

**Alignment of tonal targets: 30 years on
Bob Ladd
University of Edinburgh**

Text of a talk presented in Lund on 10 January 2007, at a symposium in Honour of Gösta Bruce's 60th birthday. References are missing and some details of the works discussed are presented more informally than would be appropriate in print. Please keep this in mind if citing this talk.

I'm pleased and honoured to be asked to present the first talk at Gösta's 60th birthday symposium. Our careers have been loosely intertwined more or less from the very beginning. To start with, we were both born within the first weeks of 1947, albeit three or four thousand miles apart, and unbeknownst to each other. During the 1970s, still three or four thousand miles apart and still unbeknownst to each other, we were simultaneously working on our doctoral theses on prosody and pitch phonology: Gösta's appeared in 1977, mine in 1978. I became aware of Gösta's work in 1980, when it was cited in Pierrehumbert's thesis. We finally met in person in 1982, after I moved to this side of the Atlantic, and we've been crossing paths regularly ever since. We keep showing up on the same platform at workshops and conferences, and I have a number of pleasant recollections of the hospitality of Gösta and his family. During the 1980s Eva Gårding even said – more than once – that we resembled one another physically. In any case, as we both head past 60, I look forward to having more time and more opportunities to see Gösta in what I sincerely hope is a long and productive old fogey phase of our careers. I use “old fogey” in the Peter Ladefoged sense, of course.

While it may be appropriate to indulge in some brief reminiscences on the occasion of Gösta's birthday symposium, it would not do justice to his scholarly importance to devote my whole talk to birthday nostalgia. Gösta's ideas have had an enormous influence on mine and on those of many others, and in my opinion they have not yet exhausted their richness. What I want to do today is discuss one of the central insights in his thesis, and show that in some ways this insight has yet to have the long-term effects on our thinking about sentence prosody – and indeed, about the relation between phonology and phonetics – that I think it will eventually have.

Now, the official theme of today's symposium, and even the official evaluation of Gösta's work found on the symposium web site, is based on one version of why his thesis was “groundbreaking”. The official version emphasises his insight about intonational phonology and intonational pragmatics in connection with the very specific question of how to analyse Swedish word accents in sentence perspective:

His insight that intonational contours in Swedish could be broken down into different tonal components: word accents, sentence accent (associated with focus) and terminal juncture (boundary tones) which realize different combinations of two phonological level tones H and L was a seminal contribution to our understanding of intonational patterning that was subsequently applied to many other languages.

In this “official” assessment of why Gösta’s work is important, the emphasis is on two things: the **linear analysis of pitch contours into pragmatically and grammatically distinct types of elements**, and on the idea that in many languages **the most appropriate phonological description of pitch level can be expressed in terms of local maxima and minima** – the Highs and Lows – not 3 or 4 or 7 distinctive levels.

There’s no question that these ideas have been extremely important. They form one of the central tenets of what I’ve called the autosegmental-metrical theory of prosodic structure. Obviously, the name most prominently associated with the autosegmental-metrical theory is Janet Pierrehumbert’s, not Gösta’s, but Pierrehumbert’s thesis, important as it is, was very much a question of “standing on the shoulders of giants” – and one of those giants was Gösta. Specifically, Pierrehumbert’s thesis draws together three key ideas and weaves them into a coherent whole that has dominated research on intonation ever since. Those ideas are: the notion of “pitch accent” from Dwight Bolinger, the notion of metrical structure from Mark Liberman, and the notions of the phrase accent and of two-level pitch phonology from Gösta. So I really haven’t got any basis for quarrelling with the official version of why Gösta’s thesis is groundbreaking and important – phrase accents and two-level pitch phonology are now a central part of the way we think about intonation, and they were first clearly articulated by Gösta in his thesis.

