Notes on English Intonation.

Ken de Jong, Indiana University
2/23/06

Introduction.

Following are some very general notes to introduce you to an analysis of the intonation of English which is currently quite popular among both American linguists, as well as a burgeoning community of computational linguists and engineers.

The general outline of this analysis has been developed through introspection and examination of individual productions in such works as Liberman's thesis and Pierrehumbert's monograph based on her thesis work. Since that time, it has also been used in several approaches to both parsing running speech and synthesizing speech with intonational expressiveness. I have used the system in the latter way, and have become convinced that Pierrehumbert's analysis captures at least part of the expressive possibilities which the real English system has. Her analysis, if anything, falls short of the complexity of the real thing.

Be aware that there are several intonational schools of thought in the literature which predate the system presented below, including ones by Americanists such as Kenneth Pike, and a well-developed British school. For a relatively straightforward presentation of the issues which separate these various approaches, see Ladd (1994).

Two useful abstractions.

To understand how intonational transcription works, you must understand two different kinds of abstractions which the system relies on.

The first is a phonetic abstraction, namely that there is something which we can call intonation, a well-defined set of linguistic phenomena all working together to determine the pitch pattern of an utterance. This abstraction is very useful because it is fairly easy to get a good measure of what listeners perceive as the pitch pattern. We can do this by extracting the fundamental frequency of the voiced parts of the utterance, a task which is computationally quite easy. We can then take the fundamental frequency pattern, and analyze it as the result of a set of linguistic categories with a number of specific purposes, and an algorithm which implements the categories as events in the pitch of the utterance. Two points to note here: 1) not all intonational categories have the same function; being an intonational category only means that the category has a specific and categorical effect on the pitch pattern. 2) these categories do not determine all aspects of the pitch pattern; various other non-linguistic differences, such as emotional state, degree of involvement in the speech, and individual differences such as ones due to sex, also affect aspects of the pitch pattern.

The second is a functional abstraction. These intonational categories can be classified with respect to the two major types of prosodic functions. Prosody can be described as consisting of 'head' mechanisms and 'edge' mechanisms. Head mechanisms are those which act to pick out one piece of an utterance as different than its

neighbors, while edge mechanisms indicate which items go with which by marking the edge of a larger grouping. Intonational categories in the English system similarly function either to pick out syllables which are more stressed than their neighbors, or to mark the final edge of a piece of an utterance which is to be interpreted as a group.

Edge marking tones – boundary tones and phrase tones.

The intonational categories which you will likely find most intuitive are the ones which are used to mark edges. One reason for this, I believe, is that the English orthography actually writes some of these differences. For example, consider the following pair of sentences.

1) *This is a test sentence.*
2) *This is a test sentence?*

If you convert these into speech (by reading them out loud), you will note a very salient difference in the pitch contour at the end. In 1) the pitch falls throughout the last word, often ending with a little bit of creaky voice, while in 2) the pitch rises throughout the last word, perhaps ending higher than anywhere else in the entire sentence. Such differences in pitch pattern reflect discourse-related differences such as is captured by the use of the question mark in 2).

At a full stop, our system indicates the possibility of four different contours, the two which appear in likely renditions of 1) and 2), and two more, one which you will likely produce in the non-final members of a slowly rendered list, and one which you might produce when calling someone in for dinner. In the transcription system, you will see these represented in the following way (more or less). The fall in 1) is low throughout, and so is indicated as LL% (two lows with the % indicating the final boundary). The rise in 2) is high throughout, with a very brief rise to a super-high at the end, and so is indicated as HH% (two highs). The so-called list boundary starts low and rises slightly at the end, and so is indicated as LH%. The last one which appears in calling chants is basically high throughout, and differs from the HH% (question marker) in that it does not rise to a super high. Thus, since it is high to start with, it starts with a H, and since it is not as high as the super high at the end, it is relatively low, and so is indicated with a L%. This makes for a neat 4-way distinction as below, given with stereotypical examples of places where you might find them. (Note these are not the only places you will find them!)

LL% Terminal fall – statements.
HH% High plateau with upped high at end – covert questions.
LH% Low plateau with little rise at end – internal to lists.
HL% High plateau with no rise to a super-high – end of calling chants

Note that the Pierrehumbert system and later revisions such as ToBI also allow for the separation of the first tone indicating a possible plateau before the final edge(H or L) from the final edge tone (H% or L%). Such tones (called phrase tones and marked in various ways in different versions of the system) can then be found all by themselves in

the middle of a phrase. In my experience, the presence of these are relatively difficult to identify, and also make the identification of head-marking tones more difficult. Part of training in ToBI would be to get you to recognize such ambiguities and figure out how to resolve them. For this brief introduction, we will not discuss the bare phrase tones any further.

Head marking tones – pitch accents.

