Phonology of intonation
By Carlos Gussenhoven

1. Introduction
Pierrehumbert’s (1980) thesis marked the beginning of a new period in intonational research by presenting a model which separated the phonological representation from its phonetic implementation, thus allowing a characterization of the notion ‘possible prosodic structure’ as distinct from an account of the phonetic details of intonation contours. In this state-of-the-art article, I sketch the historical background of this model, its role in redefining the phonetics-phonology interface, and what I see as the main developments. Introductions to the model are Ladd (1996), who labelled it the Autosegmental-Metrical (AM) model, Shattuck-Hufnagel & Turk (1996) and Beckman (1996). I leave treatments in other frameworks, notably Hirst & Di Cristo (1998), out of account.

The model is autosegmental because it has separate tiers for segments (vowels and consonants) and tones (H,L). It is metrical because it assumes that the elements in these tiers are contained in a hierarchically organised set of phonological constituents, as depicted in (1), to which the tones make reference in several ways. Amongst themselves, tones are organized into pitch accents and boundary tones. Tones may or may not be associated with Tone Bearing Units, whose nature varies across languages and ranges from the sonorant mora, as in Japanese, to the accented syllable, as in English. Representation (1) gives a possible pronunciation of the English proverb Too many cooks spoil the broth. The pitch accents have a starred tone to indicate their association with the accented syllable, as in English. Representation (1) illustrates a commonly adopted set of prosodic constituents, but other constituents have been postulated in intonation, notably the Intermediate Phrase and the Accentual Phrase, while for Japanese, a mora-tier would also be relevant to intonation. A non-crucial assumption is that boundary tones are optional, the first \( i \) in (1) not having any.

As a surface representation, (1) is a theory of a mental construct for a speaker of British English, who also possesses a phonetic implementation module which, among other things, translates every tone into an F₀ target. The phonetic ‘alignment’ of the target with the segmental tier, its timing, will to some extent be language-specific, as will its ‘scaling’, its F₀. Over and above the effects of these implementation rules, the speaker’s psychological condition and communicative purpose will influence the overall pitch range, so that the number of pronunciations of (1) is infinite. In the schematic implementation (2), the targets are given as bullets, which are connected by line segments that represent the F₀ interpolations between them. The boundary tone is attached to the bracket, as per convention (cf. Hayes & Lahiri, 1991).

My concern here is with substance (representations, contours), not with functions (focus, intonational meaning). A complete analysis of an intonational system will comprise a phonology as well as the morphology. Researchers usually feel more at ease with the phonological analysis, where the smallest units are given by the model. The elements in a morphological analysis are not given \( a \ priori \), however. In principle, the \( i \)-wide contour could be a single morpheme, any sequence of tones in the contour could be, or each tone could be a morpheme. A whole-contour proposal occurs in Liberman & Sag (1974), where the ‘contradiction contour’, for instance, consists of a two-accent sequence, a position which

Carlos Gussenhoven, Centre for Language Studies, University of Nijmegen, Erasmusplein 1, 6525 HD Nijmegen, The Netherlands, c.gussenhoven@let.kun.nl
has not been followed (Bolinger, 1986, 245). Morphemes consisting of pitch accents plus boundary tones occur in proposals for English intonational meaning (cf. Cruttenden, 1997) and Hayes & Lahiri (1991) for Bengali. Pierrehumbert & Hirschberg (1990) propose an analysis in which (virtually) every tone is a morpheme, a position defended in Bartels (1997). Recently, Dainora (2001, 2002) advanced the argument that the relatively high predictability of transitions between pitch accents and following tones in American English suggests that these elements are not morphemically independent. Of course, the same phonological analysis may be compatible with a large number of morphological analyses, just as different phonological analysis may be compatible with a large number of morphological analyses, just as different phonemes consisting of pitch accents plus boundary tones occur in proposals for English intonational meaning (cf. Cruttenden, 1997) and Hayes & Lahiri (1996), an additional intonationally defined constituent was introduced, the Intermediate Phrase (ip), ranked immediately below the i. In this way, two degrees of depth became available for an intonational boundary. Two situations given by Beckman & Pierrehumbert (1986) are illustrated in (3) and (4), where the square brackets enclose ip’s and the curly brackets the i. In (3), the two adjectives are considered to be followed by just an ip-boundary, because the disjuncture with what follows is less complete than that observed for a full-fledged i-boundary. In (4), the ip-boundary after nine is motivated by the high F0 peak on eighty, which is due to an interruption of the downstepping pattern shown by the preceding F0 peaks on one and nine. The F0 contours are given in panels (a) and (b) of Figure 1, respectively.