Nevertheless, in the rest of my talk I want to focus on another less widely appreciated feature of Gösta’s originality. Let’s begin with a diagram that I’m sure is familiar to most of you [Figure 1]: Gösta’s diagram of the pitch contours on Accent 1 and Accent 2 words, broken down into word accent fall, phrase accent or sentence accent rise, and terminal fall. That division into word-accent, sentence accent, and terminal fall is the groundbreaking feature of Gösta’s analysis that is highlighted on the symposium web site. But what’s not highlighted on the symposium web site is that, given this analysis, one of Gösta’s central claims was that the distinction between Accent 1 and Accent 2 resides primarily in the timing of the word accent fall relative to the lexically stressed syllable : in Accent 1 this word-accent fall occurs earlier than in Accent 2 . This is said to be valid across all sentence contexts and across all dialects that have the tonal distinction – again, I’m sure this is familiar to most of you. What I want to emphasise is that with his characterisation of what is invariant about the distinction between Accent 1 and Accent 2, Gösta also introduces a fundamentally new way of thinking about the phonetic description of linguistic pitch. Instead of looking at individual accented syllables and describing the pitch patterns that span them, Gösta’s analysis identifies linguistically significant pitch events in terms of local minima and maxima that can be defined independently of syllables. It then describes linguistic distinctions in terms of the temporal alignment of the local minima and maxima with specific syllables – such as the accented syllables.

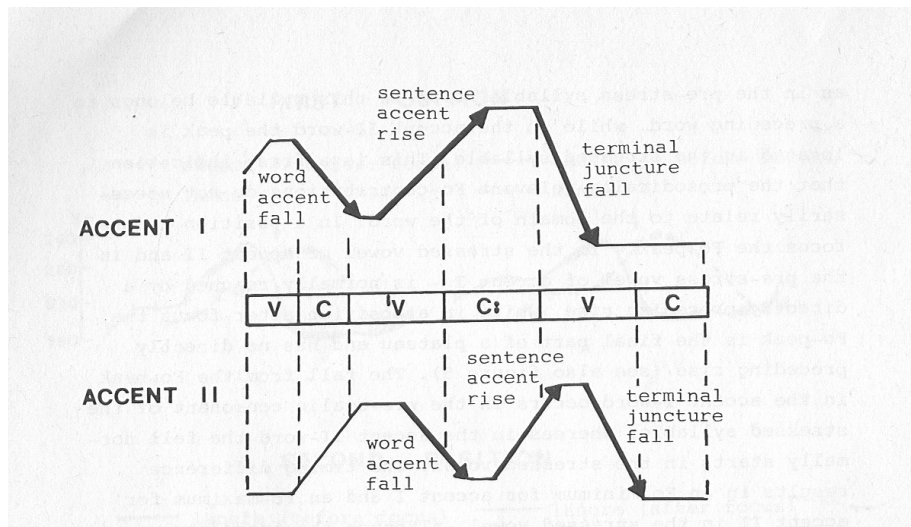


Figure 1 (from Bruce 1977, Fig. 10)

If we just look at the accented syllable contours, they look different, but they look different in ways that are not consistent from one context to another or from one dialect to another. But if we look at the whole contour independently of the syllables and if we assume that the temporal coordination of the contour and the syllables is what is important, we see that the contours are essentially identical in shape, and differ only in alignment. Whether a specific accented syllable has a rise or a fall is not the key property we are interested in, but a consequence of the key property. Specifically, it depends on how the syllable is aligned with the Hs and Ls of the pitch contour. As Gösta succinctly put it: **“reaching a certain pitch level at a particular point in time is the important thing, not the movement (rise or fall) itself”** (1977: 132).

So I think we need to add to the list of Gösta’s key contributions. Not only did his thesis clarify some crucial aspects of the structure of intonation contours and the function of their various component parts, but it also showed us something fundamentally new about how to describe the phonetics of pitch contours, at least in languages like those of Europe. The key ideas are (1) that the key elements of pitch contours need to be identified independently of syllable boundaries, and (2) that one of the important phonetic dimensions of these independently identified elements is the way in which they are aligned in time with syllables and other elements of the segmental string. If we look at what early and mid-20th century phoneticians and dialectologists said about the Scandinavian accents in the light of what we now know from Gösta’s work, we can readily see the value of this way of looking at pitch phonetics.

