LOCAL SHAPES AND GLOBAL TRENDS

Mary E. Beckman
Department of Linguistics, Ohio State University, Columbus, OH, USA
and ATR Interpreting Telecommunications Research Laboratories, Kyoto, Japan

ABSTRACT

The topic of this symposium assumes answers to two more basic questions: What is intonation? How should we understand phonological structure in general? Several phenomena basic to the debate over superpositionality (e.g., variation in overall pitch range and accent realization in the context of variation in prominence and downtrend) are reviewed in the context of these questions.

INITIAL DEFINITIONS

It will be difficult to discuss whether the structure of intonation is linear or superpositional unless we lay out our background assumptions on two more general questions: *What do we mean by "structure"? What is "intonation"?* So let me begin here. I think we all agree that the "structure" is phonological or phonetic. We are talking about the structure of **sound**, and not syntactic structure or discourse topic structure or any of the other kinds of structure relevant to understanding the sound structure of intonation. On "intonation", though, we are in less agreement. Where Grønnum [1] proposes to exclude, for example, microprosody, my prejudice is to make the definition as inclusive as possible. In modeling intonation, I want to understand **all** aspects of the perceived pitch pattern that the speaker intends for the hearer to use in understanding the utterance, or that the hearer does use whether intentionally controlled by the speaker or not.

ON REPRESENTATION

With these definitions, then, the topic of this symposium is: *What kinds of structures should we use to represent the speaker's and hearer's knowledge of these relevant aspects of the pitch pattern?* In other words, the question becomes one of the representation of knowledge about sound. And here again, I think we will find points of agreement and points of disagreement among the participants in this symposium.

One salient point of agreement is that we have chosen as our primary phonetic representation that most convenient measure, fundamental frequency. All of us on this panel have worked almost exclusively with this representation — because by comparison to perceptual or physiological records, F0 is extremely easy to get. Also, (although we explicitly invoke this reason less often) the F0 computed from a complex harmonic signal is known to be psychoacoustically close to the pitch perceived for it.

Note that in the above restatement of the title of this symposium I have deliberately rephrased it in the plural. I have said "kind<u>s</u> of structure<u>s</u>" rather than "<u>the</u> structure", because I am prejudiced to think that useful representation is not monolithic. Every representation is a model, a hypothesis about some aspect of the speaker/hearer's competence. There are many different dimensions along which we can study speech, and a representation that is useful for modeling competence in one dimension is not necessarily useful for any other; nor does it necessarily map easily to useful representations for other dimensions. For example, there is no representation of tongue shape that is useful for modeling the speaker's control of vowel constrictions and which also maps simply onto any useful representation of the hearer's knowledge of the resulting variations in timbre.

To justify this prejudice specifically for intonation, let me entertain a proposal that is not as silly as it may sound. We all use the F0 contour as a convenient phonetic representation. But suppose we were to take it not as <u>a</u> representation, but as <u>the</u> representation of intonation. Then the answer to our debate would be trivial: the structure of intonation is linear, because it is the linear sequence of F0 values calculated for the succession of sampling intervals in the utterance. I think the other participants in this symposium would join me in quickly rejecting this trivial solution. And this immediate response would be at least in part because of a set of phenomena that I think we all would agree requires a superpositional representation, involving what I will call "overall pitch range".

ON PITCH RANGE

There are several kinds of variation that fit into this set. First, different speakers have different ranges of F0 values that they can produce comfortably. This is largely an artifact of laryngeal anatomy. However, hearers do identify a common intonation pattern in the sequence of relatively low F0 values on the phrase *Thank you* in an adult male speaker's production of *Marco, say "Thank you."* and in the sequence of much higher F0 values when 4-year old Marco obeys his father's injunction. So hearers clearly can factor out the speaker differences in parsing the intonation pattern — just as they normalize to different vocal tracts in perceiving the timbre of vowels produced by different speakers — and all of our theories of intonational structure include at least an implicit representation of the speaker's overall pitch range in our models of the hearer's competence. (I would argue for including it in our models of the speaker's control as well, since any phonetically competent speaker knows how to speak in a higher than comfortable pitch range in order to sound like a younger cuter speaker, or in a lower than comfortable pitch range to sound like an older more authoritative speaker.)

A second kind of variation in overall pitch range involves global variation in vocal effort. Unless trained to do otherwise, speakers use higher F0s when speaking up to be heard over distance or above ambient noise. Two recent studies suggest that this increase is caused by the higher subglottal pressure of the louder voice [2, 3], so we may not want to include it in the same way as other aspects of F0 variation in our models of the speaker's control. However, it has been convenient to model the hearer's competence at parsing it using the same representation as for inter-speaker variation (e.g. [4]).