![Figure 1](image-url)

(3) { [ It's eleven and one and nine ] [ and eighty ] }

(4) { [ It's eleven and one and nine ] [ and eighty ] }

Analyses of West Germanic languages, including English, that do without ip and the ‘phrase tone’ that has been related to it (see below) are common (van den Berg, Gussenhoven, & Rietveld, 1992; Féry 1993; Grabe, 1998a; Gussenhoven, Terken & Rietveld, 1999). In those analyses, the first internal ip-boundary of (3) would be an i-boundary, while the second as well as that in (4) would not be an intonational boundary; the exemption from downstep would be attributed to the fact that it is the last (‘nuclear’) pitch accent of the i (Ladd, 1983, 735; Gussenhoven, 1983b).

Third, the separation into pitch accents, symbolized T*(T), and boundary tones, symbolized T%, harks back to Trager & Smith Jr.’s (1951) juncture phonemes (# ‘falling’, - ‘sustained’, || ‘rising’), which existed by the side of the pitch phonemes. A division between pitch accents and boundary tones avant-la-lettre can also be recognised in the work in the 1970s and 1980s by Hans ‘t Hart, René Collier and Antonie Cohen on...
Dutch (t Hart, Collier, & Cohen, 1990), as argued by Ladd (1996). They divided up the pitch movements of Dutch into ‘accent-lending’ and ‘non-accent-lending’, but stopped short of equating the latter with boundary features. An example is given in (5), a non-final contour, where ‘i’ is an accent-lending rise, ‘B’ a non-accent-lending fall, ‘A’ a steep accent-lending fall, and ‘2’ a non-accent-lending rise. (When two accent-lending movements appear on the same syllable, only one accent is produced.)

\[
(5) \quad 1\&B \quad 1\&A \quad 2
\]

We gaan vanavond naar de schouwburg
‘We are going to the theatre this evening, …’

The chief motivation in Pierrehumbert (1980) for ending i with two tones, a phrase accent, T-, and a boundary tone tout court, T%, was that in many contours two targets can be identified after the last (‘nuclear’) pitch accent. For instance, in (6), the L* pitch accent is followed by a high target at the end of theaccented Manotowoc, as well as a final high target, for which H- and H% were postulated, respectively. I reproduce her contour 2.29 as (6). This example would be natural in a conversation where someone had just asked the speaker if he knew of any towns with bowling alleys.

\[
(6) \quad \{ [ \text{Does Manotowoc have a bowling alley?} ] \} \quad \begin{array}{c} \text{L*} \\ \text{H} \end{array} \quad \begin{array}{c} \text{H} \\ \text{H%} \end{array}
\]

In Beckman & Pierrehumbert (1986), T was reanalysed as a boundary tone of the ip. As a result, ip’s ended in T-, and i’s in T-T% in the new analysis, since the right edge of every i coincides with that of an ip. This analysis was transferred to the practically oriented transcription system for American English that was derived from the Pierrehumbert model, Tones and Break Indices (ToBI), cf. Beckman & Ayers, 1994, see also below).

Fourth, the notion that a contour is an interpolation between levels, such that only the beginning and end points of a movement result from a target and thus from a tonal specification, was inherent in the analyses by Pike (1945) and Trager & Smith Jr. (1951), who used numbers to indicate pitch levels. Pike’s (7), for instance, shows interpolations by means of dashes. In his system, ‘1’ was the highest pitch phoneme, ‘4’ the lowest. The notion of a starred tone was present in that accented levels, which begin a ‘primary contour’, are marked with the degree symbol. In this case, the representation translates quite readily into Pierrehumbert’s (8). Other theories identified pitch movements as the basic elements instead of the interpolations between level pitches, although the idea of non-specification was present to the extent that some movements were stretchable, their duration being determined by the length of the segments over which they were pronounced, like the rising head of O’Connor & Arnold (1973) or the gradually rising pitch movement ‘4’ of t Hart, Collier, & Cohen (1990).

\[
(7) \quad \text{I wanted to do it, but I couldn’t} \\
\begin{array}{c} 4- \quad 2- \quad 4-3/4- \quad 2-4 \end{array}
\]

\[
(8) \quad \text{I wanted to do it, but I couldn’t} \\
\begin{array}{c} \\ H^+ \end{array} \\
\begin{array}{c} L-H\% \quad H^+L-L\%
\end{array}
\]

Fifth, the idea that the string of tones contained lexical and intonational tones forms the hallmark of Bruce (1977), who isolated the contribution of the lexical tones of Stockholm Swedish from that of the intonational tones, representing them as a string of pitch levels that were timed with the stresses and phrase ends much as in an autosegmental description (Pierrehumbert, 2000; Ladd, 2000). Ladd (1983a) characterized this type of description as a Tone Sequence model, to distinguish it from descriptions that superimpose accentual contours on phrasal intonation contours, termed Contour Interaction models by Ladd, as represented by Gårding (1983), Thorsen (1978, 1983) and Vaissi`ere (1983), as well as by Fujisaki’s (1983) model. The integration of lexical and intonational tones played an important role in the description of Japanese in the work by Pierrehumbert & Beckman, and later in that of Norwegian, varieties of Basque and Dutch dialects.