Scholars had puzzled for decades over the question of what makes Accent 1 Accent 1 and Accent 2 Accent 2. From one dialect to another, from one sentence context to another, even in words with different numbers of syllables, the phonetic manifestations of the word-accent distinction seemed bewilderingly diverse, and older descriptions are full of qualified generalisations, and approximations that work only most of the time. For example, here are some quotes from Kerstin Hadding’s 1961 monograph:

“[Malmberg] draws the conclusion that there is a relevant opposition between a pronounced fall in the first syllable of Accent 1 and a slight rise (or, sometimes, level pitch) in the first syllable of Accent 2.” (Hadding 1961, p. 64)

“...some stresses with Accent 2 [coded as exhibiting rises] may end in rather marked falls” (ibid. p. 66)

“Among the 329 monosyllables [in the corpus] 165 were falling ... , 115 rising ... , 15 level, and 34 ‘crescent’-shaped... .. It would seem reasonable to question whether monosyllables and disyllabic words with Accent 1 should, as is usually done, be classed together as having ‘acute’ tonal accent ...” (ibid. p. 66)

With hindsight, we can see that the key flaw in these descriptions just quoted is that they concentrate on **stretches of pitch contour defined by the limits of the accented syllable**. This is exactly what Gösta’s analysis doesn’t do.

I don’t mean to suggest that the idea of looking at pitch contours in terms of the relative alignment of independently identified pitch points emerged fully-formed from Gösta’s brow. He also stood on the shoulders of giants. In particular, in their 1953 paper “Tone and Intonation in East Norwegian” Haugen & Joos almost freed themselves from concentrating on the accented syllable, and almost saw what Gösta made explicit 25 years later. It is worth quoting them at some length :

“... the movement [of pitch] is everywhere continuous, with an up-and-down alternation ... It appears that if one did not know (by auditory means) where the stresses are located, it would not be possible to detect the characteristic word tones. If we compare the tonal movement of [two specific words from their corpus], we find that the first two syllables of each have almost identical appearance ... Yet we know that the first has accent 1 on the second syllable, while the second has accent 2 on the first ... Wherever we have an accent 1, its stress falls near the low point of the curve; in accent 2, the stress comes earlier, and usually includes the preceding high point, while the low point follows the main stress. ... *The melody is not in itself distinctive, but acquires distinctive value when it is associated with stress in a particular way.*” (Haugen & Joos 1953 (1972): 425f, emphasis added).

However, Haugen & Joos continued to devote considerable attention to describing overall contours for the two accent types and to the interaction of word accents with expressive intonation. Their suggestion that the pitch contour “acquires distinctive value when it is associated with stress in a particular way” anticipates one of Gösta’s key claims, but Haugen and Joos don’t seem to have appreciated its potential for radically reshaping the way we think about the phonetics of pitch. Only once this idea was combined with Gösta’s structural analysis of sentence contours did it begin to have a wider influence.

Let me summarise what I see as the implications of Gösta’s idea for the phonetic description of pitch. If we want to capture the linguistically

significant phonetic parameters of pitch in our description, we must avoid a syllable-by-syllable segmentation of the overall pitch contour, and avoid talking about the rises and falls of pitch that happen to result from such a segmentation. Instead, we must identify locally significant points in the pitch contour – these are often local minima and maxima, the Highs and Lows of Gösta’s phonological description. Then – and only then – we must describe the way in which those locally significant points are coordinated in time with the phonetic events of the segmental string. Essentially, the implication of Gösta’s approach is that a useful phonetic description of pitch can be expressed in terms of two principal dimensions or parameters: **the “scaling” or F0 level of the linguistically distinctive pitch points, and their alignment or temporal coordination relative to landmarks in the segmental string.**

What I’m suggesting, in other words, is that scaling and alignment are the “right” descriptive dimensions for talking about linguistic pitch, right in the sense that they are the ones that give us insight into the phenomena. In the case of Accent 1 and Accent 2, this claim seems incontrovertible – if we consider the pitch contour on the accented syllable we can make only rough and internally contradictory generalisations about what characterises the two word accents, but if we consider the alignment of pitch points relative to the accented syllable we see the regularity across the whole Scandinavian system. This is what made Gösta’s thesis an instant must-read in Nordic prosody circles. It’s also what drew the attention of various investigators of other languages – like Klaus Kohler on German and Janet Pierrehumbert on English – to the existence of intonational distinctions based mainly or entirely of differences of alignment.

