

1

Preliminaries

1.1 Prosodic features

Phonetics, in the mind of the ‘man in the street’, nurtured on *Pygmalion* and *My Fair Lady*, generally consists of sounds and the transcription of sounds: he thinks, for example, of the word *nice* being transcribed as /nais/. Such a transcription might be made for various purposes including, for instance, showing the varying relationships between sound and spelling, or indicating how to pronounce a particular word in a language or dialect. This sort of transcription is usually limited to sounds (which are represented as discrete) that follow one another in a fixed order: in the case of *nice* an /n/ is followed by an /aɪ/ which in turn is followed by an /s/. Such sounds are usually referred to as segments and the sort of transcription that represents them is consequently referred to as a segmental transcription. But there are clearly other features involved in the way a word is said which are not indicated in a segmental transcription. The word *nice* might be said softly or loudly; it might be said with a pitch pattern which starts high and ends low, or with one which begins low and ends high; it might be said with a voice quality which is especially creaky or especially breathy. Such features generally extend over stretches of utterances longer than just one sound and are hence often referred to as suprasegmentals (and a type of transcription which indicates how any of them are used is therefore called a suprasegmental transcription). Alternatively, the shorter term PROSODIC is sometimes used and I shall generally prefer this term in this book. Prosodic features may extend over varying domains: sometimes over relatively short stretches of utterances, like one syllable or one morpheme or one word (the tones of tone languages are generally relatable to such shorter domains); sometimes over relatively longer stretches of utterances, like one phrase, or one clause, or one sentence (intonation is generally relatable to such longer domains). Of course this distinction is not always as clear-cut as it first appears: a sentence, for example, may consist of one word. Since this book is principally about intonation, I shall for the most part be concerned with features relating to the longer domains.

Preliminaries

1.1.1 **Length, loudness, and pitch**

The prosody of connected speech may be analysed and described in terms of the variation of a large number of prosodic features. There are, however, three features which are most consistently used for linguistic purposes, either singly or jointly. These three features are pitch, length, and loudness. Pitch concerns the varying height of the pitch of the voice over one syllable or over a number of successive syllables; length concerns the relative durations of a number of successive syllables or the duration of a given syllable in one environment relative to the duration of the same syllable in another environment; loudness concerns changes of loudness within one syllable or the relative loudness of a number of successive syllables. The terms pitch, length, and loudness refer to features perceived by listeners; before we go on to consider the linguistic functions of these features, we must spend a little time considering the physiological and acoustic correlates of these perceived features.

1.1.1.1 **Length.** LENGTH is in one way the simplest of the features: it makes little difference whether we view it as the length of time a speaker decides to continue to produce a linguistic unit, as the duration of the acoustic correlates of the unit on a spectrogram, or as the length of time during which a listener hears that unit. But in other ways it is the most complex feature. If, for example, we wish to measure the duration of particular syllables in order to judge whether varying degrees of accent involve varying degrees of lengthening, we will initially have to make some decisions about syllable boundaries which are to some extent arbitrary. Where are we to place the boundary between the two syllables of *extra*? While the solutions /éks-trə/ or /ék-strə/ may seem the most likely, the decision between these two solutions is not easy. Again, where are we to place the boundaries of the accented syllable in *potato*? If we decide that the first /t/ belongs with the accented syllable but the second does not, do we then include the compression stage of the first /t/ in the duration of the accented syllable? The relevance of length as a prosodic feature is also difficult to assess because there are often many different influences on the absolute duration of a segment or syllable. If we wish to show that accented syllables are longer than unaccented syllables, we have firstly to discount such influences as the 'innate' length of vowels (e.g. the vowel of *peat* is generally longer than that of *pit*) and the fact that the last syllable before a pause is often lengthened.

1.1.1.2 **Loudness.** LOUDNESS as perceived by the listener is related to the breath-force which a speaker uses. A famous theory concerning the phonetic basis of syllables once asserted that speech is divided into syllables by the ebb and flow of increasing and decreasing breath-force on vowels and consonants.

