

1. Introduction. **Models and Measurements in the Study of Prosody**

D. Robert Ladd and Anne Cutler

1.1 **Concrete and Abstract Accounts of Prosody**

There are two broad traditions in the study of prosody that may be characterized - or caricatured - by their methodological preferences for one or the other of the scientific activities mentioned in the title: making measurements and constructing models. On one side of the dichotomy stand instrumental and experimental studies that seek to quantify acoustic features and investigate perceptual responses. On the other are descriptive and theoretical studies of prosodic structure and its relation to other aspects of grammar and phonology. In a great deal of past work these two traditions have simply ignored one another.

Occasionally, however, the differences between the two traditions have surfaced in explicit condemnation of the model-builders by the measurers. A pointed example can be found in Ohala's review of Lehiste 1970, in which he criticizes Lehiste for "faithfully set[ting] forth the full repertory of clever solutions which linguists have been able to come up with The catalog of such facile inventions by taxonomic linguists . . . seems out of place in a book otherwise devoted to what scientists of language have proven and demonstrated empirically about the behavior of speech sounds" [Ohala, 1975, pp. 737-738]. The implication of such criticisms is that the two traditions are based on fundamentally different assumptions about what constitutes empirical work.

We disagree. We believe that the measurers and the model-builders have fundamentally different assumptions not primarily about methodology, but about function and representation of prosody. These assumptions can be explicitly stated and carefully examined: the issues are potentially empirical ones that can be tested in ways that will satisfy a range of methodological preferences. In particular, it seems to us that two of the main points of substantive disagreement implicit in the differences of methodology have to do with the *abstractness* of prosodic representation and the *directness* of the link between prosodic function and prosodic form.

One approach - let us call it the "concrete" approach - defines prosody more or less in physical terms, as *those phenomena that involve the acoustic parameters of pitch, duration, and intensity*. This approach conceives of the link between form and function in prosody as a relatively direct mapping

between concrete meanings or functions and specific acoustic shapes or variables. Prosodic cues convey a relatively heterogeneous range of specific meanings - from "anger" to "phrase boundary" to "emphasis" - and constitute a kind of parallel channel or accompaniment to the central linguistic message of the utterance. Notions such as "continuation rise", "ironic intonation", and even "contrastive stress" are theoretical constructs of this general point of view. The main aim of "concrete" research is to identify the correlations between specific messages and specific acoustic parameters.

Compared to the concrete approach, the "abstract" approach sees prosody more from the point of view of its place in linguistic structure than its phonetic nature, and tends to class as prosodic *any phenomena that involve phonological organization at levels above the segment*. In practice, of course, the two definitions cover much of the same ground - intonation, stress, phrasing, etc. - but the different basis of the two approaches can be seen by the fact that certain phenomena, such as vowel harmony, might be considered "prosodic" by the abstract definition but not by the concrete one, while the reverse is true for phonemic vowel length, for example.

The difference of definition entails clearly distinct assumptions about function. In particular, the abstract approach does not assume that the relation between form and function is so straightforward as to justify the search for direct correlates. Prosody is not a phonetically defined component of language with a set of heterogeneous functions like "phrase boundary" and "question"; such messages are not conveyed by the prosody as a peripheral channel running parallel to the text, but as part of the whole linguistic structure, of which the prosodic features are one aspect. Acoustic parameters are seen as cues to abstract formal categories, not to concrete functional ones, and the main aim of prosodic research is to model the formal system.

1.2 Illustrating the Two Approaches

In order to illustrate the interaction of theoretical assumptions with methodological preferences, let us consider examples of the concrete and abstract approaches to two specific topics in prosodic research: the function of intonation, and the nature of stress.

