Intonation

Francis Nolan

1   Introduction

The term intonation refers to a means for conveying information in speech which is independent of the words and their sounds. Central to intonation is the modulation of pitch, and intonation is often thought of as the use of pitch over the domain of the utterance. However, the patterning of pitch in speech is so closely bound to patterns of timing and loudness, and sometimes voice quality, that we cannot consider pitch in isolation from these other dimensions. The interaction of intonation and stress — the patterns of relative prominence which characterise an utterance — is particularly close in many languages, including English. For those who prefer to reserve 'intonation' for pitch effects in speech, the word 'prosody' is convenient as a more general term to include patterns of pitch, timing, loudness, and (sometimes) voice quality. In this Chapter, however, intonation will be used to refer to the collaboration of all these dimensions, and, where necessary, the term 'melody' will be used to refer specifically to the pitch-based component.

Intonation is used to carry a variety of different kinds of information. It signals grammatical structure, though not in a one-to-one way; whilst the end of a complete intonation pattern will normally coincide with the end of a grammatical structure such as a sentence or clause, even quite major grammatical boundaries may lack intonational marking, particularly if the speech is fast. Intonation can reflect the information structure of an utterance, highlighting constituents of importance. Intonation can indicate discourse function; for instance most people are aware that saying 'This is the Leeds train' with one intonation

constitutes a statement, but, with another, a question. Intonation can be used by a speaker to convey an attitude such as friendliness, enthusiasm, or hostility; and listeners can use intonation-related phenomena in the voice to make inferences about a speaker's state, including excitement, depression, and tiredness. Intonation can also, for instance, help to regulate turn-taking in conversation, since there are intonational mechanisms speakers can use to indicate that they have had their say, or, conversely, that they are in full flow and don't want to be interrupted.

Intonation is not the only linguistic device for which pitch is recruited by languages; many languages use pitch to distinguish words. In languages around the world as diverse as Thai, Hausa (Nigeria), and Mixtec (Mexico), words are distinguished not only by vowels and consonants but also by the use of one of a limited set of distinctive pitch patterns or heights on each syllable. Such languages are called tone languages. A number of other languages, such as Swedish and Japanese, make a more limited use of pitch to distinguish words. These languages might best be called lexical accent languages. All tone languages and lexical accent languages also have intonation, but in general the greater a language's use of pitch for distinguishing words, the less scope it has to develop an elaborate intonation system. English, on the other hand, is not a tone language or lexical accent language, and is generally agreed to have relatively complex intonation.

This Chapter is set out as follows. Section 2 gives an introduction to what intonation consists of, and how we can visualise it and analyse it phonologically. It also draws attention to the aspects of prosody which are characteristic of English. Section 3 gives some examples of the kinds of information which intonation can carry and the intonational forms which are used in English. Section 4 looks at the variation in intonation to be found in dialects[1] of English. Section 5 concludes the Chapter with some general observations. In no respect does this Chapter attempt to give a comprehensive account, which would be impossible within its

scope; rather it samples the phenomena of English intonation to provide an overview. Readers who want more comprehensive accounts, both of English intonation and intonational theory, can follow up references in the 'Further Reading' Section as well as specific references cited in the text.

The examples of intonation patterns given in the Chapter assume, unless otherwise stated, a variety of pronunciation which has sometimes been termed 'standard Southern British English' (SBE) – the prestige variety of the south east of England which also serves in varying degrees as a prestige norm elsewhere in the British Isles. However the patterns used for examples will be similar to patterns in General American, and so the examples should be accessible not only to the large number of speakers of those two varieties but also to the much larger population of English speakers who have passive knowledge of those pronunciations.

2.   What is intonation made of, and how can we represent it?

2.1   The acoustics of intonation

Figure 1 shows two acoustic analyses of the utterance 'But Melanie's never been *near* the manuscript', spoken as a sharp retort to someone who might have said for instance 'Melanie doesn't think the manuscript's genuine'. The top analysis is a spectrogram, showing how the resonances and other acoustic components of speech evolve and change over time. A phonetic transcription has been added to show roughly which parts of the signal correspond to which linguistic elements. The bottom analysis shows a plot of the fundamental frequency, the acoustic consequence of the rate at which the vocal cords are vibrating in voiced speech. The fundamental frequency contour is more or less what we hear as the changing pitch of the speech. The contour is not continuous because voiceless sounds inevitably interrupt it; and

furthermore whenever the vocal tract is obstructed the fundamental frequency is perturbed. However the general trend of the pitch is clear. The utterance starts mid-low on 'But', goes low on 'Mel(anie)', rises to a peak on 'near', and falls sharply and thereafter stays low and level. This of course is not the only way the sentence could be said, but it is one appropriate way given the context described above.

INSERT FIGURE 1 ABOUT HERE

Remember that intonational pitch works hand in hand with other prosodic dimensions, notably duration. It is clear from the spectrogram that the most prominent syllable in the utterance 'near' takes up a disproportionate time compared to other syllables. Other durational correlates of prominence are less straightforward, since they interact with segmental determinants of duration (e.g. phonological vowel length); but it can be seen for instance that the unstressed vowel of 'the' is shorter than the immediately following vowel, that of 'man(uscript)'. Note too that the trisyllable 'manuscript' is more than 50% longer than 'Melanie', also trisyllabic; this is partly as a result of the former's more complex syllable structure, but also because a lengthening of sounds (a *rallentando*) is found at the end of an intonation pattern.

2.2   *General characteristics of English prosody*

All languages have ways of making given linguistic elements stand out in the stream of speech, of making them 'prominent'. One or more syllables in a word may be stressed ('di<u>ver</u>sifi<u>ca</u>tion'); and some words in an utterance will be more prominent than others ('I <u>told</u> you to go <u>home</u>'). Languages differ, however, in what might be termed their 'prominence gradient', the steepness of change between prominent and non-prominent elements. At the syllabic level, English is characterised by a steep prominence gradient. Prominent syllables have full vowels, i.e. vowels which are not schwa or unstressed /ɪ/ (as in