A third seemingly related but functionally different phenomenon is variation in pitch range to reflect different emotional states. For example, a happy voice might involve an upward shift of the pitch range, whereas an angry voice involves a "pressed" voice quality often associated with a lower pitch range. Trained actors can produce this sort of variation at will in order to simulate emotional states that they do not feel, but comparisons between productions of "aprosodic" patients and of untrained controls (e.g. [5]) suggest that this professional ability is a honing of a skill within any normal speaker's competence.

A final related variation is the use of increased overall pitch range to signal greater emotional involvement with the content of an utterance. The expanded pitch range seems to be associated typically with a louder voice as well, but perhaps context helps hearers distinguish this use from increased vocal effort to project above ambient noise. This is an aspect of pitch range variation that can be quasi-conventionalized to distinguish alternative interpretations of some intonation patterns. For example, there is an intonation pattern in Korean that involves a sharp pitch rise localized to the last syllable in the utterance (i.e. a H% edge tone), which can be interpreted either as a straightforward yes-no question or as an incredulous echo question. Jun & Oh [6] have shown that the expanded pitch range of emotional involvement is a strong cue biasing the hearer toward the latter interpretation. Since Korean morphology does not distinguish WH-question words from corresponding indefinite pronouns, the expanded pitch range can be the salient cue distinguishing the two grammatical categories in utterances with this tune. Thus, the sentence [nuka wajo] produced with a H% and in a neutral pitch range might be interpreted as 'Is anyone coming?' but in an expanded pitch range as 'WHO did you say is coming?!' A relatively nonlocal increase in F0 also distinguishes the literal question interpretation from the rhetorical question interpretation of the analogous tone pattern in Osaka Japanese [7], the "incredulity" from the "uncertainty" reading of the scooped rise-fall-rise pattern in English [8], and the incredulous rhetorical question interpretation from the declarative interpretation of utterances ending in a L% boundary tone in Kipare [9]. In all these cases, it seems possible to model the phenomenon as an expansion of pitch range over the whole utterance — in other words, as affecting the same dimension of intonational structure that we posit to account for how hearers accommodate to different speakers and to the effects on F0 of increased overall vocal effort in background noise.

I think even the most radically linear of us would agree that we should model the way in which this dimension interacts with other dimensions of the hearer's competence by adopting a superpositional representation of pitch range. We model the hearer as deriving some kind of backdrop graph paper abstracted away from the actual pitch contour in order to represent more local events (such as the high- versus mid- versus low-level lexical tones of Taiwanese or the components contrasting different pitch accents in English) as invariant across productions by different speakers or by the same speaker in different "voices". Our points of disagreement, rather, involve the ways in which the clearly local events interact with kinds of variability other than variation in overall pitch range. In particular, we have published rather divergent ideas about the best way to represent certain patterns of variability associated with such functions as signaling local prominence or discourse topic structure. Before I discuss these phenomena, however, I must admit to one last prejudice about phonological structure in general.

ON ABSTRACTION

In the preceding section, I used the notion "abstraction" to describe what the hearer does in factoring out certain effects of change in speaker, in vocal effort, and in type and degree of emotional involvement in the utterance. The hearer abstracts away from the actual pitches, a representation of the utterance's overall pitch range. Degree of abstraction has sometimes been equated with different levels in a model of grammar whereby different rule modules apply in sequence to generate from the speaker's intentions some sound signal. That is, a more abstract representation has been taken to mean something further upstream in the generative process, and phonological representation, in particular, has been contrasted with phonetic representation as being more abstract. By this usage, the representation of overall pitch range abstracted away from the F0 contour might be considered "phonological".