1.2.1 Intonation

A good deal of "concrete" research on the relation between intonation and grammar treats fundamental frequency and duration as the phonetic accompaniment of certain types of syntactic units or constituent boundaries, and

measures the acoustic characteristics of prosodic features in different syntactic environments. The most ambitiously reductionist work in this tradition is undoubtedly that of Lieberman [1967] who claimed that all the linguistically relevant uses of intonation could be reduced to a single, innate contrast between "marked and unmarked breath group", plus a prominence feature of increased subglottal pressure. Less controversial work along these general lines concentrates on specific syntactic environments such as yes/no questions, non-terminal boundaries, parenthetical phrases, echo questions, and so on. A great deal of work on French has taken this approach [Delattre, 1963; Martin, 1980; Grundstrom and Leon, 1973], as has recent work whose emphasis is on developing prosodic data for use in speech synthesis [O'Shaughnessy, 1976; Klatt, 1979]. Similar assumptions are made in Lehiste's [1975] work on paragraph cues and in much of the recent work on "declination" and its relation to syntax [Collier, 1975b; Maeda, 1976; Cooper and Sorensen, 1981].

In the same way, a direct connection is often assumed to exist between the prosodic characteristics of speech and various affective messages. Research on the vocal expression of emotion often treats the entire non-segmental part of the speech signal as a kind of overlay directly communicating emotional state. Acoustic correlates of the emotions signalled in this manner are frequently analyzed in terms of parameters that include such things as average fundamental frequency and overall frequency range [Williams and Stevens, 1972; Scherer, 1981]. Note that this work, when taken together with the work cited in the preceding paragraph, implies the existence of potential conflicts or interactions between the grammatical and emotional specifications of prosody. While this has not, to our knowledge, been seriously investigated, it is generally assumed that on the one hand, the two types of effects are often manifested in different ways (e.g. in the distinction between intonation and voice quality), and on the other hand, that interactions between the two uses, in particular on fundamental frequency, will ultimately be specifiable in precise acoustic terms.

The relation between grammatical and affective functions of intonation is seen quite differently in a wide range of "abstract" research, in particular the work done by a long line of scholars of intonation whose native language is English. Even a superficial look at English intonation is enough to make one suspect that the apparently different grammatical and emotional functions may actually be expressed by the same set of formal distinctions: a contour that is neutral in one context may express marked affect in another. Consequently, much work on English intonation does not start from a basic set of intonational *functions*, but assumes instead the existence of an inventory of more abstract intonational *categories* (e.g. "nuclear tones"). This emphasis has meant that functional labels have been used mostly as unsystematic glosses of the linguistic distinctions under investigation; most older work on English (e.g. the British pedagogical tradition of Kingdon [1958], O'Connor and Arnold [1961], etc.) simply mixes grammatical and affective descriptions in

discussing intonational meanings (e.g. peremptory request, surprised question).

However, such descriptions are consistent with more recent explicit attempts to explain the surface heterogeneity of intonational function in terms of some deeper unity, specifically those explanations which rely (in current terminology) on the notion of pragmatic inference. This approach was first suggested by Pike [1945], who assigned approximate abstract meanings to intonation contours and left the context to narrow the range of specific interpretations; Pike specifically argued against the use of concepts like "question intonation". Explanations in this vein have been advanced more recently by Bolinger [e.g. 1958, 1982], Liberman and Sag [1974], Menn [1976], Ladd [1980], Brazil et al. [1980], and Cruttenden [1981]. Note also that this abstract approach is not inconsistent with the more experimentally oriented work of 't Hart and Collier [e.g. 1975] or Pierrehumbert [1981], which assumes that any description of intonational meaning must be in terms of the abstract phonological units of the intonational analysis, not in terms of specific grammatical or affective functions.

1.2.2 Stress

Similar disagreements between the concrete and abstract approaches can be found in studies on stress: here the issue is not so much the directness of the link between form and function, as the abstractness of the linguistic categories under investigation. The observation to be accounted for is that one or more of the syllables in a word or utterance are somehow more prominent than the rest. A great deal of "concrete" research (starting with Fry [1955] and a number of others in the 1950s and continuing through such recent work as Gay [1978]) has shown clearly that increases or changes in pitch and increases in intensity and duration on a given syllable are highly correlated with judgments that that syllable is the most prominent in a word or utterance. A great deal of "abstract" research, meanwhile (going back to Newman [1946] and continuing through Chomsky and Halle [1968] to the recent work inspired by Liberman and Prince [1977]), has concentrated instead on the patterns of prominence in lexical derivation, and on relative prominence of syllables other than the most stressed. In the first view, stress is a relatively concrete feature realized by specific acoustic cues; a syllable either has it or does not have it. In the second view, stress is a relatively abstract feature that specifies the relative prominence of syllables in a prosodic structure; the whole structure - not individual syllable features - determines the acoustic details of pitch, duration, and so on.