If you go back and reproduce the items in 1) and 2) again, and this time concentrate on the area around *test*, you will very likely notice a large difference in pitch pattern in this region in addition to what is going on at the end. The word *test* is a critical portion of the utterance in most prosodic analyses of English, because it is the last item which bears some degree of stress, usually called tonic or sentence stress. I chose this sentence because the words *test sentence* form a compound, and one of the peculiarities of English compounds is that they are most stressed on the first half. Thus, *test* is the most stressed syllable in the last content word in the sentence. In stressed locations such as this, English speakers also implement tonal events. Such events are often called pitch accents,, pitch because they involve parts of the pitch pattern, and accents because they are involved in making a particular syllable more prominent. Stressing this syllable makes it stand out from its neighbors. Thus, the tonal events on *test* are head-marking events.

Here, like the boundary tones just discussed, there are tonal differences associated with different discourse conditions. In 1) you very likely will produce the stressed item with a high pitch somewhere on it, while in 2) you very likely will produce the stressed item with a relatively low pitch. Thus, the difference between vanilla statements and covert questions is not only in the presence of LL% boundary tones in one and in HH% boundary tones in the other, but also in the presence of a H accent in one, but a L accent in the other. Since there is a categorical difference in how you use pitch to stress the tonic item, you need to have a categorical difference between H* and L* accents. (The star here indicates that the tone is associated with the stressed syllable.)

In addition to using relatively high and low pitch, there are more complicated rising and falling pitch accents which differ from the simple low and high accents in what they indicate. Our system captures these differences in the local use of pitch in the accent by combining H's and L's in various ways to get rises and falls. Thus, in addition to H* which indicates a generally high pitch around the stress and L* which indicates a generally low pitch around the stress, we can also have H+L's (falling accents), and L+H's (rising accents). To illustrate the difference between a simple H and a L+H, consider the following two conditions:

3) *We will be having you read bunches of utterances for some obscure reason related to why anyone would be interested in linguistics. The first is a test sentence. It's just there for practice.*
4) *The first is not a real sentence, the first is a test sentence.*

In producing *test sentence* in 3), it is likely there will not be an appreciable rise in pitch, while in 4), where it explicitly contrasts with the preceding *real*, it is likely that there will be an appreciable rise in pitch from the *is a* to *test*. In fact, it is a general property of contrasting items that they get rendered with a relatively low pitch on the material preceding the stressed item and a sudden rise to a peak on the stressed syllable. If you read over 4) several times, emphasizing the contrast more and more each time, this rising pitch event associated with *test* will become more and more apparent.

There is one further contrast which must also be understood before we can spell out an inventory of English accents, namely a difference between rising accents as to how they are timed with respect to the stressed syllable. In 4) the rising accent is seen in the relationship in pitch between the items immediately preceding the stressed syllable and the pitch on the stressed syllable itself. However, there are other examples of rising pitch accents in which the low pitch predominates in the stressed syllable, and the high does not become realized until very late in the syllable or in the following syllables. Pierrehumbert & Hirschberg (1991) discuss fairly clear examples of this accent such as the following:

5) *A: Alan's such a klutz.*
   B: He's a good badminton player.

Here the intended meaning of the second response is that B is not sure that playing badminton qualifies one as not being a klutz. In the intended rendition there is a low pitch on *bad* and a rising pitch on the immediately following syllable, and then another fall to a general low ending in LH% phrase tones. Another example they discuss is the following:

6) *A: Did you take out the garbage?*
   B: Sort of.
   A: Sort of!?!

Here, the intended rendition of *Sort of* starts low in *sort* and rises, and then falls and rises again at the end. The intended meaning is very much like that in 5), namely, B is not really sure what she did counts as taking out the garbage. A's rendition of *sort of* in the last line has exactly the same pattern as B's, a rise through *sort* followed by a fall and a rise at the end, though the rises and falls are more exaggerated. What's important in each of these cases, *badminton* in 5), and both *sort of*'s in 6), is that the stressed syllable exhibits a distinctly low pitch and the rise which comes much later than the rise in 4).

In order to annotate this difference, Pierrehumbert used the * to indicate which part of the contour is to be associated with the stressed syllable. Thus, the contour in 4) is annotated as a L+H*, since the H part appears on the stressed syllable, and the L part simply comes some time before it. By contrast, the contour in 5) and 6) is annotated as a L*+H, since the L part happens on the stressed syllable, and the H part appears some time thereafter.

If we combine the parameters mentioned above, we get the six possibilities listed below, which are listed along with likely places where you will find them. Four of them we have already discussed:

H* - vanilla high accent – with new information in declaratives
L* - vanilla low accent – with stressed items in covert questions
L+H* - early rising accent – on items explicitly contrasted with something else
L*+H – late rising accent – on items whose link to the discourse is in doubt

And two more are included in the system:

H + L* - fall onto accent – on rhetorically stressed items
H* + L – downstepping high – in lecturing, finger-wagging statements.