I am not sure enough of how the grammars of real speakers and hearers work to say confidently what is the difference between "phonological" and "phonetic" representation, but I am convinced that abstractness is irrelevant. The representation at the heart of many applications of the acoustic theory of speech production — that of the vocal tract as a lossless uniform tube — is an extremely abstract idealization, but who would argue that this is nearer to the speaker's intent than a representation of a lossy bent tube that better approximates the physical reality? To be sure, by abstracting away from the F0 contour to a graph-paper representation of overall pitch range, I think we do get closer to the phonological in the sense that we are making a plausible hypothesis about what it is the hearer does to parse the signal for the speaker's intent to produce phonologically contrasting tone levels. However, I do not think this makes the representation of pitch range itself phonological, because the dimensions of the "graph paper" are still continuously variable. I think it is useful for other reasons (e.g., for understanding how sociolinguistic variation develops into sound change) to say that any such continuously variable representation derived from a physical dimension is "phonetic" and only becomes "phonological" when it is reanalyzed in terms of arbitrarily discrete categories. Therefore, I would like to reserve "phonological" to refer to categories of conventional paradigmatic contrast that arbitrarily discretize relations of difference along some (possibly very abstract) phonetic dimension, and to refer to discrete categories of metrical organization derived from syntagmatic constraints on the occurrence of these paradigmatically contrasting categories. In short, "phonological" for me means the symbolic representation of such paradigmatic content categories as [+Hightone] (which in the pitch accent system of English is in opposition to the category [-Hightone] — see below), and of such syntagmatic categories as "stress foot" (which in English describes a metrical structure constraining where pitch accents can occur in the alignment between intonational melody and segmental string).

I think the other participants in this symposium would agree with me in calling the graph paper representation of pitch range "phonetic" rather than "phonological". However, they may not all agree with me on my reasons for this, and so it is worth belaboring the point by giving one more illustration of my prejudice about abstractness. The example involves the representation of microprosody — the so-called "intrinsic" F0 variation associated with voiced versus voiceless obstruents. My reading of the literature on phonological development convinces me that all abstraction comes with a cost, and that new speaker/hearers acquiring a language for the first time will not abstract away a symbolic representation of knowledge unless it is unavoidably useful in speech communication. In particular, they will not abstract away a set of discrete phonological categories unless there is a compelling phonological reason to do so, some reason such as the categories' participation in the "dual structure" of a small inventory of meaningless entities of contrast that are used to compose a larger inventory of meaningful entities. So although I would include microprosody within the set of phenomena that we must model when we model intonation, I would use a superpositional phonetic representation to do so, and would certainly reject any proposal to specify an intermediate phonological structure between the pitch contour and the symbolic phonological representation of the categorical contrast between [+voice] and [-voice]. That is, I think we should assume that if speakers intend to produce these "intrinsic" effects, they are controlling them in terms of the task of producing the [+voice] or [-voice] category, and when listeners factor out the various influences on the pitch pattern, they perceive these "intrinsic" effects in terms of the voicing contrast directly without parsing any intermediate phonological categories of [raised pitch] versus [lowered pitch] . (Of course it is possible that sociolinguistic pressure might induce a particular speech community to begin parsing such intermediate phonological categories to represent how different subsets of the speaker/hearers sound different, but that is another story — the story of reanalysis and incipient tonogenesis.)

ON STRESS AND ACCENT

I have belabored this point, because it is relevant for the discussion of accent and accentual prominence, a set of phenomena concerning which we are perhaps in most disagreement. My discussion of microprosody is superficially similar to the argument that Grønnum [1] uses against a symbolic representation of pitch accent in Danish. That is, if I understand her correctly, she is saying that the rise in pitch anchored at the beginning of each stress group is predictable from the stress pattern in much the same way that the raised pitch onset associated with a preceding voiceless obstruent is predictable from the representation of the [-voice] category, and therefore there is no need to include in the model of intonation any structure representing that stress-group initial rise.

To explain why I disagree, I need to make clearer the distinction I understand between the two types of phonological category — those of metrical organization and those that discretize some dimension of phonetic content. Let me illustrate this with a discussion of vowel timbre categories. In English there are phonotactic constraints on where certain types of vowels can occur, constraints which are economically represented by positing an abstract metrical category "stress foot" , defined as a grouping of syllables that is headed by an obligatory initial "stressed" syllable where the full range of vowel contrasts is possible. These kinds of metrical properties are about the most abstract kind of phonological structure imaginable. The child acquiring the language must in effect abstract them from the phonotactic constraints on the content features defining them. Now imagine a language in which the foot is defined similarly, but the obligatory initial syllable always contains [a] and the optional trailing syllables always contain [i]. By Grønnum's argument, we would need no representation of vowel timbre categories in this language because the difference between [a] and [i] is already represented in the stress contrast. True, but I think a more plausible model of what children would do in acquiring this language is to build a phonetic representation of the vowel timbre space that allows them to abstract away from such things as speaker-related variability to a categorical contrast between [a] and [i], and then abstract away a metrical property "stress" to represent the constraint on where [a] and [i] occur. The analogous hypothesis about Danish is that children build a representation of pitch range to abstract away something like the categories [+Hightone] (henceforth "H") and a contrasting [-Hightone] ("L") so that they can then abstract away an understanding of the stress group in terms of the distribution of these tones. The child acquiring the Copenhagen dialect, for example, might decide that L only occurs on the group-initial stressed syllable and H only on the next syllable in the stress group.