the first and last syllables of 'decided' in those dialects where schwa is not used in this context), and have relatively long durations. Non-prominent syllables often have reduced vowels (most commonly schwa). By contrast in French, for example, the average gradient between a prominent and a non-prominent syllable is less steep; French unstressed vowels are generally not reduced, and stressed syllables are less salient.
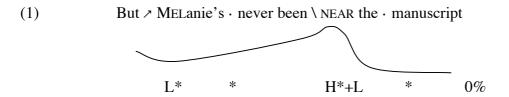
The fact that English is characterised by a steep prominence gradient is central to its intonation. One of the few things on which there has been a consensus among intonation analysts is that, put simply, interesting things happen to the pitch around prominent syllables; such syllables are associated with a pitch *landmark*. This is seen most clearly in Figure 1 in the case of the word 'near', which coincides with a high point, a *peak*, after which the pitch drops sharply over the whole range used in the utterance. 'Mel(anie)' coincides with a low point, a *trough*, after which the pitch climbs steadily to the peak. Could we look at prominence the other way round, and say 'these syllables are prominent *because* they are associated with pitch landmarks'? The factor which breaks the circularity is that the prominence pattern of a word is independent of pitch. A word's stress pattern, or metrical prominence pattern, is often predictable from its phonological and morphological structure; and it is also realised, mainly through timing relations, even when spoken without a pitch accent. The word 'manuscript' in Figure 1 has no pitch landmark associated with it (it's low and level), but it is still apparent from the rhythm that the syllable *man-* is the stressed syllable (we will return in Section 3 to why this word should be accentually neglected in this way). In fact, if we were to resynthesise the utterance on a monotone, the prominence relations would still be completely clear. In describing English intonation, the 'association' of a pitch landmark with a particular stressed syllable is crucial; it is termed a *pitch accent* (or often just *accent*). The melody of an utterance consists to a large extent of the sequence of

its pitch accents, and the description and classification of these landmarks forms a central part of current models of intonation.

English, then, is a language in which there is a relatively sharp difference between prosodically prominent events and those which lack prosodic prominence. The melodic part of intonation involves tonal events associated with points of prosodic prominence, and additionally with boundaries of intonational phrases.

2.3   The phonology of intonation

Using the term 'phonology' with respect to intonation implies that there are discrete, contrastive linguistic units[2] underlying the continuously variable melody of speech, and that these units do not have meaning (any more than a phoneme has a meaning), but can function in context, singly and in combination, to convey meaning. These implications are now widely accepted.

In (1) below there are two alternative phonological (or 'intonological') analyses of the intonation of the utterance in Figure 1, the melody of which is now represented as a stylised pitch curve[3]:

(1)   But ↗ MELanie's · never been \ NEAR the · manuscript

                    L*   *   H*+L   *   0%

Embedded in the sequence of words (in which small capitals indicate pitch accents) are symbols from an analysis within what has become known as the British tradition, developed in works such as Palmer (1922), Kingdon (1958), O'Connor and Arnold (1961/1973) and Crystal (1969). The intonational elements are shown by the diacritics before the stressed syllables of words (the symbols used vary from author to author, but the ones chosen here illustrate the general point). Before 'near', for instance, there is a sloping line which indicates

a fall. The fall is specifically the *nucleus*, that is, the accent which occurs last and often constitutes the most salient point of the utterance. The stressed syllable of 'Melanie' initiates a pre-nuclear rise, represented by the diagonal up-arrow. The elements of the system, then, are generally pitch movements; the exception in this example being the dots before 'nev(er)' and 'man(uscript)' which mark a stressed syllable within an existing pitch trend (here rising and low level respectively).

Below the stylised pitch curve is an equivalent 'autosegmental-metrical' (AM) analysis (for the term AM, see Ladd, 1996 pp. 2-4). AM descriptions take as their atoms the H (high) and L (low) tones of autosegmental phonology, originally applied to tone languages, combining these tones when needed into 'bitonal' (or potentially larger) elements. The Hs and Ls constitute pitch targets, and pitch movements arise from interpolating between ('joining up') these targets. The 'metrical' part of the name arises because, crucially, certain tones are tied to metrically prominent events in the utterance (in effect stressed syllables) as noted in 2.2 above; this is represented in the notation by adding an asterisk to the tone. Thus the syllable 'near' in the example is stressed and associated specifically with the high tone of the H*+L bitonal pitch accent. Metrically strong syllables without a pitch accent are not marked in most AM transcription systems, but logically could be shown as here by an asterisk.

The AM framework became the dominant paradigm in intonational research under the influence of Pierrehumbert (1980) and subsequent work, for instance Beckman and Pierrehumbert (1986) (for an introduction to AM and a critique see Ladd, 1996). A modified version of Pierrehumbert's (1980) description is expressed in the ToBI transcription system which was agreed on as a unified set of conventions for transcribing American English, particularly in work on speech corpora (see Silverman et al., 1992; Beckman, 1999); and a number of language-specific adaptations such as G-ToBI for German (Baumann, Grice and

Benzmüller, 2001) and ToDI for Dutch (Gussenhoven, Rietveld and Terken, 1999). The particular variant of the AM class of descriptions used here is the IViE system (the acronym standing for Intonational Variation in English) which was developed as part of a research project[4] into the intonation of a number of urban centres in the British Isles.

Superficially the British and AM analyses look very different, but there is a high degree of compatibility. Most of the intonational phenomena which can be expressed in one can be expressed in the other, and some of the differences between specific analyses in the two traditions are incidental. One essential difference, however, concerns the boundary of an intonation unit, or *intonational phrase* (IP) as it is now commonly known. An essential task in making an intonation analysis is to divide the speech into intonational phrases. These may be separated by pauses, but more often in fluent speech the end of an intonational phrase will be marked (if at all) only by a degree of slowing (pre-boundary lengthening), and the real essence of an intonational phrase is its internal coherence in terms of intonation pattern (rather in the way that we don't expect to find a gap between syllables, but rely for their demarcation on their internal coherence in terms of lawful combinations of sounds). AM models assume that an intonational phrase boundary may (or in most versions must) have a *boundary tone* associated with it. We can illustrate this if we imagine a reply to 'But Melanie's never been near the manuscript' consisting of an incredulous 'Never?!' with an overall falling-rising contour. A 'British' analysis would classify this as a fall-rise pitch accent. IViE would regard it as H*+L H%, with the final H% indicating a tone 'belonging' to the intonational phrase boundary. On the face of it these seem equivalent, but if we add more material to the response while keeping the pattern equivalent, and leaving the main stress on 'Never', we will find that the rising part of the fall-rise is delayed to the end:

(2)   NEver?!   She's NEVer seen the manuscript?!

H*+L H%   H*+L   *   H%

Phenomena like this suggest that intonational equivalence is captured more transparently through the use of boundary tones. However it is still useful to recognise the coherence of patterns such as H*+L H%, and the combination can be called a (nuclear) *tune*.

Although IViE acknowledges the importance of boundary tones, it allows IP-final boundaries to be tonally unspecified (0%) when there is no pitch movement in the immediate vicinity, unlike most AM models which require H% or L% to be specified. In doing this, it merely extends and makes explicit the practice in other models of not specifying tone on many IP-initial boundaries. Henceforth in this chapter examples will be presented and discussed in terms of the IViE transcription system, albeit a somewhat simplified version. For the full IViE inventory of pitch accents, boundary tones, and intonational processes see for instance Grabe, Nolan and Farrar (1998) and Grabe (2001).