In my experience, the H+L* seems pretty rare, but this might be a point of dialect difference. The last one is also odd, though for a different reason. In the original formulation of the system, the H*+L is actually identical to the H*, except that it affects the realization of later accents. For this reason, the ToBI revision of Pierrehumbert's system has been in the direction of leaving out this tone from the inventory.

Pitch Range.

One final aspect of intonational modeling must also be mentioned, that is the notion of pitch range. As I noted above, the tone category sequences do not all by themselves determine the pitch contour for an utterance, but other non-linguistic (non-conventionalized) factors also affect the final realization of pitch. One approach to handling these less conventionalized effects, such as what may be due to emotional involvement, is to allow for modulation of the overall range of the pitch movements. The general approach used in most models is to specify a 'pitch window', which indicates the range of pitch to be used at any given time. The top of the 'window' is where you find the H's and the bottom of the window is where you find the L's. This window can be affected by a number of different factors, which work in different ways.

Some factors are global in that they typically affect a large portion of speech. Take, for instance, the effects of emotional involvement. When people get irate, there is a strong likelihood that the both H's and L's will be higher, and that the difference between the H's and L's will be bigger. This 'larger and higher window' will often affect entire sentences. You will also likely find such global shifts in window size if you examine how people do narratives which include parentheticals and quotations. Parentheticals often are rendered with a narrower window, while quotes often involve a larger window.

Other factors which affect pitch range can be localized to one particular location in the utterance. The most commented upon is the effect of <u>downstep</u> (sometimes called catathesis). Downstep is a very regular lowering and narrowing of the pitch range which happens in the presence of the accents. In Pierrehumbert's analysis, any tone which is

composed of two tones (the rising L+H and falling H+L accents) also trigger downstep. You can easily imagine this effect in an emphatic rendition of the following sentence.

7) *I don't want horses and dogs; I want sheep and cats.*

If you are contrasting *horses* with *sheep* and *dogs* with *cats*, you will very likely produce this sentence with L+H accents on all four items (probably L*+H on *horses* and *dogs*, and L+H* on *sheep* and *cats*). If you do so, you will also notice that the second item in each list, *dogs* and *cats*, will both be lower in pitch than the first, *horses*, and *sheep*. This conventionalized lowering is taken to be due to the downstepping effect of the complex rising accents.

One can also see this conventionalized downstepping very clearly in phrases with multiple accents rendered in a finger-wagging lecturing style where the clear intent of the style is to indicate that 'you should know this by now'. For example,

8) *You just don't seem to get it. <sigh> Insert tab A into slot B. Repeat it four times.*

In this situation, the rendition of the last two sentences, which we can assume have been rendered several times before in the extended discourse, will likely not exhibit huge rising or falling accents. Nevertheless, I have heard this sort sentence produced with clear downsteps between each accent. Due to sentences like these, one must conclude that the occurrence of downstep does not necessarily demand the obvious existence of rising or falling accents. In Pierrehumbert's analysis, this is due to the H*+L tone category which is locally the same as a plain H*, except that it triggers the lecturing downstep effect. In other systems, such as the ToBI revision, this downstepping is marked with an explicit marker (an exclamation point placed before the affected accent.

Conclusion.

Above, I have given a rough sketch of a model of English intonation. Two final points should be highlighted. 1) this model generates a large number of intonational possibilities; on a stressed syllable before a full stop, we can have 1 of 6 accent types preceding 4 phrase/boundary tone contours for a total of 24 ways of rendering a single word in isolation. When embedded in running speech, the combinatorics of this system produce an astonishing number of possibilities. This complexity, I believe, indicates just how much expressive power resides in conventionalized English intonation. 2) however, this model also specifies the number of possibilities, and hence restricts what can be done with an utterance. Actually, this is the point of a grammar, namely to specify out of the imaginably infinite a particular number of conventional categories. I believe that, here too, the model presented above helps us put our finger on what is going on in English tonal behavior, because it allows us to identify from a set of categories what speakers are doing.

References.

Ladd, D.R. (1992). An introduction to intonational phonology. In G.J. Docherty and D.R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press, pp. 321 - 334.

Pierrehumbert, Janet, and Julia Hirschberg (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan and M. E. Pollock (eds.), Intentions in Communication (pp. 271_311). Cambridge MA: MIT Press.

Pierrehumbert, J.B. (1980), *The Phonology and Phonetics of English Intonation*, Ph.D. dissertation, M.I.T., available through the Indiana University Linguistics Club, Bloomington, Ind.

Pitrelli, J.F., M.E. Beckman and J. Hirschberg (1993). Evaluation of prosodic transcription labeling reliability in the ToBI framework. In *Proceedings of the International Conference on Spoken Language Processing*, Yokohama.