1.1 Prosodic features

Later, more sophisticated equipment (the earlier equipment was balloons in stomachs, the later was electrodes in muscles) showed that increases in breath-force were only regularly present in the case of accented syllables; and even this has been called in question by experiments showing that accented syllables are more regularly indicated by length and pitch than by loudness. The acoustic correlate of loudness is intensity or the amount of energy which is present in a sound or sequence of sounds, variations in intensity being produced by variations in the pressure of air coming from the lungs. The relevance of intensity or loudness as a prosodic feature, like that of length, is often difficult to assess because there are often different influences on the absolute intensity or loudness of a syllable or sequence of syllables. For example, open vowels are acoustically of greater intensity than close vowels and listeners must in some way allow for this when interpreting relative loudness for other purposes. Again, the relationship of absolute intensity to perceived loudness is by no means linear (a sound has to be much more than doubled in absolute intensity before it will be heard as twice as loud) and moreover the relationship is different at different frequencies. Additionally, loudness may be used for a variety of linguistic purposes, some of which apply to single syllables, and some to sequences of syllables. I may shout because I am angry or I may make my accented syllables much louder than my unaccented syllables as an emphatic device.

1.1.1.3 **Pitch.** PITCH is the prosodic feature most centrally involved in intonation and it is with this feature that I shall be principally concerned in this book. Physiologically, pitch is primarily dependent on the rate of vibration of the vocal cords within the larynx (it is nowadays more usual to use the spelling 'cords' although the spelling 'chords' has often been used in the past). How such variation in the rate of vibration is brought about has been a matter of some dispute: at the moment the majority opinion is that such variation is principally produced by the length and tension of the vocal cords, which factors are themselves controlled by the intrinsic (and possibly the extrinsic) muscles of the larynx. Pressure of air below the larynx is regarded as a secondary influence on the rate of vibration.

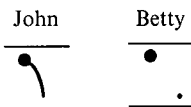
Rate of vibration of the vocal cords is reflected in the acoustic measurement of fundamental frequency. This term refers to the number of repetitions of the regular waveform within one second, such a regular waveform being typically produced when the vocal cords vibrate for voicing. So the number of times that the vocal cords completely open and close in one second is directly related to the frequency of the waveform. Fundamental frequency among male speakers varies between 60 Hz and 240 Hz and among female speakers between 180 Hz and 400 Hz. The average fundamental frequency for men is approximately 120 Hz, for women 220 Hz, and for children 265 Hz.

Preliminaries

While fundamental frequency involves acoustic measurement measured in Hz (or the number of cycles of vibration in one second), pitch is used as a perceptual term, relating to listeners' judgements as to whether a sound is 'high' or 'low', whether one sound is 'higher' or 'lower' than another and by how much, and whether the voice is going 'up' or 'down'. Such judgements are not linearly related to fundamental frequency. For listeners to judge that one tone is twice as high as another, the frequency difference between the two tones is much larger at higher absolute frequencies, e.g. 1000 Hz is judged to be double 400 Hz, but 4000 Hz is judged to be double 1000 Hz. This is also demonstrable from musical scales where, for example, a difference of a semitone at the bottom end of a piano keyboard is much less in terms of Hz than a similar difference at the top end. Fortunately, fundamental frequency values in speech are all relatively low (i.e. usually less than 500 Hz), and for most practical purposes pitch can be equated with fundamental frequency.

There are certain local characteristics of fundamental frequency which a listener must allow for if he is listening to an utterance and extracting meaning from its pitch pattern. Firstly, only voiced sounds have a repetitive waveform and hence a fundamental frequency. Around a quarter of the sounds in a connected English text are voiceless consonants and hence have no fundamental frequency; an ear listening for an overall pitch pattern learns to ignore these gaps in voicing. Whether or not consonants are voiced affects the fundamental frequency of adjacent voiced sounds: in particular vowels have a higher fundamental frequency when preceded by voiceless consonants than when preceded by voiced consonants. Moreover, the fundamental frequency peak will be at the beginning of the vowel following voiceless consonants but in the middle of the vowel following voiced consonants. It is also true that different types of vowels have inherently higher and lower fundamental frequencies: all other things being equal, open vowels will tend to have a lower fundamental frequency than close vowels. All such characteristics of fundamental frequency are in some way allowed for and discounted when listeners are listening to the semantics of an overall pitch pattern. However, they can complicate the extraction of such an overall pattern from an instrumental acoustic record of an utterance.