2.4   *Non-phonological components of intonation*

Not all intonational effects lend themselves to analysis in terms of discrete categories such as pitch accents and boundary tones. Other intonational effects are communicative in the sense that the speaker has a choice, but are essentially gradient. For instance each of the following ways of saying an utterance conveys progressively greater involvement (whether or not this is the speaker's true feeling):

(3) I'd LOVE to meet him   I'd LOVE to meet him   I'd LOVE to meet him

but identifying three gradations (rather than four, or seven, or more) is arbitrary; *pitch range* here behaves as a continuum.

This non-categorical aspect of intonation probably provides a link to the origin of intonation in very basic physical and physiological phenomena. In order to explain certain universal tendencies in the use of pitch Gussenhoven (2002), building on earlier work by Ohala (e.g. 1983, 1984), proposes three *biological codes*: the *frequency code*, the *effort code*, and the *production code* (which I will rename the *respiratory code*). For instance, small objects or animals produce high frequencies, and so high pitch is a natural way to signal submissiveness in the animal kingdom, and by (metaphorical) extension politeness or uncertainty (among other things) in human interaction – the frequency code. Greater physical effort, resulting from physiological arousal, will produce more energetic movements, and more dramatic pitch change, and by extension can naturally signal involvement (as in (3)) or linguistic emphasis – the effort code. And as vocalisation proceeds, air is used up, subglottal pressure drops, and the natural tendency is for pitch to get lower in the course of a vocalisation, so it may be natural to signal newness by high pitch and older information by lower pitch – the respiratory code. Quite possibly the categories of intonational phonology represent in some measure the *grammaticalisation* of these codes; it is tempting to see the use of H% in (some) questions as arising from the frequency code. We shall see in section 3 that the task of intonational signalling in English is shared between a discrete, clearly phonological resource and a gradient component.

Relatively little attention has been paid to systematising the description of the non-categorical part of English intonation (though Crystal (1969) does discuss many relevant

phenomena). A useful prerequisite to understanding those aspects involving pitch range (best used as a 'catch all' term) is a clear set of terminology. We can distinguish the following: *speaking tessitura*, a given speaker's range of comfortable speaking pitch; *pitch level*, the overall placement of an utterance within a speaker's tessitura; *pitch span*, the general distance between highs and lows in an utterance; *pitch excursion*, a local high–low distance, e.g. associated with a pitch accent; and *downtrend*, the lowering of pitch over the course of an utterance. In these terms the degrees of involvement in (3) are signalled by changes in pitch span (manifested in the excursion of a single pitch accent, but if the utterance were longer the changes would affect the whole of the utterance).

3.   Functions and forms of English intonation

Section 2 introduced some of the general concepts required for understanding intonation. This section exemplifies how English intonation carries a number of different kinds of information.

3.1   Grammatical structure

An important role of intonation is as the 'punctuation' of spoken languages, marking the division between grammatical units and more generally helping the listener to follow the utterance. The function is brought sharply into focus on occasion when the words used allow more than one grammatical parse, for instance 'While eating my dog my cat and I watched television'. In writing we would use a comma; after 'dog' for the more unsavoury interpretation, and after 'eating' (and probably another comma separating 'my dog' and 'my cat') for the pleasanter interpretation. An intonational equivalent of this comma in these two
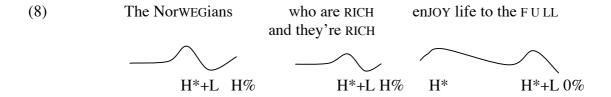
positions is transcribed in (4) and (5) respectively – a falling pitch accent followed, crucially, by a high boundary tone, along with a slowing down before the boundary:

(4)   While EATing my D O G   my CAT and I WATCHed TELevision

                          H*+L H%

(5)   While EAT i n g   my DOG my CAT and I WATCHed TELevision

                          H*+L H%

Note, however, that the relation between grammatical units and intonational units is not one-to-one. It is possible to phrase the following sentence intonationally in at least two ways:

(6)   If you're READy   we'll G O   If you're READy we'll G O

                H*+L H%   H*+L 0%   H*   H*+L 0%

without there being any corresponding change in grammatical structure. We might regard the change as a kind of 'connected speech process' like segmental assimilation, correlated with – but not directly determined by – speech rate. In general, then, we can regard grammatical structure as determining the point at which intonational phrase boundaries can occur, but whether they do or not depends on performance factors such as speech rate. The slower and more careful the speech, the more explicitly will grammatical structure be signalled in intonational phrasing.

        In some cases intonation can guide the listener to grammatical structure which is not directly to do with phrasing. For instance the intonation of the words 'The Norwegians who are rich enjoy life to the full' can signal whether the relative clause is restrictive, meaning that, specifically, rich Norwegians enjoy life to the full:

(7) The NorWEGians who are R I CH   enJOY life to the F U LL

H*   H*+L H%   H*   H*+L 0%

or whether the relative clause is non-restrictive, implying that all Norwegians are rich, and

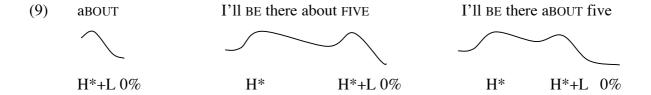having a status more like a parenthetical remark (e.g. '…and they're rich…'):

(8) The NorWEGians   who are RICH   enJOY life to the F U LL
                           and they're RICH

H*+L   H%   H*+L H%   H*   H*+L 0%

Whilst cases of intonational disambiguation such as the ones above are useful for illustrative

purposes, intonation provides guidance to the grammatical structure of all speech.

3.2   Information structure

Another thing which intonation does is to highlight points of high informational importance

in the utterance. Each word in the lexicon has a stressed syllable, or, perhaps better, a

'stressable' syllable. This means that this syllable has the potential to be the site of prosodic

prominence in an utterance. The prominence is usually manifested as greater duration,

greater intensity (the primary physical correlate of loudness), and in the majority of cases a

pitch accent. In the word 'about' it is the second syllable which is stressable. If we cite

'about' in isolation (9), the second syllable will carry a pitch accent — often H*+L. If

however we say the word as part of the utterance 'I'll be at the station about five', there will

be by default no prominence on the second syllable of 'about' beyond what may be perceived

as a result of the 'full vowel' (in this case a diphthong) it contains and its rhythmic context.

But if, again, the specified time is already present in the discourse, and the speaker wants to

focus on the approximation implied by the preposition 'about' ('no, don't buy tickets for the 5.02, it's too risky'), then that word can carry a pitch accent (rightmost example in (9)).

(9)   aBOUT   I'll BE there about FIVE   I'll BE there aBOUT five

       H*+L 0%   H*   H*+L 0%   H*   H*+L 0%

This exemplifies an important principle, that the speaker adjusts prominence according to communicational need. In the citation utterance there is no redundancy (i.e. predictability), and no word which is more important than 'about'. In the sentence uttered when the specified time is new information, 'five' is more important, and the presence of a temporal preposition is predictable from the rest of the sentence. It would be most unnatural to speak a sentence putting a pitch accent on every word, and as a first rule of thumb we can expect content words to have a pitch accent and grammatical words to lack one.