Sixth, the idea that there are only two tones, H and L, was also part of Bruce’s thesis. Earlier, Liberman (1975) had described the intonation of American English with the help of four tones, H, L, raised H and raised L, thus staying closer to the earlier descriptions. At that point, descriptions were still vulnerable to Bolinger’s (1951) criticism that four-level transcriptions of English intonation, like (7), were arbitrary, because 2–4 would not be discretely different from, say, 3–4 or 1–4.

2.1 The 1986 model for American English

Beckman & Pierrehumbert’s (1986) revised model includes six pitch accents, H*, L*, H*+L, L*+H, L+H* and H+L*. An optional initial boundary-%H precedes the i, and the ip and i are obligatorily closed by boundary tones, as explained above. The grammar can be given as in (9), where parentheses include optional elements, accolades alternative options, and subscripts stand for ‘n or more occurrences’, as usual. The part enclosed between the outermost (…) indicates the ip, of which there must be one or more, containing one or more pitch accent.

\[
(9) \quad \text{The tonal grammar of Beckman & Pierrehumbert (1986)}
\]
Perhaps more than Bruce’s (1977) ‘pitch rules’, Pierrehumbert’s implementation rules made it clear that the generation of an infinite number of $F_0$ values between the highest and lowest pitches allowed the distinction between phonological representations and phonetic contours to become very clear. The representations are there to describe what is a possible linguistic expression and thus characterize the contrasts of the language. For this purpose, two tones suffice, for English and for many other languages that have been described since. The phonetic implementation will create contextually appropriate targets for each tone. Downstep provides a suitable illustration. In (11), each non-initial $H^*$ is downstepped relative to the preceding $H^*$. In Beckman & Pierrehumbert (1986), downstep applies to $H^*$ and $H^-$ after a bitonal pitch accent. Or again, a mid tone at the end of the phrase of an Intonational Phrase is obtained by the combined working of DOWNSTEP, which lowers $H^-$ after a bitonal pitch accent, and UPSTEP (12), which raises $L^%$ to the level of a preceding $H^-$, and raises $H^%$ above the level of a preceding $H^-$. In the half-completed fall (13), therefore, the $H^-$ is downstepped and the $L^%$ is upstepped. UPSTEP is also responsible for the extra-high $H^%$ in (13b).

(10) PB DOWNSTEP: $H \rightarrow \text{!H/T*T} \ldots \ldots \text{T%}$
   (Implementation)

(11) There are many intermediate levels
   $L+H^* \quad L^-H^* \quad L+H^* \quad L^- L^%$

(12) PB UPSTEP: $T \% \rightarrow \text{raised T% / H- \ldots \ldots}$
   Implementation

(13)

a. { [ It’s warmer ] } b. { [ It’s warmer ] }
   $L+H^* \quad H-L^% \quad I^* \quad H-H^%$

In addition to the two implementations rules given in (10) and (12), and the abstract downstep-trigger $L$ of $H^*+L$ (see section 3.2), a final implementation rule was that $L^*$ of $H+L^*$ was realised as a downstepped $H$, $H+L^*$ being equivalent to $H+!H^*$. This pitch accent is used to described the spreading plus-downstepping pattern illustrated in (14). This fairly abstract system has in practice been replaced with the practically oriented ToBI transcription system, but may still be seen as the ‘real’ Pierrehumbertian analysis behind the newer system. ToBI did away with the abstract downstep trigger, and noted downstepped $H$‘s directly with !, following a suggestion by Ladd (1983b). Earlier $H+L^*$ thus also became $H+!H^*$ in the new system.

(14) It’s a common type of contour
   $H^*$ $H+L^*$ $L^-$ $L^%$

In addition to downstep, two further phenomena are involved in $F_0$ downtrends (Pierrehumbert, 1980; Liberman & Pierrehumbert, 1984; Ladd, 1984; Pierrehumbert & Beckman 1988). First, there is ‘declination’, a time-dependent, gradual $F_0$ lowering, associated with one or more $i$‘s but otherwise context-independent. Second, there is ‘final lowering’, loosely associated with the last syllable(s) of the Utterance ($\text{t}$). Downstep may be seen as a grammaticalization of declination, just as many tone languages have grammaticalizations of final lowering. Earlier, these three concepts tended to be collapsed under a single notion, usually simply referred to as ‘declination’.