This distinction between metrical properties and content features is even more critical in English, where there are several paradigmatically contrasting pitch accent types signaling different pragmatic functions. For example, the L+H* accent (where the "*" means that the rise is aligned to put the H tone on the stressed syllable) signals a choice of value along some semantic scale and a commitment to the pragmatic relevance of that value, whereas the L*+H (i.e. the same rise but aligned very similarly to the Copenhagen Danish stress-group marker) signals the same sense of scalar choice but lack of commitment [10]. This difference can have enormous consequences for the implied presuppositions of a statement. For example, in the context of the assertion *No one in his right mind works on intonation.* the response *Beckman works on intonation.* said with a L+H* accent on *Beckman* means that Beckman is a relevant counterexample, whereas the same sentence produced with a L*+H accent means that this is not really a relevant counterexample and thus implies that Beckman is dotty. Pierrehumbert & Hirschberg [11] describe contrasting pragmatic meanings associated with four other pitch accent types, including a contrast between a single-tone H* accent and a single-tone L* accent. These contrasts in pitch accent type must be included as part of the phonology of intonation. Thus, it is only the pattern of association between accents and syllables that can be relegated to the phonology of stress, to represent the constraint that an accent can occur only on the head of a stress foot, and the fact that pitch accent placement defines a categorical level of metrical prominence over and above the specification of stressed versus weak syllables in the lexicon. (There are further constraints on pitch accent placement which are related to such discourse phenomena as focus, and which are very similar to the facts that Grønnum describes about the "suppression" of the pitch rise on neighboring stress-groups to indicate emphatic contrast.)

I think most people who work on the phonology of English intonation would agree on the above characterization of pitch accents as something that must be represented in the phonology of the intonation pattern directly and independently of the relationship between pitch accent placement and stress. Also, while not everyone agrees on the exact inventory, there is a strong consensus that H and L tone levels are the right way to represent the accentual contrasts (see, [12] as well as Pierrehumbert's and Ladd's analyses). This is an important point of agreement, because some superpositional models of typologically similar intonational systems assume a different representation of pitch accent, in terms of rises or falls as holistic units (see Möbius's representation of the pitch accents of German [13]). One of the most compelling motivations for H and L as the "phonemes" of English intonation is that these occur alone as the intonational morphemes H* and L*. This analysis of single-tone accents is supported by intonation contours where several L* accents occur in succession with no intervening rise in pitch or several H* accents occur in succession with no intervening dip. (Several of the F0 contours of utterances being modeled in Möbius's paper make me think that German has comparable sequences of single-tone accents. Note, for example, the flat F0 pattern over the sequence *sieben Minuten entfernt* in Fig. 5.)

ON LOCAL PROMINENCE

The tone-level analysis of English accents is also supported by data on the way that F0 peaks associated with H* accents vary under changes in overall pitch range. In the classic experiment [4], subjects produced the sentence *Anna came with Manny* in response to contexts that induced one or another fixed pragmatic relationship between the names. The sentence was in both cases a sequence of two intonational phrases, so that both names were accented (and hence focused), but in half of the productions, *Anna* was assumed background and *Manny* the answer to the context question, and in the other half, *Manny* was the background focus and *Anna* the answer. The two renditions were produced many times at each of ten different overall vocal effort levels. When the F0 value of the second peak was plotted as a function of the first peak, the two accents showed a very tight clustering around the two regression curves for the contrasting pragmatic relationships, suggesting that the speakers were controlling the location of the accent peak within the overall pitch range. By contrast, when the extent of the F0 rise into the first peak was plotted against the extent of the second rise, there was no clear clustering. A recent study of cricothyroid muscle activity and subglottal pressure in productions of a somewhat different set of pragmatic relationships [3] offers further support for this interpretation.