In fact at the same time as associating 'about' with a pitch accent the speaker has robbed 'five' of the prominence it had the first time round. This kind of adjustment of prominence is a crucial feature of English intonation, often called *deaccenting*[5]. By de-accenting 'five', focus has been placed on 'about'; and 'five', which is *given* information, is relegated to a lower level of salience. Deaccenting happens when a word is given by virtue of being repeated (10) or being substituted by a hypernym (11):

(10)   I OFFered her a COFFee but it TURNS out she doesn't DRINK coffee

(11)   I OFFered her a BEER but it TURNS out she doesn't DRINK alcohol

In such examples it is intonationally ill-formed in English[6] — and will give rise to a perceptual double-take on the part of the listener — if the given item carries prominence

equal to that of its first occurrence. In contrast many languages, such as Italian and Romanian (Ladd 1996, pp. 176-7), do not typically have deaccenting of given information. Absence of deaccenting in a language, however, does not necessarily mean that givenness goes unsignalled. In Icelandic, which does not deaccent given information (Nolan and Jónsdóttir, 2001), it seems that the information structure is reflected in gradient prominence levels, and deaccenting may just be a grammaticalisation of a very general reflex of the effort code.

So far the use of intonational pitch accents in English seems rather logical; informationally rich items are made to stand out and other information is backgrounded by deaccenting. But it has long been remarked that the relation between information and accent is not always so transparent, as in cases such as the following:

(12)   Look OUT! That CHAIR's broken

In the context of someone about to sit down, 'chair' is contextually given, and being broken is the unexpected, crucial information. Yet, perversely at first sight, 'chair' gets the main accent. But this kind of accentuation is probably the intonational equivalent of pointing; first make sure the listener looks at the chair, because then the problem will be perceived directly.
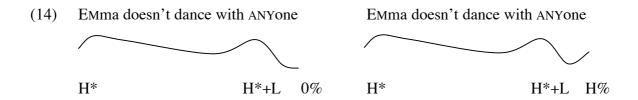
Also initially opaque is the kind of contrast between the following utterances:

(13)
The DEER was shot by JOHN   the BUTCHer   The DEER was shot by JOHN   the butcher

H*+L 0%   H*+L 0%   H*+L 0%   *   0%

In the first version 'butcher' is in apposition, and explains that John is the butcher. It's rather like a reduced non-restrictive relative clause. The pitch accent on the item in apposition ('butcher') usually echoes the pitch accent on the word to which it is in apposition ('John'), but with a less extensive pitch excursion. In the second version on the other hand 'butcher' is

an evaluative epithet, a metaphorical application of the word expressing (here) the speaker's disapproval of John's recreational pursuits. It carries a rhythmic stress, indicated here by the asterisk, but no pitch accent. This deaccenting is conventional, but not easy to explain. Conceivably it is a grammaticalised form of the reduced pitch span which often accompanies parenthetical expressions, including expressions of opinion, as in 'John — and I think he's a butcher because of it — is the one who shot the deer.'
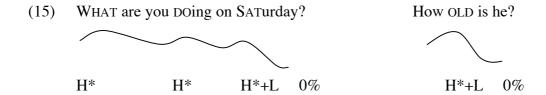
A comprehensive account of the relation of intonation to information structure would be too lengthy for the scope of this chapter, but as a final, very specific case, consider the following:

(14)   EMma doesn't dance with ANYone   EMma doesn't dance with ANYone

     H*   H*+L   0%   H*   H*+L   H%

In reply to 'why didn't she dance with Wayne?' the first version, with a low boundary tone, means that Emma will refuse all men who ask her to dance without exception. The second, with a high boundary tone, means that Emma is selective; she doesn't accept just *any* offer. The difference may arise from two broad categories of intonational meaning that have been associated with boundary tones. Low endings are thought of as assertive and non-continuative, for which Cruttenden (1997, p.163) has proposed the term *closed*, and high endings as non-assertive and continuative, or *open*. Thus the high boundary tone in the second version leaves it open for the speaker to express, or the listener to infer, a qualification, e.g. ' – but she'll say yes if the man looks rich'.
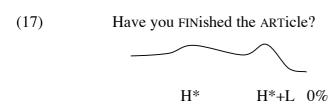
3.3   *Discourse function*

The best know fact about intonation is that questions rise. Like most well known facts it is a considerable oversimplification. Counterexamples are easy to find. English 'Wh-' questions in particular are more often falling at the end than rising:
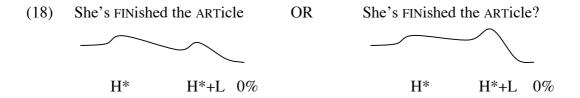
(15)   WHAT are you DOing on SATurday?   How OLD is he?

           H*   H*   H*+L   0%   H*+L   0%

Nonetheless the popular belief that the voice goes up in questions has some basis in truth. 'Yes-no' (or 'general') questions can rise:

(16)   Have you FINished the ARTicle?   OR   Have you FINished the ARTicle?

             H*   H*+L H%   H*   L*+H H%

These two versions are both common; the first has a falling rising pattern on the last word, and the second steps down to the final word but then rises to the end[7]. Ending high is in keeping with the *open—closed* distinction mentioned above, and Gussenhoven's *frequency code* and *respiratory code* (Section 2). The questioner perhaps metaphorically submits to the greater knowledge of the hearer, and leaves it open to the listener to provide completive information. However it is still perfectly well-formed to say:

(17)   Have you FINished the ARTicle?

                H*   H*+L 0%

One might nonetheless assume that if there is nothing in the words to indicate that an utterance is a question (a 'morphosyntactically unmarked question') then the phonological choice of a high boundary tone would be obligatory; nevertheless the second utterance below will be interpreted as a question:
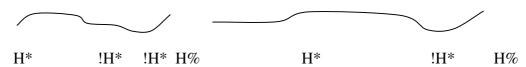

(18)   She's FINished the ARTicle   OR   She's FINished the ARTicle?

              H*   H*+L 0%   H*   H*+L 0%

The question is marked by gradient aspects of pitch range; the *downtrend* is less steep, and the *pitch excursion* of the nuclear accent is greater. In tone languages, where local pitch movements are determined lexically, intonation will rely heavily on such pitch range effects. In English there is a rich and to some extent complementary interplay in the signalling of discourse function between morphosyntactic marking, discrete intonational marking, and gradient intonational marking.