But the alignment perspective is enlightening in ways that go well beyond the concerns of intonational phonology or Scandinavian dialectology. For example, Gösta himself, in his thesis, discussed certain ways in which pitch contours can be modified by time pressure. These modifications are completely unsurprising once we accept the fundamental assumption that we are describing the alignment of pitch points in time. So for example, if pitch movements are specified on adjacent syllables in the phonology but for reasons of motor control are “too close together” for the phonetics, the result is often that one of the pitch targets is undershot. This is completely familiar from segmental phonetics, where we might expect the formant values in a rapid [iaiaia] sequence to be undershot.

More recently, the alignment perspective has led to new and more surprising discoveries. For example, work by myself and my colleagues as well as several other groups of researchers has shown that in a number of languages, what is invariant about a given linguistically significant pitch feature may reside in the way it is aligned with the segmental string. Specifically, we’ve identified the phenomenon of “segmental anchoring”, in which the alignment of specific pitch points relative to specific features in the segmental string remains roughly invariant, while the slope and duration of the pitch changes varies. This first came to light in the study on Greek that Amalia Arvaniti did with me and Ineke Mennen. What we found was that prenuclear declarative pitch accents show a rise in pitch that begins simultaneously with the

beginning of the accented syllable and ends 10 or 20 ms after the beginning of the **following unstressed vowel**. You might think that if the syllable composition changes so that the following unstressed vowel is farther away from the beginning of the stressed syllable, the alignment might change, while the duration and slope of the rise would remain constant. But that's not what happens: the local maximum at the end of the rise continues to be aligned a few ms into the following unstressed vowel, and what gets adjusted are the slope and duration of the rise. This is easy to describe if we base our description on significant pitch points like the local minimum and maximum; it's a lot harder to make sense of – or even notice – if we're basing our description on the pitch pattern of individual syllables in sequence.

Building on this finding, we've also shown that if you compare apparently identical pitch features in different languages or language varieties, you find that they may differ subtly in the way they are aligned with the segmental string. For example, if you look at similar rising prenuclear accents in similar contexts in Greek, English, and two varieties of German, you find that there are slight differences in the alignment of the beginning and the end of the rise. These differences are summarised on this slide [Figure 2]. If we describe these pitch features phonetically in terms of their alignment rather than in terms of, say, the shape of the stressed syllable contour, we have a simple way of making phonetic comparisons across languages and language varieties. Here once again the alignment perspective on pitch phonetics puts us on familiar territory from segmental phonetics. For example, the cross-language comparison of phonetic details of alignment is similar to the cross-language comparison of vowel formant spaces or voice onset time.

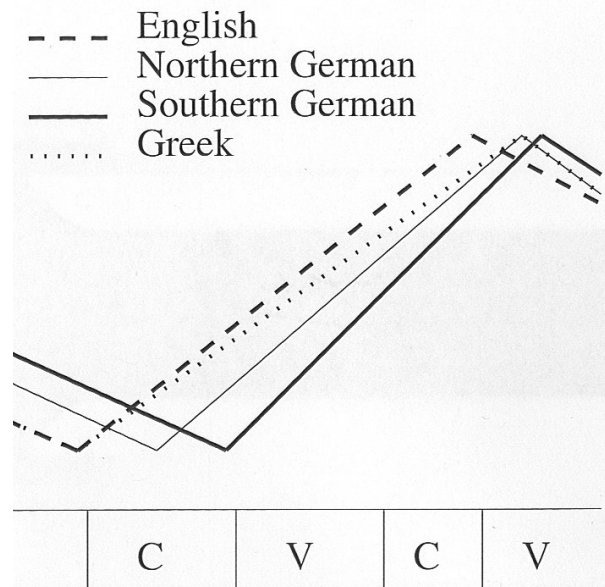


Figure 2. From Atterer & Ladd 2004, Fig. 2

Still, the alignment perspective on the description of pitch phonetics keeps throwing up surprises. In an earlier study of English nuclear accents, my colleagues and I accidentally discovered that the peak of a nuclear accent in short English sentences was aligned very slightly earlier in sentences with two pitch accents like *Her father's a miner* that