Questions like ‘How do you pronounce $H$ in English’ will thus require a lengthy consideration of all the contexts in which $H$ can occur. The difference between discrete (‘digital’) representation and gradient (‘analogue’) implementation as drawn for intonation by Pierrehumbert (1980) was extended to the phonology-phonetics interface generally by Pierrehumbert (1990), Keating (1990), Cohn (1990) and others. This work has crucially contributed to our understanding of speaker control in phonetic implementation (Kingston & Diehl, 1994), phonetic underspecification (see below), and the understanding of the place of paralinguistic meaning in intonation. To take up the latter point briefly, implementation rules that apply as a function of phonological context (e.g. DOWNSTEP, UPSTEP) are to be distinguished from structurally contextless variation signalling paralinguistic meanings (Gussenhoven, 2002). While English contrasts early peaks ($H^*$) with late peaks ($L^*+H$) (Pierrehumbert & Steele, 1989), the height of the peak is gradiently variable and correlates positively with degrees of urgency. Such variation may also correlate with other attitudinal meanings, like surprise. Differences in meaning are therefore no longer criterial for phonological contrasts (Ladd & Morton, 1997; Gussenhoven, 1999). Haan (2002)
shows that this type of variation may be a function of sentence categories like ‘declarative question’, wh-question, and yes-no question, each of which appears to some extent to have its own phonetic profile in Dutch.

3. Some developments
Pierrehumbert’s model has been applied to a number of different languages, like Japanese (Pierrehumbert & Beckman, 1988), Bengali (Hayes & Lahiri, 1991), varieties of Latin-American Spanish (Sosa, 1991), German (Uhm, 1991; Féry, 1993), Palermo Italian (Grice, 1995b), European Portuguese (Frota, 1998), varieties of Korean (Jun, 1993), Basque (Elordieta, 1997; Hualde, Elordieta, Gaminde, & Smiljanic, 2002), French (Post, 2000; Jun & Fougeron, 2000), Bern Swiss German (Fitzpatrick-Cole, 1999), and European Spanish (Face, 2002), among others. This section discusses some noteworthy developments.

3.1 Phonetic underspecification
From Pierrehumbert (1980) onwards, a distinction has generally been made between interpolation, the creation of F0 values in the phonetic implementation between the targets of phonological tones, and specification through spreading. The point was particularly clearly made in Beckman & Pierrehumbert (1986, 263), Pierrehumbert & Beckman (1988). Japanese Accentual Phrases (henceforth x), often no more than word-sized prosodic constituents, are either accented or unaccented. Both of these typically begin with a pitch rise, while accented x’s are characterized by a subsequent sharp fall, from the accented syllable. Unaccented x’s lack this fall, and after the initial rise the pitch slowly descends. Earlier descriptions described this slow descent as fully high pitch, and assumed that it was to be explained by a spreading H. This is shown schematically in (15a), after Poser (2001), which represents an x plus the rise of a following x. In (15b), the non-spread, phonetically underspecified situation is given, after Pierrehumbert & Beckman (1988), with an interpolation between the high target in the first x and the low target in the next. The evidence that (15b) is the superior theory comes from the dependence of the slope between the high and low targets on their distance. Pierrehumbert & Beckman showed that the longer the first x was, the less steep its sloping F0. By contrast, theory (15a) would predict that the slope remains high up till the last syllable of the first x, regardless of its distance from the first x’s beginning.

3.2 Unassociated tones
Pierrehumbert (1980) applied the Africanists’ distinction between associated and unassociated lexical tones to intonational tones. There are two different uses to which unassociated (or ‘floating’) tones have been put. The first is based on the role that they have as a trigger for downstep in Bantu languages, while remaining without a target themselves (Stewart, 1966; Clements & Ford, 1980). Thus, a disyllable may have three tones, HLH, of which the first H is realized as high pitch on the first syllable, and the second as a downstepped H on the second syllable, L itself obtaining no target. Similarly, Pierrehumbert (1980) describes one of the downstepped patterns of English by H*L H*, whereby the bitonal pitch accent causes the second H* to be downstepped, even when there was no low target in between. The nonrealization of L in H*L+L needs to be specified in the implementation rules, and was specific to Pierrehumbert’s description of American English.