3.4   Attitude and the speaker's state

From the brief survey above concerning questions it can be seen very clearly that there is no one-to-one mapping between discourse function and intonation pattern. Some of the reason for this is that intonation is also doing other, less linguistic, work, conveying information for instance about the speaker's attitude. The example (17) of a question ending in a fall is unambiguously a question (because of the syntax), but a rather less genial, more demanding one than those in (16). Furthermore although we have tacitly assumed that statements are *closed* and are associated with low endings, not every statement ends low. Most famously, the spread through many varieties of English of the 'high rising terminal' (see e.g. Fletcher, Wales, Stirling and Mushin, 2002) – the trend to end intonational phrases

on a high and rising pitch – has made rising intonation on non-question utterances commonplace, as for instance in examples like the following (where !H* indicates a lowered or *downstepped* high accent):

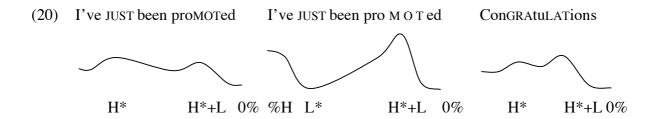(19)   MY name's JOHN SMITH.   I've got an apPOINTment with Dr SANDerson.

        H*   !H*   !H* H%   H*   !H*   H%

The speaker is not asking for information, but is more probably using a signal for non-assertiveness (the *frequency code*) as a politeness strategy.

There is no denying the role of intonation in conveying attitude, as witness both the common observation that the problem was 'not what he said but the way that he said it', and the large amount of attention devoted to the attitudinal function of intonation in books tutoring learners of English. However with attitude we are entering particularly difficult territory. Not only is someone's attitude hard to describe (much harder, say, than the linguistic description of an utterance as a declarative consisting of two clauses and functioning as a question), but also a person's attitude shades into their psychological state. Whilst choosing a 'polite' or 'informal' intonation is primarily a matter of attitude, a person whose intonation might be described as 'angry' may be genuinely experiencing that emotion and expressing it unchecked, may be trying with only partial success to hide it, or may be feigning anger to signal that the matter in hand is one which deserves condemnation. There is a large body of work on how speech is affected by actual emotions and psychological states (see e.g. Scherer, 2001), but these non-linguistic determinants lie outside the scope of this Chapter.

As we have seen in (16-17) and (19), categorical choices are available in English to convey attitude. But as we would expect from the link between attitude and psychological

state, the deliberate communication of attitude also employs devices which directly reflect Gussenhoven's (2002) *biological codes*. The gradations of *pitch span* in (3) on the words 'I'd love to meet him' directly mimic (or indeed are) the effects of physiological arousal, and convey progressively greater involvement. It is tempting to say 'greater enthusiasm', but we must beware of attributing specific meanings to intonational effects; if we impose a similar continuum of increasing pitch spans on the reply 'I rather *not* meet him', the strength of feeling is mapped in a similar way, but we can no longer label it enthusiasm.

One aspect of attitude is *accommodation*, the degree to which a speaker matches the speech of an interlocutor. Undoubtedly prosodic accommodation occurs widely. For instance if one person uses whispery phonation and a reduced pitch span, their interlocutor may well do the same. Failure to accommodate pitch span, for instance, can lead to ill-formed exchanges; if the intention of the third utterance in (20) is genuinely to congratulate, the response is appropriate to an utterance in the manner of the first, both in terms of phonological choices and pitch span, but not the second, against which it will sound somewhat grudging:
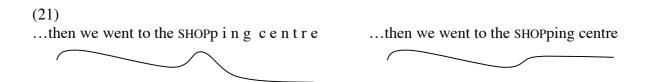
(20)   I've JUST been proMOTed   I've JUST been pro M O T ed   ConGRAtuLATions

　　　　H*　　　　H*+L 0% %H L*　　　　H*+L 0%　　　　H*　　　　H*+L 0%

3.5   *Discourse regulation*

In a successful conversation *turn-taking* by the speakers happens smoothly. Depending on the type and degree of formality of the interaction, interruptions may be appropriate, but they will be recognisable as such by the participants, as will the point at which a speaker has

finished what he or she has to say. The 'traffic signals' which regulate a well-formed interaction are mainly intonational.
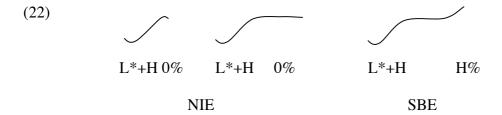
End-of-turn markers include low pitch, reduced loudness, and rallentando (lengthening of turn-final elements). The low ending and lengthening (indicated by the stretched spacing of the text) in the first utterance in (21) give it an air of finality. This does not preclude further comment on the topic (e.g. a question about it from the listener), but it does open the floor to another speaker. On the other hand the lack of slowing (or even accelerando) in the second utterance, combined with sustained final high pitch often used in listing items, indicates that more is to come and the speaker is not willing to yield the floor.

(21)
…then we went to the SHOPp i n g c e n t r e   …then we went to the SHOPping centre

Again we can relate this intonational use of pitch to Gussenhoven's (2002) biological codes. The *respiratory code* ('production code' in his terms) links low pitch and finality by virtue of the reduction in subglottal pressure as air is used up in speaking, and this link could be extended metaphorically to a conversational turn. Conversely attempts to wrest the floor from the speaker will be characterised by high pitch and loudness.

4.   Intonational variation

Varieties of languages are marked not only by their vowels and consonants but also by their prosody. The intonation of some varieties is often remarked on by outsiders using terms such as 'sing-songy' or 'flat'. One of the most distinctive dialects of English from the intonational point of view is Northern Irish English (NIE), which 'always goes up at the end'. The truth is a little more complex, as shown in (22).

(22)

L*+H 0%   L*+H   0%   L*+H   H%
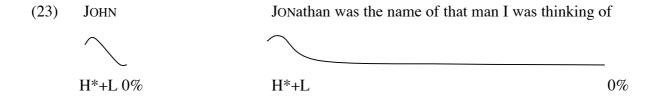
NIE   SBE

The first and second patterns show the commonest nuclear 'tune' of NIE. The first pattern shows what happens on short (usually monosyllabic) phonetic material, such as the answer 'three' to the question 'how many?'. It looks and sounds pretty much like a rise; but as soon as the phonetic material becomes longer (e.g. 'three of them') as in the second pattern, it becomes clear that the 'underlying' pattern is a 'rise-plateau'. This nuclear tune can be analysed within the IViE system as L*+H 0%. This is a pattern which seems not to occur in Southern British English (SBE), or most other dialects; as shown in the third schema a nuclear rise co-occurs in SBE only with a high boundary tone (and the tune functions as a question, not a statement).

Here we have what appears to be a phonological difference between dialects, specifically a difference – similar to a segmental phonotactic difference – determining the permissible combination of phonological elements or possible *tunes*. It is also possible in NIE to drop sharply at the end of the plateau to an L% boundary, again yielding a tune which is not available in SBE or most other dialects.