in sentences with only one pitch accent like *He's a miner*. The “time pressure” explanation invoked by Gösta to explain the behaviour of adjacent pitch movements doesn't seem to apply, because if anything we might expect that the presence of an extra pitch accent at the beginning of the sentence would push the peak on the nuclear accent slightly **later**, not attract it earlier. Nevertheless, Caterina Petrone and I have recently reproduced this effect in a more controlled experiment in Italian. We created a set of sentences in which the sentence length was systematically manipulated. We tried to distinguish mere length, expressed in terms of the number of syllables preceding the nucleus, from the presence or absence of a prenuclear accent, though we weren't completely successful in doing that, and in any case our results are still preliminary and messy. But what's clear is that there is some such effect, and that to at least some extent it depends simply on the absolute length of the sentence: the longer the sentence, the earlier the nuclear peak. Here [Figure 3] you can see results from one speaker. A dependence on mere length makes a little more sense in terms of time pressure, because in some sense you've got more room for the whole contour, but the explanation is still not very obvious, and we hope to pursue this line of investigation soon.

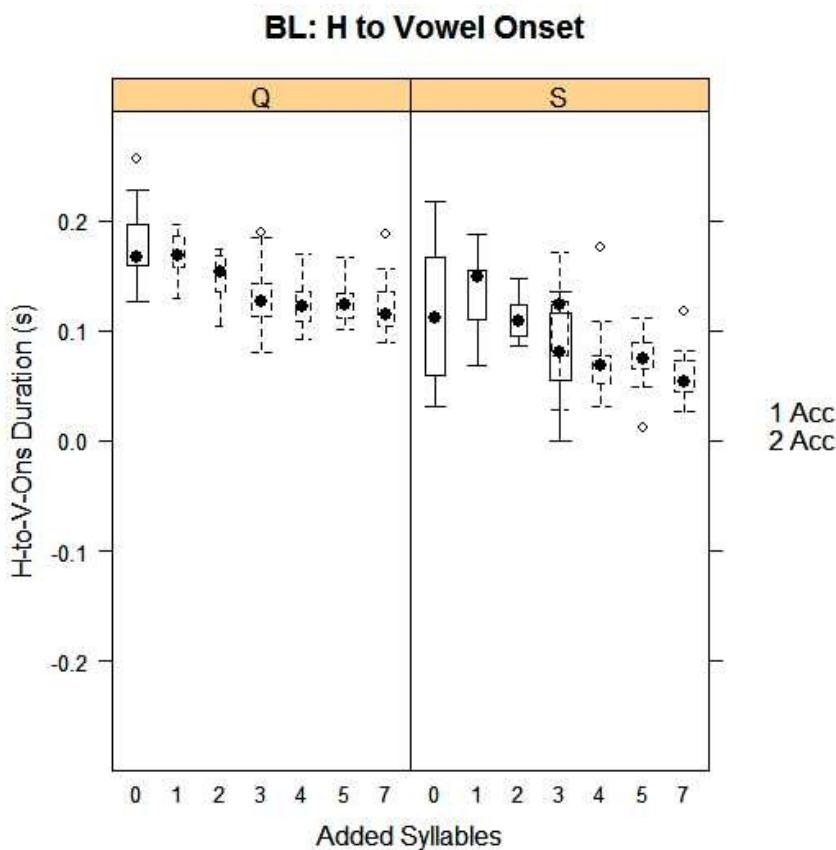


Figure 3. From Petrone & Ladd (in preparation)