A slightly different sort of influence that the composition of individual segments and syllables may have on an overall pitch pattern is illustrated by the following words said with the same 'tone':



1.2 Auditory and instrumental

This tone is sometimes called a high-falling tone and semantically it represents a straightforward unadorned statement of someone's name. But notice that the tone is in fact 'realised' differently in each case. *John* consists of one syllable, which is voiced throughout: the fall is more or less continuous throughout the word. Whereas *Betty* consists of two syllables including two short vowels interrupted by a voiceless consonant; in this case the fall is realised as a step between a level high pitch and a level low pitch. As we shall see in chapter 3, some people actually consider the 'essence' of the pattern to be a sequence of high and low tones, rather than considering it a fall. This is a dispute between levels and contours as primes which has been argued for many years, has not been resolved in a principled way, and indeed may ultimately turn out to be a trivial question (see chapter 3, sections 3.3 and 3.9). For the moment, however, all the reader need know is that what is essentially the same pattern will actually look rather different depending on the number and make-up of the syllables which go with it.

1.1.1.4 **Summary.** Sub-sections 1.1.1.1 to 1.1.1.3 have considered the prosodic features of length, loudness, and pitch. In each case we looked at the way in which speakers physically controlled the feature, at the acoustics of the feature, and at listeners' perception of the feature. We noted that there is no one-to-one relation between acoustics and perception. In particular, different types of segment directly influence the acoustics of a prosodic feature and such segmental effects have to be discounted by listeners interpreting the meaning of a longer prosodic pattern, e.g. the different effects of close and open vowels on fundamental frequency have to be discounted when listening for the semantics of a longer pitch pattern. Moreover, some prosodic features may be used for two linguistic effects simultaneously, e.g. some syllables may be loud because accented syllables are being made louder than unaccented syllables to give an emphatic effect, while at the same time a whole stretch of syllables is louder than usual because the speaker is angry. It was also mentioned that of the three prosodic features considered, our chief concern will be with pitch, this being the principal perceptual correlate of intonation.

1.2 Auditory and instrumental

The previous section discussed the disparity between acoustic records and perception. This disparity has been reflected in a duality of approaches to the study of prosodic features, in particular to the study of pitch patterns. The most common labels attached to the two approaches are the AUDITORY and the INSTRUMENTAL. There has been a long history of auditory analyses of pitch patterns and their meanings. Very detailed studies have been made of the meanings involved in English intonation, using auditory methods on

Preliminaries

large bodies of data. Critics of such auditory analyses often claim that the methods are unscientific, that they are too impressionistic, that even those trained to listen to pitch patterns in language will hear only what they have been trained to hear, and that in this way myths are propagated. It is said that such impressionistic listening will be even less reliable in listening to a foreign language, which will be filtered through the listening habits developed in a mother-tongue. Such criticisms are of course made by those who prefer a wholly instrumental approach to prosodic features. Instrumentalists claim that their methods, unlike those of auditory analysts, involve precise and verifiable measurements and are consequently more scientifically respectable. The response of auditory analysts is that such precise instrumental measurements can only by their very nature be carried out on small amounts of data. It is also pointed out that such methods usually involve a very crude approach to meaning: typically a speaker may be asked to read a sentence with a 'statement intonation' and a 'question intonation'. Moreover, it is said that we know too little about perception to rely entirely on acoustic records.

In this book I shall try to use the best of both approaches. In explanation, an analogy with segmental description may be helpful. Phonologists operate with consonants like English /p/, /t/, /k/ and vowels like English /i/, /a/, /u/. If we look at acoustic records, we find very complex, inconsistent, and sometimes even intangible correlates of such sounds. Not only would we have difficulty in isolating such sounds if we relied entirely on acoustic records, but even with the benefit of many years of acoustic analysis, we cannot always tell what 'sound' has been uttered just by looking at the acoustic records. The essential point is that all linguistic units are in varying degrees abstractions and are perceptual rather than acoustic realities; because we cannot find invariant acoustic counterparts, we should not, on such grounds alone, discard such units. On the other hand, acoustic research has made considerable contributions to our understanding of sounds. One of the most famous contributions involved the acoustic nature of [p], [t], and [k]: the place of articulation of a plosive is principally indicated by the transition phase between the plosive and an adjacent vowel. There is no reason to regard the analysis of prosodic patterns any differently from the analysis of segmental patterns; both auditory and instrumental analysis have something to offer and should, ideally, be combined.

