2 Many grammatical items (often known as function words) consist of a single weak syllable. This includes determiners, auxiliary verbs and pronouns. In this sentence, at a fast speech rate the definite article *THE* is likely to lose its vowel, leaving perhaps just a weakly articulated and voiceless dental fricative before the voiceless plosive at the beginning of *TELEPHONE*: [θ^h]. The indefinite article *A* is quite likely to be devoiced too, and even to disappear almost entirely, perhaps just audible as a brief aspiration, sandwiched between the voiceless plosive at the end of *LIKE* and the plosive at the start of *BANANA*.
- 3 The first vowel of *BANANA* is likely to disappear at a fast speech rate, a common fate for an initial unstressed syllable that immediately precedes the stressed syllable in the word. It is well known that syllables in this position are particularly vulnerable in the speech of young children: ¹NANA is a very common child pronunciation of the word.
- 4 Similarly, the middle, unstressed vowel of *TELEPHONE*, sandwiched between the stressed syllable and a relatively prominent final syllable, is likely to be lost at a faster rate.

One possible outcome of the increase in rate is therefore something like:

(1.8.1)

TH ¹TEL PHONES LIKE B ¹NANA
 [θ^ht^hɛl^hɪ^hʃfəʊzlaɪkə^hˈpnanə]

You can listen to a pronunciation of this sentence at slow and fast rates.

(1.8.2)

|| [ðə^ht^hɛl^hɪ^hʃfəʊnzlaɪkəbə^hˈnanə] ||

(1.8.3)

|| [θ^ht^hɛl^hɪ^hʃfəʊzlaɪkə^hˈpnanə] ||

In the slower pronunciation, there are two stressed syllables, the second forming the Tonic, and seven unstressed syllables. In the faster pronunciation, there are again two stressed syllables, the second forming again the Tonic, but just five unstressed syllables. Because the stressed syllables are the same in both versions, both retain the same basic intonation structure:

(1.8.4)

|| THE ¹TELEPHONE IS LIKE A BA ¹NANA ||