While the experiment supports the representation of accent production in terms of H versus L target values at values relatively high or low within the overall pitch range, it also raises knotty questions about how to model the variability within the overall pitch range for the same accent in the two renditions. The peak was higher when the accent was the answer focus as compared to when it was the background focus (although the first peak was always at least as high as the second, whether answer or background). Liberman & Pierrehumbert themselves first modeled this variability by representing the overall pitch range in terms of a choice of scalar value for a "reference line" and then representing the two H* tones as differing in local phonetic "prominence" — another scalar value specified accent by accent and governing the distance of the accent's target tones above the reference line. Others have proposed alternative models, however. For example, a revision to Liberman & Pierrehumbert's original model suggested in [14] specifies a reference line for each intonational phrase (in effect allowing each phrase to have its own local pitch range), so that the answer/background peak relationship is modeled as a speaker strategy for choosing particular values for the reference lines of the two phrases.

The original proposal to represent the accent/background relationship in terms of choice of accentual prominence values was based on the desire to limit the degrees of freedom in the model, since accentual prominence was already being used to model one component of "declination". That is, Liberman & Pierrehumbert found that (for another large set of utterances by the same speakers) they could model the cumulative exponential decline over utterances with varying numbers of accents produced in three different overall pitch ranges, simply by specifying a proportional reduction within the overall pitch range of the prominence of each successive accent (as suggested earlier by Pierrehumbert [15]). This was an encouraging result for linear models, because Bruce [16] had just shown that the same local "downstep" worked better than Gårding's superpositional model [17] in accounting for the decline in the portion after the focal accent in a large dataset of Swedish utterances.

Pierrehumbert & Beckman's revision to Liberman & Pierrehumbert's original proposal introduced a new parameter — local pitch range — and was motivated by our difficulties in modeling downstep and its interaction with focus in a large dataset of Japanese utterances. We found that in order to model L tone targets when one accentual phrase is downstepped and pragmatically subordinated to its neighbors, we had to specify a local pitch range "topline" for each accentual phrase, a value like the prominence value specified for each pitch accent in Pierrehumbert's original model but affecting all of the tones within the phrase rather than just the tones of the accent.

This approach is reminiscent of superpositional models such as Grønnum's [1] or Gårding's [17] in that the phonetic specification of local pitch range is embedded within the specification of the speaker's overall pitch range for that degree of vocal effort. It differs from these more radically superpositional models, however, in that the embedding is phonologized in the metrical representation of stress and phrasing. The phonology of intonation itself is not superpositional, because the local graph paper is not phonologized. Rather, local pitch range is represented by a single continuously variable phonetic value (the reference line height) locally specified for the phrase.

Ladd's [18] reanalysis of the *Anna came with Manny.* data similarly involves more local pitch range specification. However, he differentiates the more local kind of variation from the specification of overall pitch range by making only the latter a continuous phonetic dimension. That is, he phonologizes local pitch range by constraining the values it can take in terms of l/h vs h/l relationships specified on branching nodes in a hierarchical phonological structure with the accents as leaf nodes. My own prejudice is that this is phonologizing relationships that belong to be represented elsewhere. For example, "answer/background" is really a relationship of pragmatic subordination between the two phrases. The F0 data that we are modeling when we examine this relationship give an impression of discretely constrained relationships between the local pitch range values, but this is an artifact of contrasting mini-discourses that felicitously allow only two choices for the degree of pragmatic subordination between the compared phrases. Differentiating my interpretation from Ladd's would require something like an experiment in which speakers are successfully induced by the context to produce more degrees of pragmatic subordination between the two accents.

WHERE FROM HERE?

There are many other directions in which this comparison of models could go. For example, in the discussion above I have referred to large datasets from three typologically unlike languages: English, Swedish, and Japanese. We could extend the comparison of models if we gathered comparably comprehensive datasets from yet other typologically different languages. Shih's [19] data for Mandarin Chinese is one such dataset, and presents problems for all of our models. Her results show a lowering of successive tone targets, but not after tone 1, and to different extents elsewhere, so that targets are lowered more after tone 3 than after tone 2 or 4. This "differential downstep" is difficult to accommodate in any superpositional model that represents the lowering by a declination built into the graph paper at any level of grouping of tones. Shih's data also show focus having the general effect of locally raising high tone targets (an effect that linear models should accommodate ideally). However, in strings of tone 1 the targets following the focus also are raised and only gradually decline back to a more neutral value. That is, tone height does not immediately return to neutral (so that focus could be modeled as affecting only the focused tone's "prominence") nor does it stay at the higher level until some metrical domain edge (so that focus could be modeled as increasing the local pitch range value).