Generally, however, floating tones are realized in intonation. They differ from associated tones in that the timing of their targets is not attributable to a specific point in the segmental tier, but is rather with reference to the targets of other tones. Typically, while one tone in a pitch accent, T*, associates with the accented syllable, the leading or trailing T receives a target which occurs some fixed distance before or after that of T*. As a result, the H of L+H as used on Ríganamarole will occur on –ma-, the second syllable after the accented Rig-7, while in Stein, it will be realized on the same syllable as L* (Beckman & Pierrehumbert, 1986). Bruce (1987) showed that the focus-marking tone of Stockholm Swedish, which is a single H, is floating, and is pronounced after the associated lexical tone complex HL. Since the lexical tone contrast depends on whether the L (Acc 1) or the H (Acc 2) associates with the stressed syllable, the focal H is pronounced later in Accent-2 words than in Accent-1 words. This is shown in the minimal pair in (16a, b), where the target for focal H falls in the [nd] in ‘the duck’, but in [c] in ‘the ghost’. In (16c), where Accent 1 occurs on a final syllable, H occurs inside the stressed syllable (Bruce, 1987; Bruce, 1990). (The L-boundary tones define the i; the initial L in (16a) is truncated for lack of segmental space; see also below.) Because the focal H does not associate, the term ‘pitch accent’ has been avoided for this tone. Pierrehumbert & Beckman (1988, 251) have in fact suggested it may be a boundary tone which is pronounced early.
3.3 Bitonal pitch accents

A decision to designate one of the two tones in a bitonal pitch accent as T* may be based on nothing more than that its target is closer to the accented syllable. Not all pitch accents appear to be timed in this asymmetrical fashion, however. Work on Standard Greek has shown that the distance between the two targets of the prenuclear LH pitch accent depends on the duration of the stressed syllable, since the low target occurs just before the syllable and the high target just after it (Arvaniti, Ladd, & Mennen, 2000). As a result, the rise is longer when the syllable has more consonants; a further implication is that it is no longer clear which is the starred tone, L or H. This pitch accent could be interpreted as a branching structure, which as a unit associates with the accented syllable, as the authors suggest. The internal structure of the bitonal pitch accent had earlier been discussed by Grice (1995b), who recognised a ‘cluster’ by the side of a ‘contour’. This distinction requires that a tonal node is introduced between the syllable and the tones. In the contour, the two tones are gathered under the tonal node, a representation originally proposed by Yip (1989) for tone complexes in south-east Asian languages, where they are underlying, while in the cluster the tones are directly dominated by the syllable, a representation Yip used for tone in African languages, where contour tones are typically derived. Grice claimed that in English, pitch accents with trailing tones, like L*+H are contours (cf. (17a)), but that pitch accents with leading tones, like L+H* are clusters (cf. (17b)). One of her arguments is that leading tones tend to be truncated when the accent is i-initial, while i-final trailing tones are not. To return to the Greek case, a structure like (17a) could be provided with a star for the whole complex to serve as the prenuclear LH (Arvaniti, Ladd, & Mennen, 2000).

(17)  

Timing characteristics were used by Frota (2002) to decide between an analysis of the European Portuguese focal and non-focal pitch accents as either bitonal H*+L and H+L* or as HL combinations of boundary tone and single-tone pitch accent. The evidence leads her to conclude that both are bitonal pitch accents, but that only the targets of H*+L are characterized by a constant interval.

Detailed phonetic studies of the timings of targets have on the one hand revealed cross-linguistic and cross-varietal variation, and on the other a tendency for targets to be coupled to ‘segmental landmarks’, like the syllable offset, as in Mandarin lexical tones (Xu, 1998), or the CV-boundary (Ladd, Faulkner, Faulkner, & Schepman 1999). They can be sensitive even to the tenseness of the vowel (Ladd, Mennen, & Schepman 2000). Evidence for the influence of the length of the onset and the sonorant status of consonants is provided in Prieto van Santen, & Hirschberg (1995) and Rietveld & Gussenhoven (1995), while the role of constraints like the speed of F0 movements and the time difference between implementation and articulatory effect are discussed in Xu (2002).