Let me just emphasise that none of this would have come to light if we hadn't been measuring alignment. More generally, it seems clear from all these examples that by looking at pitch phonetics in terms of the alignment of pitch points relative to the segmental string we are discovering new phonetic phenomena and helping to make sense of the

intonational phonology of specific languages. Now, this doesn't mean that we now understand everything about this general topic and the way ahead is clear. In fact, a number of people have picked up the idea of alignment and segmental anchoring in ways that I think may be somewhat problematical – some people have proposed what to me are excessively fine notational distinctions based on alignment, while others have contended that the whole idea of alignment focuses on quantitative phonetic detail to the exclusion of insightful phonological analysis. Because these various proposals revolve around the difficult theoretical issue of the relation between phonetics and phonology, I'd like to spend a little time talking about what I've described as the difference between **association** and **alignment**.

I first proposed the idea of such a distinction in the early 1980s, based directly on my reading of Gösta's work. Association is phonological and abstract: a particular pitch feature (say, a H tone or local maximum) can be associated with a particular segmental feature (say, a stressed syllable) in the sense that the location of one depends on the location of the other. Association is also categorical and works in terms of discrete phonological elements: a particular pitch feature may be associated with a stressed syllable or some other phonologically defined tone-bearing unit, but not with the middle of the stressed syllable, or ten milliseconds before its end. Alignment, on the other hand, is phonetic and concrete: the H tone or local maximum associated with a stressed syllable may occur consistently in the following unstressed syllable. Alignment is also continuously variable and measurable: a pitch peak that in one context is consistently aligned on average 10 ms before the end of a syllable may in another context be aligned on average 20 ms after it. In my opinion, distinguishing association from alignment in this way makes pitch phonetics and phonology well behaved instead of mysterious. Specifically, it makes it possible to apply long-standing and well-understood ideas from segmental phonology to the description of intonation, and gives us a clear basis for understanding phonological categories and their varying manifestations in the signal. I'll give three examples.

First, distinguishing association from alignment allows us to talk about phonological distinctions within a given syllable that are realised in phonetic properties of some other syllable. This is clearly the case with Accent 1 and Accent 2 in many contexts: the accent, as a phonological feature, is on one specific syllable, but the phonetic cues that allow us to identify it may be primarily part of the realisation of some other syllable. This may seem like unjustifiably unconstrained abstraction, but analogous abstractions are completely familiar in segmental phonetics and phonology. For example, nobody would raise an eyebrow at the statement that the main perceptual cue to the voicing distinction in English final stops in pairs like *bat* and *bad* is actually the duration of the preceding vowel.

Similarly, the fact that alignment is physical and measurable while association is categorical and abstract should not be a cause for theoretical or methodological concern. In segmental phonology and phonetics it is absolutely uncontroversial to posit some categorical distinction in the phonology even though we know that the physical

correlate of the distinction involves a phonetic continuum: for example, we know that voice onset time and the vowel formant space involve physically continuous, quantifiable and measurable dimensions, but this does not in any way prevent us from operating with categorical distinctions of voicing and vowel quality in the phonology. On the contrary, understanding voice onset time and the vowel formant space has given us a new understanding of what information is present in the speech signal, and if anything it gives us a new respect for the human perceptual and speech processing systems that convert this continuous mess into the discrete and orderly strings that we apprehend as language.

Finally, cross-dialect and cross-language comparison is enhanced by understanding alignment as a physical continuum, and by the assumption that different languages or different language varieties can divide up physical continua in different ways by. Bruce & Gårding give the example of the Stockholm and Göteborg varieties of Swedish, which differ in overall alignment in much the same way as Northern and Southern German discussed earlier. Overall, Göteborg aligns its pitch movements later than Stockholm – but with each variety, Accent 2 is aligned later than Accent 1. This means, among other things, that the Accent 2 fall in Stockholm is aligned about the same way as the Accent 1 fall in Göteborg. The simplicity of this description does not provide proof that it is correct, of course, but it does provide good reason to think that it is superior to one based on impressionistic categories of rising and falling and crescent-shaped accented syllables. And again, when we compare this case to segmental phonetic differences between languages and language varieties we see that we are on familiar theoretical ground. For example, many languages have a two-way stop voicing contrast which is cued primarily by voice onset time in syllable initial position, but, as is well known, the details are such that one language's voiced stop can be similar to another's voiceless stop. Describing the phonetics of stop voicing in terms of VOT provides a clear and precise basis for explaining cross-language and cross-variety confusions and misperceptions. In the same way, describing cross-language and cross-variety differences of pitch contour in terms of alignment gives us a basis for predicting confusions and misperceptions of the same kind.