Another area where we need much more data is the behavior of intonational events that are low in the pitch range. Here our phonetic representation serves us ill, because F0 is not well-defined for the irregular phonations often seen in extremely low-pitched parts of the utterance. In the original English model described in [4], the degrees of freedom were limited by having a constant "baseline" value representing the bottom of the speaker's overall pitch range. This was motivated by the data's apparent reconfirmation of the long-standing finding that F0 values at the end of declarative contours in English seem relatively invariant compared to peak values. Hirschberg & Pierrehumbert's simulation of varying endpoints to reflect degree of topic embedding in a monologue [20] suggests that the invariance is an artifact of measuring endpoint values at the last point before creak makes F0 a bad measure of pitch. We need to find a better representation of pitch in these regions, perhaps by first doing basic perception studies to see how perceived pitch varies for different kinds of creaky phonation.

Also, I should like us to begin looking at physiology. What we know of F0 control makes it extremely doubtful that we will not find any very simple mapping between particular model parameters and particular muscles or parts of muscles. On the other hand, our preliminary data on subglottal pressure and EMG activity levels [2, 3] suggest that the variation in overall pitch range associated with changes in vocal effort levels can be separated from some sorts of variation in local accentual prominence in being created in large part by the increased subglottal pressure at louder vocal effort levels. It would be interesting to see whether physiological data support an understanding of more local pitch range variation as a local change in vocal effort.

I am sure the other participants in this symposium can add other areas in which they would like to see more data to enrich our models and our debate. So to respond to Ladd's question, I think we should leave this debate where it stands right now, and go have a beer so we can start fresh tomorrow in the laboratory.

REFERENCES

[1] Grønnum, N. (1995). "Superposition and subordination in intonation — a non-linear approach." This symposium.
[2] Erickson, D., Honda, K., Hirai, H., Beckman, M. E., & Niimi, S. (1994). "Global pitch range and the production of low tones in English intonation," *ICSLP '94*, vol. 2, pp. 651-654.
[3] Beckman, M. E., Erickson, D., Honda, K., Hirai, H., & Niimi, S. (1995). "Physiological correlates of global and local pitch range variation in the production of high tones in English," *13th ICPhS*.
[4] Liberman, M., & Pierrehumbert, J. (1984). "Intonational invariance under changes in pitch range and length," In M. Aranoff & R. Oehrle, eds., *Language Sound Structure*, Cambridge, MA: MIT Press.
[5] Ross, E. D., Edmonson, T. A., Seibert, G. B., & Chan, J.-L. (1992). "Affective exploitation of tone in Taiwanese: an acoustical study of 'tone latitude'," *J. Phon.*, vol. 20, pp. 441-456.
[6] Jun, S.-A., & Oh, M. (1994). "A prosodic analysis of three sentence types with 'WH' words in Korean," *ICSLP '94*, vol. 1, pp. 323-326.
[7] Miura, I., & Hara, N. (in press). "Production and perception of rhetorical questions in Osaka Japanese," *J. Phon.*
[8] Hirschberg, J., & Ward, G. (1992). "The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English," *J. Phon.*, vol. 20, pp. 241-251.
[9] Herman, R. (1995). "Final lowering in Kipare," *Ohio St. U. Working Papers in Linguistics*, no. 45, pp. 36-55.
[10] Ward, G., & Hirschberg, J. (1985). "Implicating uncertainty: the pragmatics of fall-rise intonation," *Language*, vol. 61, pp. 747-776.
[11] Pierrehumbert, J. & Hirschberg, J. (1990). "The meaning of intonation contours in the interpretation of discourse," In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in Communication*, pp. 271-311. Cambridge, MA: MIT Press.
[12] Gussenhoven, C. (1984). *On the Grammar and Semantics of Sentence Accents*, Dordrecht: Foris.
[13] Möbius, B. (1995). "Components of a quantitative model of German intonation," This symposium.
[14] Pierrehumbert, J. B., & Beckman, M. E. (1988). *Japanese Tone Structure*, Cambridge, MA: MIT Press.
[15] Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*, Doctoral dissertation, MIT.
[16] Bruce, G. (1982). "Developing the Swedish intonational model," Working Papers, Lund, no. 22, pp. 51-116.
[17] Gårding, E. (1979). "Sentence intonation in Swedish," *Phonetica*, vol. 36, pp. 207-215.
[18] Ladd, D. R. (1995). "'Linear' and 'overlay' descriptions: an Autosegmental-Metrical middle way," This symposium.
[19] Shih, C.-L. (1988). "Tone and intonation in Mandarin," *Working Papers, Cornell Phonetics Laboratory*, no. 3, pp. 83-109.
[20] Hirschberg, J., & Pierrehumbert, J. (1986). "The intonational structuring of discourse," *24th ACL*, pp. 136-144.