1.3 Implications of the Two Approaches

By now it should be clear how one's approach to investigating prosody reflects not only methodological preference, but also the very conception of the questions to be investigated. The model-builder is interested in establishing an inventory of abstract categories - a formal *representation* - of prosodic function and prosodic form. The goal of the model-builder's enterprise is to describe the systematic structure underlying prosodic distinctions; the basic assumption is that there are well-defined abstract levels of representation that mediate between specific prosodic functions like "phrase boundary" and specific acoustic traits.

By contrast, for many of those who take the measurer's approach, the primary concern is not representation, but *realization*. The question being asked is: What are the physical correlates of this or that prosodic message? To the extent that such investigators have constructed explicit models of prosodic representation, they have tended to think in terms not of linguistic categories, but of interacting parameters; their models assign acoustic correlates to individual functions such as word stress, sentence stress, sentence modality, affective use of pitch range, and so on, and attempt to specify the interaction of all these effects on individual parameters like fundamental frequency.

When the two positions are expressed in this way, it becomes clear that one of the main issues between them is what we might call "the place of prosody in language". Some of the functions normally considered under the rubric of prosody do seem to represent a separate communicative channel that accompanies the verbal message but is peripheral to it: the least controversial example is perhaps the differences of voice quality that accompany different emotional states. The disagreement arises over the extent to which this peripheral-channel model can properly be applied to other prosodic phenomena. Descriptions of prosody that contain abstract phonological categories or assign a central role to pragmatic inference take notions that have proven useful in the description of other aspects of language and apply them to prosody; this implies a belief that in at least certain respects prosody is an integral part of language. The experimental view, insofar as it assumes direct links between physiology and prosodic function, or describes prosodic form in terms of interacting parametric variables rather than abstract categories, treats prosody more as a separate channel than as part of the basic structure itself.

The issue of whether prosody is "central" or "peripheral" is surely not a simple empirical question with an either/or answer, but it does suggest more specific empirical questions and hypotheses that could reflect on the validity of the assumptions that underlie the different approaches to prosody. For example, clinical data and neurolinguistic research are a valuable source of empirical evidence whose implications are seldom taken into account. Such

evidence suggests that in some ways prosodic phenomena are indeed peripheral to normal language function: for example, it is well known that even global aphasics can often express emotion vocally, and there is some experimental evidence [Blumstein and Cooper, 1974; Van Lancker and Fromkin, 1973] to suggest that intonational cues are processed differently from other linguistic stimuli. Other experimental evidence [Zurif and Mendelsohn, 1972] casts doubt on the view that prosody is processed separately, and other clinical data suggest that certain prosodic distinctions may be lost in aphasia with the rest of language (for a review see Weniger [1978]). An important contribution to resolving the issue of prosody's role in language could be made by experiments and clinical studies designed to test specific hypotheses.

In an introduction of this sort we cannot properly take sides on the general issue of concrete and abstract models of prosodic function and representation, since the contributors whose work we are introducing clearly represent different points of view. We have, however, tried here to place the papers in a context that emphasizes the interdependence - even the inseparability - of making measurements and constructing models. Those whose methodological preference runs to experimental work must acknowledge that experimental data are not theoretically neutral, but presuppose a frame of reference that can and should be made explicit. Those who are concerned with building models, meanwhile, must acknowledge the importance of experimental testing and of reinterpreting experimental data based on different points of view. The key to progress in understanding prosody is to examine assumptions and, as far as is possible, test them empirically. That is the single most important point that we hope will emerge - both by example and by juxtaposition of examples - from the contributions collected here.