There are (at least) two other ways in which dialects can manifest a difference in their intonational phonology. First, dialects can differ in terms of what intonological elements they have in their inventory, just as a dialect may lack a phoneme (SBE does not have the voiceless labial-velar which distinguishes 'what' from 'watt', while Scottish does, for instance). The intonational inventory will, of course, depend on analytic assumptions; one could dispose of the segmental difference just mentioned by treating the voiceless labial-

velar as the combination of /h/ and /w/. Within the IViE framework, which assumes that an intonational phrase boundary tone T% will be manifested by pitch movement directly adjacent to the boundary, it seems that SBE lacks an L% boundary in its inventory. Nuclear falls are accounted for as H*+L, reflecting the fact that as material is added after the nuclear syllable, the low pitch is still attained shortly after the accented syllable and not at the boundary, as in (23). There are no cases where a fall can be associated unambiguously with the boundary and not with a prominent syllable, contrary to the NIE pattern discussed above.

(23) JOHN   JONathan was the name of that man I was thinking of

H*+L 0%   H*+L   0%

Second, the association of intonational elements with functions and meanings shows considerable variation between dialects. Grabe and Post (2002) examined read statements and inversion questions in the IViE corpus and found the distribution of nuclear tunes (last pitch accent and boundary tone) shown in Figure 2 for SBE (Cambridge) and NIE (Belfast). It can be seen that Belfast uses predominantly the rise-plateau L*+H 0% pattern in statements, and overwhelmingly in questions, revealing that these utterance types are generally not phonologically distinct. In Cambridge, statements mostly have a falling nucleus – a straightforward difference in usage. Almost half of the inversion questions also have this pattern, but the option exists to use a rise (L*+H H%) or a fall-rise (H*+L H%). As an aside, informal polling of students in Cambridge by the author, involving presenting a polite inversion question with each of these two patterns and asking 'which is more old-fashioned', has consistently shown the rise to be perceived as the 'old-fashioned' alternative. The subtlety of intonational variation is underlined by Ladd (1996: 122) who notes that the fall-rise nuclear tune H*+L H% on a request such as 'Can I have the BILL please?', which is

perfectly polite in British English, may be heard as condescending or peremptory by a speaker of American English.

So far we have looked at intonational variation that can be analysed in terms of discrete phonological categories. There are also differences which are a matter of phonetic realisation. One such is the way a dialect behaves under 'tonal crowding', that is, when there is only a very short time, because of limited phonetic material, to achieve several intonational targets (Hs and Ls). Idealising somewhat, there are two possibilities: to 'compress', and try to squeeze all the targets into the available time; or to 'truncate', and give up on achieving one or more targets. These strategies are schematised in (24):

(24)　　NINE　　　　　SIX　　　　　　NINE　　　　　SIX

　　　　H*+L　　　　H*+L　　　　　H*+L　　　　H*+L

　　　　　　compressing　　　　　　　　　truncating

The compressing dialect on the left attempts to realise the full fall despite the very short vocalic nucleus of 'six' (short because of the phonologically lax vowel, and pre-fortis clipping) by making the pitch change steeper. The truncating dialect on the right does not alter the rate of pitch change, and 'runs out of road' leaving an incomplete fall[8]. Hungarian has been described as a 'truncating' language (Ladd 1996, 132-136), while English is thought of as 'compressing'. Grabe (1998) showed that German truncates falls but compresses rises.

Table 1 summarises results in Grabe, Post, Nolan and Farrar (2000) for four dialects of English (with German added for comparison). It can be seen that SBE conforms to the stereotype of English as a compressing language, as does Newcastle. Leeds, despite being similar to SBE in terms of its intonational phonology, is truncating when it comes to

realisation, as is Belfast (which as we have seen is phonologically unusual, and lacks the rises on which to test this parameter).

	RISE	FALL
SBE	compresses	compresses
Newcastle	compresses	compresses
German	*compresses*	*truncates*
Leeds	truncates	truncates
Belfast	truncates	–

Table 1. Summary of truncation and compression of nuclear pitch accents in four English dialects.

Another source of realisational differences is the way in which intonational targets align with segmental material. In the extreme, alignment differences pretty much oblige us to recognise a phonological difference, as in the case schematised in Figure 3 of Connaught and Donegal Irish (Gaelic) reported by Dalton and Ní Chasaide (2003). It is tempting, and probably realistic, to speculate that the Donegal pattern might have developed as a result of a progressive historic drift rightwards of the intonational targets relative to the segments; but within our current intonational models it would be stretching credulity to do other than recognise Connaught as having H* accents and Donegal as having L* accents.

INSERT FIGURE 3 ABOUT HERE

On the other hand, take the comparison schematised in Figure 4 between SBE and the Scottish dialect of Anstruther, Fife (based on Aufterbeck 2003). In SBE (thin line), peaks are aligned with the accented syllables. In Fife, the peaks lag, and the accented syllable itself manifests a perceptually salient upglide. After the nuclear peak, the pitch also declines more gradually than in SBE. This impressionistic description captures the difference, but is the difference phonological (as we decided Donegal vs. Connaught Irish had to be), or realisational?[9] Aufterbeck argues that the difference here is realisational, and, in effect, that both dialects are associating prominent syllables with high rather than low pitch. This view is

in keeping with Farrar and Nolan (1999) who demonstrated that (utterance initial) H* peaks in SBE were not tightly tied to the accented syllable, being allowed to lag when there were no utterance-initial unstressed syllables – a case of tonal 'allophony'. Recognising SBE allophonically lagged peaks as H* opened the way to treating more severely lagged peaks in other dialects, such as Newcastle, as H* despite relatively low pitch on the accented syllable.

INSERT FIGURE 4 ABOUT HERE

Acknowledging that a substantial amount of intonational variation is realisational rather than a difference of phonological system may explain why there is relatively good between-dialect comprehension of intonation – occasional misinterpretation of affect notwithstanding. As we move from varieties of English which are historically indigenous to the British Isles to those which have emerged world wide, however, we find cases of fundamental prosodic differences influenced by substratum languages. These may give rise to comprehension difficulties. I will focus on one, potentially interrelated, cluster of prosodic differences.

It has long been recognised that languages can differ in terms of rhythm, and this is sometimes discussed in terms of *syllable-timing* and *stress-timing* (cf. Abercrombie 1967: 96-98). In the ideal syllable-timed language, each syllable would take up the same amount of time, or be *isochronous*, whereas in the ideal stress-timed language, it is the *stress-foot* which would be isochronous (the stress-foot consists of a stressed syllable plus any unstressed syllables which intervene before the next stress). According to this view French is a good example of syllable-timing, and English is a good example of stress-timing. In reality, however, experimental phonetics has failed to support either isochrony in any strict sense or a polar division of languages into two types. Nevertheless the impression which these terms sought to capture is real, and can be quantified. Recently progress has been made using a number of measures including the Pairwise Variability Index (PVI). The PVI simply

expresses the average difference between successive pairs of phonetic units – in duration, intensity, or vowel quality. It turns out, for instance, that as expected French has a lower durational PVI value for vowels and consonants than English (Grabe and Low 2002), reflecting more evenly timed syllables (well short of isochrony, of course).