This example brings me to a final point that I think it's important to emphasise. It's important to emphasise it because Gösta's original description wasn't really fully clear about this, and because my original distinction between association and alignment wasn't fully clear about it, and because this lack of clarity may also be the source of the idea that the study of alignment is all about quantitative phonetic detail and nothing more. The clarification is this: I am absolutely not saying that the **location** of a Swedish word accent on a specific syllable is "association" and hence phonology, while the **distinction** between Accent 1 and Accent 2 is "alignment", and hence phonetics. Rather, what I am saying is that there are two **phonological** categories, Accent 1 and Accent 2, which are primarily distinguished by the **phonetic** dimension of alignment, just as in, say, English and Italian there are two **phonological** categories of stops, voiced and voiceless, which are primarily distinguished in syllable onsets by the **phonetic** dimension of VOT. The **quantitative** details of alignment, like the **quantitative** details of VOT,

are irrelevant to our phonology, and for most purposes to our phonetic notation. The only phonetic fact that counts is the **relative** alignment or **relative** VOT.

So if we're describing English or Italian, we can appropriately use the symbols /b/ and /p/ to indicate the two members of the labial stop voiced-voiceless pair, even though the phonetic manifestations of Italian /p/ sometimes overlap with the phonetic manifestations of English /b/. The only phonetic fact we are committing ourselves to in both cases is that, in both languages, the VOT of /p/ is later than that of /b/. Similarly, in describing Swedish, we can appropriately use the labels Accent 1 and Accent 2 – or, if you prefer, H+L* and H*+L – to indicate the two members of the word accent distinction, even though the phonetic manifestations of Accent 1 in one dialect sometimes overlap with the phonetic manifestations of Accent 2 in another dialect. The only phonetic fact we're committing ourselves to in both cases is that, in both dialects, the alignment of the pitch fall relative to the stressed syllable for Accent 2 is later than that for Accent 1. I especially want to emphasise that this is true of the ToBI-style notations H+L* and H*+L: these do **not** need to imply any specific alignment details on their own, any more than /b/ or /p/ imply a specific VOT. All they imply is that H*+L is aligned later than H+L*. In fact, this was one of the main points of the Atterer and Ladd paper to which Kohler so objected: we wanted to caution against using notations like H*+L and H+L* to convey some specific phonetic interpretation independently of the context of a phonological distinction. Phonetic interpretations can best be described in quantitative terms; notations for phonological distinctions are, ultimately, arbitrary.

To sum up: what I've tried to do today is to celebrate Gösta's idea that pitch contours are best described phonetically in terms of identifiable pitch points that are aligned and scaled in specified ways. This is a contribution to phonetics, related to but clearly distinct from his structural and functional insights about Swedish intonation contours. A syllable-by-syllable segmentation of the pitch contour of an utterance may seem like the most natural and neutral point of departure for phonological analysis, but, at least in the languages of Europe, it is not. It obscures the true regularities and makes for puzzling descriptive paradoxes like the pitch contour on Swedish monosyllables. If many of us are now busy describing the functional distinctions and the fine phonetic detail of European intonation systems in terms of the scaling and alignment of pitch targets, much of the credit goes to Gösta.

Acknowledgements

My collaborators Amalia Arvaniti, Michaela Atterer, Ineke Mennen, Caterina Petrone, Astrid Schepman, Rebekah Stackhouse, and Laurence White, who are partly responsible for some of the results I summarise here.

The UK Economic and Social Research Council and the British Academy, which funded some of the research reported here.

Carlo Ladd, for expert assistance with the PowerPoint® presentation.