1.4 The Contributions

The foregoing sketch allows us to place the individual contributions to this volume in a common context. The first three - by Garding, Thorsen, and Ladd - deal with the representation of intonation, and they nicely illustrate the range of approaches we have discussed so far. Thorsen accepts many of the assumptions of the "concrete" approach, Ladd's model is rather "abstract", and Garding falls somewhere in between, though perhaps closer to Thorsen in most respects. Garding presents a summary of the Lund model of prosody, and illustrates how it can be used to generate different prosodic characteristics in three different languages. Pitch is specified by the interaction of grid lines for sentence intonation, tone sequences for accent, and various modifications for rhythmic and other effects. The grid lines typify the

intermediate position that Garding takes: their height and shape directly represent the pragmatic effects of intonation, but they also play a role in specifying the phonetic detail of the rather abstract tone sequences by which Garding represents accent. Thorsen's model is more "concrete", representing both sentence intonation and accent by interacting configurations. The main concern of her paper is to demonstrate the necessity for such a model in Danish; in particular, she argues that the more abstract notions of sentence stress and tone sequence are unable to account for her instrumental data. Ladd, finally, illustrates a feature analysis of tone sequences, and argues that such a relatively abstract description can actually give more insightful and more accurate accounts of overall slope than models in which overall slopes are directly specified as lines or configurations.

The next three contributions, by Vaissiere, Brown, and Cutler, are all concerned in one way or another with prosodic function. Vaissiere explicitly adopts the "concrete" point of view and the assumption of a direct link between function and form, arguing that prosodic function is to a great extent physiologically natural. She discusses several specific functions - marking stress groups, marking phrase boundaries, marking questions, and so on - that appear to be found in a wide variety of languages. For each such function she identifies common acoustic correlates and emphasizes the importance of physiological data. Brown, by contrast, assumes a relatively abstract approach to prosodic function and concentrates on a specific problem, namely the role of prosody in signalling discourse structure. She treats a single formal category, pitch prominence, and identifies it with a functional label, "new information". In keeping with the assumption of indirect connection between prosodic form and utterance interpretation, she is at pains to show that there are *not* intonational correlates for the whole range of possible discourse "statuses" posited by E. Prince [1981] and others: pitch prominence is a "generalized 'pay attention' marker", and more specific categorization of discourse status is inferred by the speaker on the basis of a variety of syntactic and other cues. Cutler, finally, reports on her own corpus of speech errors involving prosodic features. Different types of errors produce different patterns of correction; in general, speakers are more likely to correct prosodic errors the more seriously they disrupt the interpretation of the utterance. Other details suggest that language users behave as if prosody were definitely a peripheral, separate, and in many cases dispensable component of language.

The final three contributions are all relevant to recent abstract models of prosodic structure proposed by Selkirk [e.g. 1980] and others. Hirst in particular is concerned to show that even research on fine phonetic detail may be relevant to the broadest theoretical questions of the organization of language. He shows how, within the general framework of Selkirk's model of prosody, it is possible to describe intonation in a way that represents both its relation to the hierarchical organization of utterances and its phonetic linearity. His linear phonetic representation bears a considerable resemblance to

Garding's, while his hierarchical structures are essentially the same as those discussed at greater length by Nespors and Vogel.

Buxton describes several experiments in which fine details of speech timing were manipulated in stimuli. Her results point clearly to the existence of temporal structure that is critical for the fluent processing of speech. Among other things, these results can be read as evidence for the centrality of prosodic structure in language: timing is not simply a matter of, for example, adding length to a syllable because of an accompanying feature of stress, but rather of fitting all the sounds, all the syllables, into an overall frame that is part of the phonological representation of the utterance.

This view is given a more explicit form by Nespors and Vogel, who build on Selkirk's model of prosody and make predictions about prosodic disambiguation in Italian. They also report experimental results that bear out their predictions. What is of particular relevance to the question of the centrality of prosody is the fact that the prosodic structures which Nespors and Vogel discuss affect many different physical aspects of the speech signal, including phenomena such as "radoppiamento sintattico" which are not normally considered "prosodic". That is, their model implies that it is not possible to identify certain functions or acoustic features as prosodic and treat them apart from the rest of language, but rather that everything in the realization of an utterance is potentially affected by its prosodic organization.