The first application of the PVI was in fact to dialects of English, in a comparison of SBE and Singapore English – the latter of which has been described as 'syllable-timed'. Low (1998) and Low, Grabe and Nolan (2000) showed that, compared to SBE, Singapore English had less pairwise variability in vowel duration[10], vowel intensity, and vowel spectral dispersion (how peripheral a vowel is in the acoustic vowel space). To a large extent this reflects the fact that Singapore English is much more reluctant than SBE to reduce unstressed vowels to schwa. Singapore English could be said to have on average a less steep *prominence gradient* between syllables than SBE.

Separately, Low (1998) demonstrated that speakers of Singapore English do not deaccent given information (see section 3); they are quite happy to say things like *I OFFered her COFFee but she doesn't DRINK COFFee*, with a full accent on the second occurrence of 'coffee'. The strategy of backgrounding less important parts of the utterance by intonational means seems not to be grammaticalised. It is intriguing to speculate that at the level of pitch accents, too, Singapore English has a less steep prominence gradient; there may be a scaling of pitch accents according to information, but radical reduction (to zero) is not an option. It remains to be investigated whether there really is a systematic scaling of pitch accents according to information structure (short of deaccenting), or whether this kind of intonational signalling of informational value is simply absent. What is clear is that the lack of vowel reduction and the lack of deaccenting conspire to make Singapore prosodically radically different from (e.g.) SBE, and create problems for speakers of SBE in lexical access and comprehension. Deterding (1994:71) notes that the British model of intonation 'is

inappropriate for [Singapore English], because there is no clear nucleus acting as the focus of information or anchor for information within each intonational phrase', and it is almost certain that other world varieties of English will pose a similar challenge to our ingrained assumptions about English intonation.'

5. Conclusion

One of the 'design features' of speech is that pitch is variable independently of the sounds being produced. This is possible because the rate of vibration of the larynx does not have to match a resonant frequency of the vocal tract (unlike a brass player's lips, the vibration of which is coupled to a resonance of the tube which makes up the instrument). As a consequence pitch can be recruited to carry information over and above that borne by the vowels and consonants of language, functioning (as we saw in Section 1) either as lexically significant tone, or non-lexically as intonation. In doing so, pitch operates in tandem with durational factors and loudness.

Intonation, as an information channel independent of the words chosen, carries a number of quite distinct strands of information. We have seen that the ways in which it does so include signals mirroring physiologically-determined changes in pitch, on the one hand, and abstract phonological categories on the other. The latter may originally derive from grammaticalisation of biologically-determined frequency effects, but the status of phonological intonational categories as members of an abstract linguistic system means that their relationship to information is potentially arbitrary.

This arbitrariness should lead us to expect variation across languages and dialects, and Section 4 discussed such variation between dialects. Even those intonational effects whose basis in biology is more transparent are highly conventionalised, and so can vary.

Intonation, then, is just as significant a component of a dialect as the pronunciation of its vowels and consonants.

English is generally regarded as having a complex intonation system. English, of course, is not a tone language and so intonational categories can flourish without competition for the resource of pitch variation. But even among non-tone languages English seems to rely rather heavily on intonation for signalling. Schubiger (1965) compares English to German, which often uses pragmatic particles where English uses intonation. For instance, she cites 'rejoinders with the connotation "by the way you talk (or act) one would think you didn't know (or were ignorant of the circumstances)"', which in German naturally include the particle 'doch', for instance 'Ich bin doch eben erst aufgestanden' ('I've only just got up'). 'Doch' does not readily translate lexically, but the connotation is achieved in English by a low pre-nuclear accent:

(25)   I've ONLY just got U P

                        L*   H*+L 0%

It may seem to be stretching a point to claim unusual complexity for English intonation from this one little corner of information signalling, but whether or not the claim that English is unusual in the richness of its intonation can be proved there is no doubt that English intonation remains a highly elaborate and flexible communicative resource. This Chapter has sought to give an overview of some of the ways English intonation is used to convey a wide variety of information.

6.   Further reading

For an accessible and wide-ranging all-round introduction to the forms and functions of intonation, focusing on English, see Cruttenden (1997), while Ladd (1996) provides an objective overview and critique of the autosegmental-metrical approach to intonational phonology, and explores several problematic areas in the description of intonation including the definition and use of pitch range.

Gussenhoven (2004) deals with the tonal and intonational use of pitch across languages and discusses what is universal or language specific, and Hirst and Di Cristo (1998) offers a compendium of descriptions of the intonation of a large selection of languages.

A classic (and highly detailed) analysis of the prosody of British English within the 'British' descriptive framework is to be found in Crystal (1969), while more pedagogically oriented descriptions within the same tradition are provided by O'Connor and Arnold (1961/1973) and Couper-Kuhlen (1986).

Pierrehumbert (1980) is pivotal in theoretical terms, marking as it does the first comprehensive application of autosegmental mechanisms to the description of English intonation. It also provides wide overview of patterns found in American English. The ToBI transcription system, based on Pierrehumbert (1980), and information about its adaptation to other languages can be accessed on the web at http://www.ling.ohio-state.edu/~tobi/. Information about IViE, a further adaptation aimed for English dialect intonation, and references to work on intonational variation in the British Isles, can be found at http://www.phon.ox.ac.uk/~esther/ivyweb/.

References

Abercrombie, D. (1967). *Elements of general phonetics.* Edinburgh: Edinburgh University Press.

Aufterbeck, M. (2003). Scottish English intonation: a phonetic analysis of a Fife dialect. Doctoral dissertation, University in Cambridge.

Baumann, S., Grice, M. and Benzmüller, R. (2001). GToBI - a phonological system for the transcription of German intonation. In: Puppel, S. and Demenko, G. (eds) *Prosody 2000: speech recognition and synthesis*. Poznan: Adam Mickiewicz University, pp. 21-28.

Beckman, M. (1999). *ToBI.* http://www.ling.ohio-state.edu/~tobi/

Beckman, M. and Pierrehumbert, J. (1986). Intonational structure in English and Japanese. *Phonology Yearbook* 3, 255-309.

Couper-Kuhlen, E. (1986). *An introduction to English prosody*. Tübingen: Niemeyer.

Crystal, D. (1969). *Prosodic systems and intonation in English*. London: Cambridge University Press.

Cruttenden, A. (1997). *Intonation* (2nd edition). Cambridge: Cambridge University Press.

Dalton, M. and Ní Chasaide, A. (2003). Modelling intonation in three Irish dialects. *Proceedings of the 15th International Congress of Phonetic Sciences,* Barcelona, 1073-76.

Deterding, D. (1994). The intonation of Singapore English. *Journal of the International Phonetic Association* 24, 61-72.

Farrar, K. and Nolan, F. (1999). Timing of F0 peaks and peak lag. *Proceedings of the 14th International Congress of Phonetic Sciences,* San Francisco, 961-64.