The assumption that the targets of bitonal pitch accents are close together is not generally shared. The timing of trailing tones in English was made dependent on the distance to the next accent in Gussenhoven (1983a, 1988, 1999), Féry (1993) and Grabe (1998a). Part of the motivation was functional, as illustrated by (18a), after Gussenhoven (1983a), which has two occurrences of what appears to be the same neutral pitch accent. The target of the trailing L of the non-final H*L is timed rightmost, and bounded by the following associated tone. The trailing L of the final pitch accent is however constrained so as to occur immediately after the target of its T*. Because the timing of the trailing tone is context-dependent, the ‘+’ is avoided, as it suggests that the tones are always realized close together. The representation of contour (18) in Pierrehumbert’s theory is (18b), which has different pitch accents, H* and L+H*. A second argument for assuming (18a) is that both the timing and the scaling of the low target before the second peak is not precise, and could be higher and earlier with no appreciable perceptual difference, which suggests that L*’s rightward drift is somewhat imprecise. This is not what is suggested by the pitch accent L+H*, however. A third argument is that the right-moving trailing tone, or ‘displaced’ tone, to use Grabe’s (1998a) term, also occurs in non-final L*H and H*+LH (Gussenhoven, 1983a). The rightward displacement was termed a ‘partial linking’ in Gussenhoven (1983a), where ‘linking’ referred to the coherence of the two pitch accents, and was seen as a step towards ‘complete linking’, the deletion of the trailing tone.

(18)

3.4 Boundary tones

Boundary tones mark the edges of prosodic constituents. Prosodic phonology holds that speech is
produced in batches of segments that are hierarchically ordered: within any such batch except the lowest a smaller batch can be identified. The gestural integration correlates inversely with rank: syllables are highly integrated articulations, while Utterances may contain noticeable pauses. In (1), which represents a widely adopted view of this hierarchy (Selkirk, 1978; Nespor & Vogel, 1986; Hayes, 1989) the syllable (σ) is included in the foot (F), the foot in the phonological words (ω, also ‘prosodic word’), the phonological word in the phonological phrases (φ) or Accentual Phrase (α), depending on language, the phonological phrase in the intonational phrase (τ) or Intermediate Phrase (ip), depending on language or analysis, and the intonational phrase in the Utterance (ω).

The reality of these prosodic constituents is apparent from a number of phenomena. The context for segmental processes like assimilation is often defined by the boundaries of specific prosodic constituents (e.g. Nespor & Vogel, 1986). Second, their boundaries reveal themselves through lengthening at the end (Wightman, Shattuck-Hufnagel, Ostendorf, & Price 1992; Gussenhoven & Rietveld, 1992) and consonantal strengthening at the beginning (Fougeron & Keating 1997; Fougeron, 2001; Cho & Keating, 2001). For instance, in Tiptoe through the tulips, the v-initial [t] in tip- will be longer and have a more extensive articulatory contact than the ω-initial [t] in tu- which in its turn will be longer and stronger than the F-initial [t] of -toe. The presence of a pitch accent independently increases the durations of segments in, and to some extent near, the accented syllable (Beckman & Edwards, 1990; Cambier-Langeveld & Turk, 1999; Cambier-Langeveld, 2000), due to ‘acentual lengthening’. Abstracting away from the presence of pitch accents and segmental effects on duration, -toe, the last syllable of a phonological phrase, will be longer than tu-, the first syllable of a prosodic word, due to final lengthening. Third, syntactic movement rules may be sensitive to the size of the constituents they manipulate (Inkelas, 1989; Inkelas & Zec, 1990).

In intonation, the prosodic structure may play a role in the context of phonological or phonetic rules (e.g. downstep, which is always confined to some prosodic constituent), in the rhythmic distribution of pitch accents, and in the presence of boundary tones, the topic of this section. One, sometimes two, higher prosodic constituents may come with boundary tones initially and/or finally. These are now reported so frequently that they may well be universal. Some languages appear to have only boundary tones, like Unangan, which has $L_ω$ at the beginning and $H_φ$ at the end of every $φ$, and the only function of intonational structure in this language is thus to signal phrasing (Taff, 1997).

Boundary tones may be complex, i.e. consist of a tone sequence. Jun (1993) gives an analysis of Seoul Korean where every x has a string LH-LH, which is reduced to just an initial LH if the x has only three syllables (Jun, 1998). In the final x, any second occurrence of LH is preempted by one of the $r$-final boundary tones $H_φ$, $L_ω$, $L_ω$, or $H_φ L_ω$, which express different intonational meanings. An example is (19). Formally and functionally, Korean is thus more complex than Unangan, but like Unangan lacks pitch accents.

Bengali combines boundary tone with pitch accents, and closes the $τ$ either by $L_ω$ or by the boundary complexes $L_ω H_φ$, for continuation, and $H_φ L_ω$, for yes-no questions. This is illustrated in (20), where $L^*$ marks the accented syllables. Evidence that $H_φ L_ω$ both end the $τ$ is provided by the contrast with $L^* H_φ L_φ$, which contour is used to mark narrow focus with declarative intonation, as shown in (21), and is realized with a lower peak, as shown in (22). The $F_0$ peak signalling the question is always $τ$-final and high.