1.3 Prominence

The physical bases of three prosodic features have so far been considered. There are a number of other prosodic features of speech which have not been discussed, in particular TEMPO and PAUSE. I shall consider pause in some detail in chapter 3, sub-section 3.2.1 and tempo more briefly in chapter 6,

1.4 Intonation

sub-section 6.1.1. The reasons for the initial emphasis on length, loudness and pitch are twofold. Firstly, the relationship between their measurable attributes and their linguistic function is often complex; some examples of this complexity have already been given. Secondly, these three features conspire in varying degrees in many languages to give some syllables PROMINENCE when compared with other syllables. Such prominence (variously called stress or accent by different authors – I define my own use of these terms more precisely in chapter 2) is on one level a feature of words as stored in our mental lexicon (word-stress or word-accent) and on another level a feature of connected speech (sentence-stress or sentence-accent). Such prominences are often themselves linguistically important: they may be involved in distinguishing different lexical meanings, cf. *belów* and *billow*, or different grammatical classes, cf. *insult* and *insult*; or they may be involved in making certain syllables stand out in sentences, and hence make the word containing those syllables stand out as more important: cf. *John didn't dó it* and *Jóhn didn't do it*. Not only are the prominences produced by some combination of length, loudness, and pitch themselves linguistically important, they are also important because sequences of prominent and non-prominent syllables form the framework of connected speech. In many languages such patterns of prominent and non-prominent syllables produce a particular rhythmical effect. Additionally, and most important of all for this book, such patterns are the backbone of intonation. Intonation concerns which syllables are prominent, how they are made prominent, and to what extent they are made prominent; it also concerns how the movement from one prominent syllable to the next is accomplished. In chapter 2 I shall deal with matters of accent and rhythm, since it is impossible to describe and discuss intonation without first establishing a descriptive framework for accent and rhythm. In chapters 3, 4, and 5 I deal with intonation 'proper'.

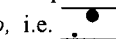
1.4 Intonation

INTONATION involves the occurrence of recurring pitch patterns, each of which is used with a set of relatively consistent meanings, either on single words or on groups of words of varying length. Grammatical constituents of any level up to at least the sentence may be treated as separate intonation-groups having their own meaningful tune (and indeed some features of intonation may even link such groups together into 'paratones', a word meant to indicate an analogy with paragraphs in the written language). For example, a common tone in English is the fall-rise, and one of its common meanings involves a contrast within a limited set of items either stated explicitly or, more usually, just implied. This tone is used on all the following examples:

Preliminaries

- (a) (Isn't his name Jim?) No / [∨]John
 (b) The old man didn't come / whereas the [∨]young man / did come and actually enjoyed himself
 (c) [∨]John didn't do it

(/ indicates an intonation-group boundary, although boundaries are not marked at the beginnings and ends of examples, nor where a change of speaker is involved; they are automatically present in such cases. The [∨] mark indicates a fall-rise tone spread over all syllables before the next boundary.) In example (a) the fall-rise occurs just on the single-word response *John*; in example (b) on the two words *young man*; and in example (c) on the whole sentence. Intonation-groups (which are sometimes also called intonational phrases) generally correspond with constituents of sentences in a somewhat loose way. For example, it is not uncommon for the noun-phrase subject of a clause to be given a separate intonation-group in English as in (b) above, but notice that the conjunction *whereas* has been incorporated into the group. Notice also that the fall-rise tone does not begin until the word *young*, indicating the focal point of the group. Such matters will be discussed in detail in chapters 3 and 4. The examples and explanation given so far are simply to illustrate the domain of intonation.

Intonation may convey several types of meaning. It may indicate a discursive meaning like inviting a listener to make a contribution to the conversation, or an attitudinal meaning like being condescending. If I were to say to a colleague *I'll show you how to do it* (the ^l mark indicates a high level pitch beginning on *show* and the / mark means a rise starting low on *do*, i.e. , that colleague might well feel that I was being condescending, such a tune being frequently used to children. In some languages (not English) the meanings associated with intonation may come nearer to being grammatical: this is the case where the use of particular tunes is closely tied to functional sentence-types, e.g. where statement, yes/no question, and command regularly involve certain tunes. From most of the descriptions of intonation in languages other than English, one might imagine that this was the principal use of different tunes in languages. It may indeed be true that many languages do use intonation less for attitudinal purposes than English, but the suspicion exists that the correlation of tunes with sentence types is merely an easy way to investigate intonation and often more sophisticated attitudinal and discursive uses remain undocumented.