Fletcher, J., Wales, R., Stirling, L. and Mushin, I. (2002). A dialogue act analysis of rises in Australian English map task dialogues. *Proceedings of Speech Prosody 2002*, Aix-en-Provence, 299-302.

Grabe, E. (1998). *Comparative Intonational Phonology: English and German.* Doctoral dissertation, MPI Series in Psycholinguistics 7. Nijmegen: Max Planck Institute for Psycholinguistics.

Grabe, E. (2001). The IViE labelling guide, version 3.
http://www.phon.ox.ac.uk/~esther/ivyweb/guide.html

Grabe, E., Nolan, F. and Farrar, K. (1998). IViE – A comparative transcription system for intonational variation in English. *Proceedings of the 5th International Conference on Spoken Language Processing*. Sydney, Australia.

Grabe, E, Post, B., Nolan, F., and Farrar, K. (2000). Pitch accent realisation in four varieties of British English. *Journal of Phonetics* 28, 161-185.

Grabe, E. and Low, E.-L. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. In: Gussenhoven, C. and Warner, N. (eds), *Papers in laboratory phonology 7*. Berlin: Mouton de Gruyter. pp. 515-546.

Grabe, E. and Post, B. (2002). Intonational variation in the British Isles. *Proceedings of Speech Prosody 2002*, Aix-en-Provence, pp. 343-346.

Gussenhoven, C. (2002). Intonation and interpretation: phonetics and phonology. *Proceedings of Speech Prosody 2002*, Aix-en-Provence, pp. 47-57.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Gussenhoven, C., Rietveld, T., and Terken, J. (1999). *ToDI: transcription of Dutch intonation*. http://lands.let.kun.nl/todi/todi/home.htm.

Hirst, D. and Di Cristo, A. (Eds). (1998). *Intonation systems: a survey of twenty languages*. Cambridge: Cambridge University Press.

Kingdon, R. (1958). *The groundwork of English intonation*. London: Longman.

Ladd, D.R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

Low, E.-L., Grabe, E. and Nolan, F. (2000). Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. *Language and Speech* 43, 377-401.

Nolan, F. and Jónsdóttir, H. (2001). Accentuation patterns in Icelandic. In van Dommelen, W.A. and Fretheim, T. (eds), *Nordic prosody: proceedings of the VIIIth Conference, Trondheim 2000*. Frankfurt-am-Main: Lang, pp. 187-198.

O'Connor, J.D. and Arnold, G.F. (1961/1973). *Intonation of colloquial English*. London: Longman. (2nd ed. 1973)

Ohala, J.J. (1983). Cross-language use of pitch: an ethological view. *Phonetica* 40, 1-18.

Ohala, J.J. (1984). An ethological perspective on common cross-language utilization of F0 in voice. *Phonetica* 41, 1-16.

Palmer, H.E. (1922). *English intonation, with systematic exercises*. Cambridge: Heffer.

Pierrehumbert, J.B. (1980). *The phonology and phonetics of English intonation*. PhD dissertation, MIT. Published 1988 by Indiana University Linguistics Club.

Scherer, K. (2001). Vocal communication of emotion: a review of research paradigms. *Speech Communication* 40, 227-256.

Schubiger, M. (1965). English intonation and German modal particles: a comparative study. *Phonetica* 12, 65-84.

Silverman, K., Beckman, M. E., Pitrelli, J., Ostendorf, M., Wightman, C, Price, P., Pierrehumbert, J. and Hirschberg, J. (1992). ToBI: a standard for labeling English prosody. *Proceedings of the Second International Conference on Spoken Language Processing,* Banff, Canada, 867-870.

Figure Captions

Figure 1

Acoustic representations of 'But Melanie's never been *near* the manuscript'. Top, spectrogram revealing segmental timing information; and bottom, time-aligned fundamental frequency contour.

Figure 2

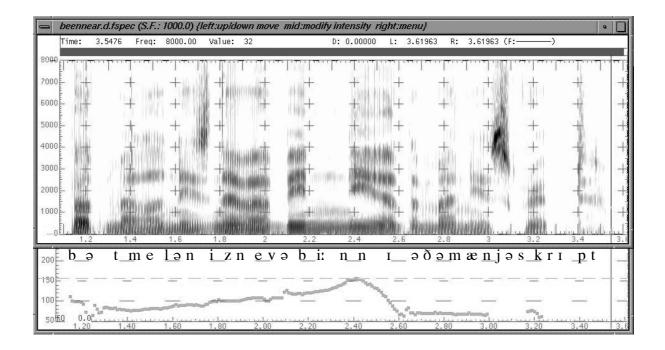Distribution of patterns between statements and inversion questions (after Grabe and Post 2002).
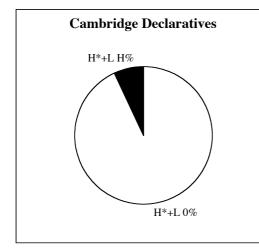
Figure 3

Schematic representation of the alignment of intonation pitch relative to prominent syllables (shown as boxes) in two dialects of Irish (Gaelic). The rectangles indicate the alignment of accented syllables.
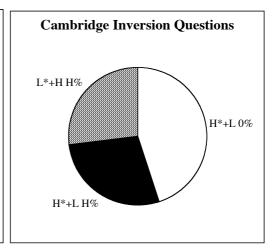
Figure 4

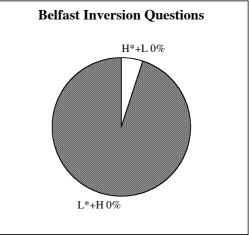Schematic comparison of lagged peaks (heavy line) as in Scottish English and aligned peaks. The rectangles indicate the alignment of accented syllables.
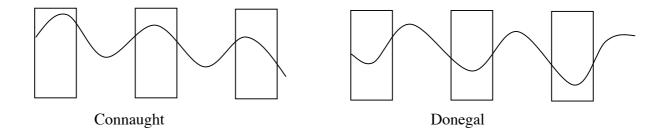
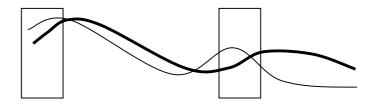beennear.d.fspec (S.F.: 1000.0) {left:up/down move mid:modify intensity right:menu}

Time: 3.5476   Freq: 8000.00   Value: 32   D: 0.00000   L: 3.61963   R: 3.61963 (F:————)

b ə t me lə n i zn ev ə b iː n n ɪ ə ð ə m æ nj ə s k r ɪ p t

Cambridge Declaratives

H*+L H%

H*+L 0%

Cambridge Inversion Questions

L*+H H%

H*+L 0%

H*+L H%

Belfast Declaratives

L*+H L%

H*+L 0%

L*+H 0%

Belfast Inversion Questions

H*+L 0%

L*+H 0%

Connaught

Donegal

WHERE   is the nearest   A I R   port