These examples also illustrate that more than one prosodic constituent may come with boundary tones, since in addition to $H_φ$, Bengali has $H_φ$. Evidence that
H_φ is a boundary tone is given by the pronunciation of *amar 'my' in (20)–(22). The L* goes to the first syllable of the first lexical word in the φ, and thus skips the function word *amar. Because *amar is pronounced on a downward slope, H_φ cannot be interpreted as the leading tone of the following pitch accent, since this analysis would predict that –mar has high pitch.

For Georgian, Bush (1999) reports a complex boundary L_H, as in (23), which is part of a question intonation contour that begins with H_1. As Bush points out, the L-tone is not a pitch accent, as it spurns the stressed syllable in laparakob. Neither is it a sequence of L_H, one of his arguments is that the end of i may contribute a L, in polite speech which could not be abstracted out of the contour if it was combined with the preceding H-tone. The Dutch dialect of Venlo has boundary tones of i and v (Gussenhoven & van der Vliet, 1999): all i's end in H or L, while the utterance can have an additional H or leading to four utterance-final contours, H, L, L_H, and H_H.

3.5 Secondary association

Pierrehumbert & Beckman (1988) introduced the concept of a boundary tone associating with a Tone Bearing Unit (TBU), the sonorant mora in Japanese. They used it to explain the low target at the beginning of the a in (15) and the first a in (25). A different situation exists when an a's first mora is associated with a H-tone, as in the second a of (25), where a pitch accent occurs on the first syllable. There is still a pitch rise, but it starts from a less low target. In (15b) and the first a in (25), therefore, they assumed that the low boundary tone is associated both with the prosodic constituent (the 'primary association') and with the mora in the first syllable (‘secondary association’), while in the second a of (25) the L-tone is merely associated with the prosodic constituent. This difference in representation is translated by the phonetic implementation rules by a more precise, that is, lower target for the L-tone with the secondary association.

The notation of secondary association was applied by Grice (1995a) to account for the variation between the slight fall on r-final accented syllables in Palermo Italian question intonation, as in (26a), and the deep fall observed when the accented syllable is non-final, as in (26b).
There is no necessary implication, however, that a tonal representation like (26a) inevitably leads to a phonetic implementation whereby the final L is not or is barely pronounced. Grønnum (1991) and Ladd (1996, 133) would describe a phonetic contour like (26a) as a case of ‘truncation’, and distinguish it from ‘compression’, i.e., a contour which reaches low pitch. Languages or language varieties may differ in that one is ‘compressing’ and another ‘truncating’, as shown by Grabe (1998b) for RP and northern standard German, respectively, and by Grabe, Post, Nolan, & Farrar (2000) for Cambridge English and Leeds English, respectively.

While the difference between ‘compression’ and ‘truncation’ might exclusively be accounted for by language-specific implementation rules that are sensitive to the availability of segmental material, confirmation of the moraic association of boundary tones was provided by data from Venlo Dutch. This dialect has a privative tone contrast on stressed syllables with two sonorant moras. There are thus three prosodic types of stressed syllable: those with one sonorant mora (27a), those with two but no lexical tone (27b) (also known as Accent 1), and those with two and H on the second mora (27c) (also known as Accent 2). An intonational pitch accent H* associates to the first mora of the stressed syllable of every focused word, and a declarative L closes the i. The fall for Accent 1 is completed inside the Accent-1 syllable in (27b), which is explained by the secondary association of L with the sonorant mora in the accented syllable, a TBU which requires tone. Neither in (27a) nor in (27c) is there such a ‘free’ mora available, and as a result the falls in the latter two contours are slower (Gussenhoven & van der Vliet, 1999). Gussenhoven (2000a) gives data for all three conditions in the related dialect of Roermond. When H occurs instead of L in situations like (27b), there are clearly two targets, forming a high level stretch.

(27)

3.6 Phonological adjustments
Pierrehumbert (1980) stated that, in English, the tone string arising from the compilation of pitch accent and boundary tones was at the same time the surface representation which offered itself to the phonetic implementation module. Likewise, Stockholm Swedish required no phonological adjustments (assimilation, deletions, insertions) of the underlying tone string. Cases of adjustments in intonational tone strings have since been reported, however. Bengali disallows adjacent like tones (OCP), and violation is solved by tone deletion (Hayes & Lahiri, 1991). Thus, when intonational H* is introduced before H, H* deletes (Lahiri & Fitzpatrick-Cole, 1999). Assimilation of lexical or intonational H to L* occurs in the dialect of Roermond, if H follows L* in the same syllable (Gussenhoven 2000b).