1.5 Tone languages

Many languages use pitch to produce distinctive word-meanings. Such languages are called tone languages. Tone in tone languages is a feature of the lexicon, being described in terms of prescribed pitches for syllables or

1.5 Tone languages

sequences of pitches for morphemes or words; whereas intonation is a feature of phrases or sentences (but, as we shall see in chapter 3, some linguists now speak of an 'intonational lexicon' which stores overall contours or sequences of pitches with their own meanings, which are at some point mapped onto phrases or sentences). The tones of tone languages, then, concern the pitch patterns of words. In the simpler type of case, a change of meaning is produced if one tone is exchanged for another on one syllable, while keeping the segmental composition unchanged. This situation results in sets of words distinguished only by tone and applies to many languages of the Far East. One variety of Chinese, Szechuanese, has four tones, producing four different words when combined with the segmental sequence [ta]:

[ta] + [']: 'imitation of trumpet noise'	(Tone 1)
[ta] + [\]: 'to answer'	(Tone 2)
[ta] + [\]: 'to beat'	(Tone 3)
[ta] + [v]: 'big'	(Tone 4)

In the more complex cases involving the use of tone, words have prescribed tonal patterns, although minimal pairs are not always easy to find. For example, in Ganda, verbs fall into two tonal classes. Verbs in Class I have a high pitch on all syllables of the stem, e.g. *ku-seka* [. • •] 'to laugh'; verbs in Class II have a falling pitch, generally on the first syllable of the stem, e.g. *ku-tambula* [• ˩ ..] 'to walk'. This type of use of tone is often called 'characteristic tone'. It often involves a complicated build-up of the overall pitch pattern of a word by the use of affixes which may not only have their own inherent tone but also produce changes of tone in the stem or in other affixes. Because the changes of meaning brought about by these affixes often involve the sorts of modifications signalled by inflectional morphology alone in other languages (e.g. case in nouns and tense in verbs), it is often said that such a use of pitch involves a grammatical function of tone. Indeed in some languages (e.g. Efik, Igbo) modifications of meaning may be produced by change of tone alone, without the use of affixes. As implied by the languages mentioned, grammatical use of 'characteristic tone' is typical of many languages of Africa. But whether we are thinking of 'lexical tone' as in the case of Szechuanese, or of 'characteristic tone' as in the case of Ganda, such uses of pitch apply at the word level and produce changes of meaning quite unlike those of intonation.

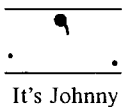
Tone and intonation are not completely mutually exclusive in languages. Languages with tonal contrasts may nevertheless make use of a limited amount of superimposed intonation. Such superimposed intonation may be manifested in four different ways: (i) the pitch level of the whole utterance may be raised or lowered; (ii) there will usually be downdrift in the absolute value of tones but

Preliminaries

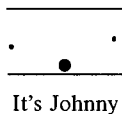
downdrift may be suspended; (iii) the range of pitch used may be narrower or wider; (iv) the final tone of the utterance may be modified in various ways. Some of these are present in the Szechuanese variety of Chinese mentioned above: emphatic statements are produced with a higher pitch; various attitudes are expressed by a wider pitch range; and questions involve utterance-final tonal variants as follows: for Tone 2, low-fall becomes low-level; for Tone 3, high-fall becomes high-level; for Tone 4, low fall-rise becomes simple low-rise.

1.6 Pitch accent languages

While the distinction between tone and intonation is a relatively clear one, the distinction between so-called ‘stress accent’ languages and so-called ‘pitch accent’ languages has never been very clearly defined. The term ‘stress accent’ is usually used to refer to languages, like English, using pitch primarily for intonational purposes. It is, however, an unfortunate term since it implies that prominent syllables in such languages are produced primarily by ‘stress’ which in this usage seems to mean breath-force or loudness. It has never been clearly shown that any language produces prominences primarily in this way and it seems certain that pitch is in some way used for accentual purposes at least in all languages which are not tone languages. What does however seem to be characteristic of ‘stress accent’ languages is that, although prominences or accents will commonly involve pitch, the actual type of prominence involved will be determined by intonational factors. In a sentence like *It’s Johnny*, the first syllable of *Johnny* is likely to be accented and hence given some sort of pitch prominence, but the type of pitch prominence may, for example, be high, as in:



or low, as in:



So-called ‘stress accent’ languages like English are generally contrasted with ‘pitch accent’ languages, of which the most well-known example is Japanese. In the standard Japanese of Tokyo words fall into two classes: accented and unaccented. Words with an accent realise the accent by a high pitch on the accented syllable followed by a low pitch on the following syllable, e.g. *óngaku* [● . .] ‘music’; *toshókan* [. ● .] ‘library’; *tamanégi* [. ● .] ‘onion’. The pitch of unaccented syllables is predictable by rule: an initial unaccented syllable is low, any