1.5 The Purpose of an Interdisciplinary Volume

While we have argued strongly that different approaches to studying prosody reflect empirical assumptions that can be stated and examined, we do not mean to underestimate the extent to which different traditions of inquiry are due simply to lack of communication among linguistics, phonetics, and psychology. Traditional disciplinary boundaries still make themselves felt most obviously in the fact that, for example, people who read *Linguistic Inquiry* do not normally read *The Journal of Experimental Psychology*, and vice versa. The effects of this can be illustrated most dramatically in the very recent history of work on rhythm, timing, and metrical organization. Within linguistics, an entire school of prosody and metrics has developed within the last five or six years, based on Liberman and Prince's article "On Stress and Linguistic Rhythm" [1977]. (Notable works in this tradition include Selkirk [1980], Kiparsky [1979], Hayes [1981], and McCarthy [1979].) Meanwhile, in psychology and experimental phonetics, in about the same length of time, there has been a resurgence of interest in the question of "isochrony" and rhythm, as represented by the work of Martin and his colleagues [e.g. Martin, 1972; Meltzer et al., 1976], the work on perceptual centres by Morton et

al. [1976], and other work on timing by Lehiste [1977], Fowler [1979], Donovan and Darwin [1979], Nakatani et al. [1981], and Scott [1982]. Despite the fact that both these approaches are directly concerned with the temporal patterning of prominence in language, however, there has been virtually no contact between them. Linguistic model building has concentrated on the assignment of lexical stress, and has worked with notions like "relative prominence" and "metrical grid" without making any attempt to show what those might mean in perceptual or acoustic terms. The psychological work, meanwhile, strongly suggests the existence of temporal structure to speech, but has scarcely gone beyond formulations like "a tendency to isochrony" in explaining experimental results. (Martin [1972], it is true, proposes a hierarchical model not unlike Liberman and Prince's, but his experiments have not really attempted to test its predictions.)

One obvious purpose of a volume like this, then, is to acquaint people who read one set of journals with work of the sort that appears in various other sets. Yet even when researchers from one field explore what has been done in another, the problems of methodology, research strategy, and background assumption remain. The thousand subtle lessons that are learned in becoming a practitioner of one field may contribute to the feeling that work in neighboring fields is uninteresting, irrelevant, or worse. The tension between model building and doing experiments - which in this context strongly tends to divide the linguists from the psychologists and the phoneticians - is still often a significant barrier to the spread of ideas.

Beyond the obvious goal of making papers from different disciplines available in the same place, our main aim in assembling this collection has been, as we said, to illustrate the interdependence of models and measurements. Specifically, we hope to have raised questions like the following:

a) Vaissiere presents data showing the universality of certain prosodic features, while Garding and Thorsen both present data showing how language-specific prosody can be. What sort of model can account for both kinds of observations?

b) We have described both Ladd and Brown as exemplifying the "abstract" approach to prosodic function. Why, then, are the details of their models so different, and what predictions do their models make that could be tested empirically?

c) Buxton suggests certain ways in which the hierarchical model assumed by Nespor and Vogel and Hirst may be inadequate to account for data from speech perception. In what ways could the hierarchical model be revised to take account of such data? How would revision affect the model's applicability to broader questions of the organization of language? How could a revised model be tested?

d) Cutler discusses the problem of uncorrected prosodic errors, while Brown's emphasis on "speaker choice" would appear to make it difficult to identify a prosodic error unless it is corrected. If we grant the difficulty of directly testing the validity of the concept "uncorrected prosodic error", then

we must instead test larger models of which that concept is or is not a part. What would such models look like, and how would we test them?

If readers are asking themselves such questions when they finish this book, then it will have fulfilled the best purpose for projects that cross traditional boundaries between fields - to enable people on opposite sides of such boundaries not just to become aware of each other's work, but to see the point of it as well.