3.7 Phrase accent
The term ‘phrase accent’ has a checkered history. Pierrehumbert (1980) applied it to the internal boundary tone T-, equating it with Bruce’s (1977) ‘sentence accent’, the focal H of Stockholm Swedish. What they have in common is that they occur between the final boundary tone and the last T*(T), an intonational pitch accent in English and a lexical pitch accent in Swedish. Functionally, the Swedish ‘sentence accent’ is equivalent to the intonational pitch accents of English. In Beckman & Pierrehumbert (1986), the ‘phrase accent’ was reanalyzed as a boundary tone of the ip, as noted above, allowing for an analysis in
which lower-ranking ip’s end T- and higher-ranking i’s end in T-T%. Grice, Ladd & Arvaniti (2000) narrowed the meaning down to a boundary tone with secondary association. This analysis explains why the English phrase accent (T-) is not in fact pronounced at the boundary, but usually immediately after the last pitch accent, and they therefore claim that T- associates with a stressed syllable inside the ip. They also suggest that the much-discussed issue of the difference between ‘fall-rise’ and ‘fall-plus-rise’ in descriptions of British English (cf. Cruttenden, 1997) can be explained in terms of ‘phrase accents’. While the dependence of T- on stress may be less obvious in German of English, a comparison between Standard Greek and Cypriot Greek makes the difference between an edge-seeking tone and a stress-seeking ‘phrase accent’ clear. Both languages have a yes/no question contour, analyzed as L* H-L%. In Cypriot Greek, the H- remains close to the i-boundary, with the peak usually falling in the last syllable, while in Standard Greek its target falls in the rightmost stressed syllable. In (29) for instance, H- associates with the unaccented, but main stressed syllable L*- in Standard Greek, while L% possibly does too, or else is timed after H-. (In (29), I place the L-target where it would appear to occur in the F0 track.) The Cypriot Greek case is like that of Hungarian (Ladd, 1983b; Gösy & Terken, 1994; Varga, 2002) and Bengali (Hayes & Lahiri, 1991), and shown in (30), where the rightmost stress is [zo], but the F0 peak is on the unstressed final [mu].

(29)

\[
\{ \text{tin' Elena} | \text{Lenosi} \} \\
L^* \quad H-L%
\]

the Elena Lenosi

‘[Are you asking about] ELENA Lenosi?’

(30)

\[
\{ \text{e'faete pu to ri'zogalo mu} \} \\
L^* \quad H-L%
\]

eat.2PL.PAST of the rice-pudding mine

‘Have you eaten of my rice pudding?’

A third possibility is for the H-tone to have two high targets forming a plateau on post-focal stretches of speech, as in Roermond Dutch (Gussenhoven, 2000a) and Transylvanian Romanian (Grice, Ladd, & Arvaniti, 2000). In Bern Swiss German, a similar plateau is due to H of the L*-H pitch accent, which in broad focus continues until the last stressed syllable of the phrase, where a L, seeks a secondary association (Fitzpatrick-Cole, 1999). This is shown in (31).

(31)

\[
\{ \text{[ Er het e Politiker ermordet ]} \} \\
L^*+H \quad L_0 \quad L_u
\]

4. Conclusion

While this brief survey may give one a feel for the cross-linguistic variation in intonational structure, the collection of languages of which fairly complete descriptions are available is still small. Future research is likely to bring many more tone systems to light couched in the autosegmental-metrical framework, with and without lexical tones. Many of these descriptions will have a practical bias in the sense that in addition to claiming to provide a tonal grammar, they are usable as transcription systems of spontaneous speech (cf. Jun (forthcoming)). More grammatically oriented accounts are likely to include a description of how information status (‘focus’) is expressed prosodically. Three aspects of the prosodic structure have been reported to be used for this purpose. First, there may be a requirement to align the beginning or end of the focus constituent with a prosodic constituent of a given rank, as happens in Japanese, Chichewa, and Bengali (Pierrehumbert & Beckman, 1988; Kanerva, 1989; Hayes & Lahiri, 1991). In each of these cases, there is a different effect on the phonetics and phonology of the expression: the suspension of downstep (Japanese), the distribution of lexical tones and of vowel quantity (Chichewa), and the location of a boundary tone (Bengali, cf. (20)–(22)). Second, there may be a special pitch accent signalling narrow focus, as occurs in Bengali and European Portuguese (Frota, 1998). Third, postfocal words may be de-accented, as is most notably the case in West Germanic languages, but also in many other languages (cf. Ladd, 1996). Further research will undoubtedly refine this picture, as exemplified by Wolof, which lacks prosodic marking of focus and relies on verb morphology for this purpose (Rialland & Robert, 2001).

Acknowledgments

I am grateful to Martine Grice for her comments on an earlier version.
A Phonology of Intonation Bibliography


