1
Introduction

MARY E. BECKMAN AND JOHN KINGSTON

While each of the papers in this volume has its specific individual topic, collectively they address a more general issue, that of the relationship between the phonological component and the phonetic component. This issue encompasses at least three large questions. First, how, in the twin processes of producing and perceiving speech, do the discrete symbolic or cognitive units of the phonological representation of an utterance map into the continuous psychoacoustic and motoric functions of its phonetic representation? Second, how should the task of explaining speech patterns be divided between the models of grammatical function that are encoded in phonological representations and the models of physical or sensory function that are encoded in phonetic representations? And third, what sorts of research methods are most likely to provide good models for the two components and for the mapping between them?

Previous answers to these questions have been largely unsatisfactory, we think, because they have been assumed *a priori*, on the basis of prejudices arising in the social history of modern linguistics. In this history, phonology and phonetics were not at first distinguished. For example, in the entries for the two terms in the *Oxford English Dictionary* each is listed as a synonym for the other; *phonology* is defined as "The science of vocal sounds (= PHONETICS)" and *phonetics* as "The department of linguistic science which treats of the sounds of speech; phonology." The subsequent division of this nineteenth-century "science of sounds" into the two distinct subdisciplines of phonology and phonetics gave administrative recognition to the importance of the grammatical function of speech as distinct from its physical structure and also to the necessity of studying the physical structure for its own sake. But this recognition was accomplished at the cost of creating two separate and sometimes mutually disaffected scientific subcultures.

We can trace the origin of this cultural fissure to two trends. One is the ever-increasing reliance of phonetic research on technology, rather than on just the analyst's kinesthetic and auditory sensibilities. This trend began at least in the first decade of this century, with the use of the X-ray to examine vowel production and

the adoption of the kymograph for examining waveforms. With such technical aids, phoneticians could observe the physical aspects of speech unfiltered by its grammatical function. With this capability, phonetics expanded its subject matter far beyond the taxonomic description of "speech sounds" found in phonological contrast, to develop a broader, domain-specific attention to such extra-grammatical matters as the physiology of speech articulation and the physics of speech acoustics, the peripheral and central processes of speech perception, and the machine synthesis and recognition of speech.

The other trend that led to the separation of the two subdisciplines was the development of more complete formal models of the grammatical function of speech than are instantiated in the International Phonetic Alphabet. This trend had its initial main effect in the 1930s, with the emergence of distinctive feature theory, as elaborated explicitly in Prague Circle phonology (Trubetzkoy 1939) and implicitly in the American structuralists' emphasis on symmetry in analyzing phonological systems (Sapir 1925). Distinctive feature theory effectively shifted the focus of twentieth-century phonology away from the physical and psychological nature of speech sounds to their role in systems of phonemic contrast and morphological relatedness.

Both of these trends undermined the alphabetic model that underlay the nineteenth-century synonymy between phonetics and phonology, but they did so in radically different ways. The analysis of "vocal sounds" into their component units of phonological contrast eventually led to new non-alphabetic representations in which phonological features were first accorded independent commutability in different rows of a matrix and then given independent segmentation on different autosegmental tiers. The use of new technology, on the other hand, questioned the physical basis originally assumed for alphabetic segmentation and commutability, by revealing the lack of discrete sequential invariant events in articulation or acoustics that might be identified with the discrete symbols of the IPA. These radically different grounds for doing away with a strictly alphabetic notation for either phonological or phonetic representations produced an apparent contradiction.

Modelling the cognitive function of speech as linguistic sign requires two things: first, some way of segmenting the speech signal into the primitive grammatical entities that contrast and organize signs and second, some way of capturing the discrete categorical nature of distinctive differences among these entities. A direct representation of these two aspects of the grammar of speech is so obviously necessary in phonological models that it is hardly surprising that the early, rudimentary phonetic evidence against physical segmentation and discreteness should elicit the reaction that it did, a reaction caricatured in Trubetzkoy's declaration that "Phonetics is to phonology as numismatics is to economics." A more benign form of this prejudice recurs in the common assumption among phonologists that nonautomatic, language-specific aspects of phonetic repre-

sentations and processes should share the discrete segmental nature of phonological symbols and rules.

This apparent contradiction induced also a complementary prejudice on the part of phoneticians. Instrumentally-aided investigation of speech has resulted in decades of cumulative progress in phonetic modeling, including the monumental achievement of the acoustic theory of speech production (Fant 1960). A great deal of this research has necessarily been concerned with the details of mapping from one extra-grammatical system to another – for example, from acoustic pattern to cochlear nerve response or from motor excitation to articulatory pattern. This research into the relationships among different phonetic subcomponents has derived little direct benefit from advances in phonological theory. As a result, it has often been assumed that arguments about phonological representations and processes are irrelevant to the phonetic component as a whole, a prejudice that could be expressed in its most malignant form as "phonology is to phonetics as astrology is to astronomy."

We have caricatured these prejudices at some length because we feel that they are a major impediment to answering our three questions concerning the relationship between phonology and phonetics. They distort our pictures of the two linguistic components and of the shape of the mapping between them. One set of theories describes the mapping as a trivial translation at the point where the linguistically relevant manipulations of discrete symbolic categories are passed to the rote mechanics of production and perception. Another set of theories places the dividing line at the point where the arbitrary taxonomy of linguistic units yields to experimentally verifiable models of speech motor control, aerodynamics, acoustics and perception.

Such distortions are inevitable as long as the relegation of aspects of sound patterns between the two linguistic components is guided by unquestioned assumptions about what research methods are appropriate to which field. Therefore, we ask: how can we use the physical models and experimental paradigms of phonetics to construct more viable surface phonological representations? Conversely, what can we learn about underlying phonetic representations and processes from the formal cognitive models and computational paradigms of phonology? Determining the relationship between the phonological component and the phonetic component demands a hybrid methodology. It requires experimental paradigms that control for details of phonological structure, and it requires observational techniques that go beyond standard field methods. The techniques and attitudes of this hybrid laboratory phonology are essential to investigating the large group of phonic phenomena which cannot be identified *a priori* as the exclusive province of either component.

An example of such a phenomenon is fundamental frequency downtrend. It is a common observation that F_0 tends to fall over the course of an utterance. Phonologists have generally assumed that this downtrend belongs to the

phonological component. They have postulated simple tone changes that add intermediate tone levels (e.g. McCawley's 1968 rule lowering High tones in Japanese to Mid tone after the first unbroken string of Highs in a phrase), or they have proposed hierarchical representations that group unbroken strings of High tones together with following Lows in tree structures that are interpreted as triggering a downshift in tonal register at each branch (e.g. Clements 1981). Phoneticians, on the other hand, have typically considered downtrend to belong exclusively to the phonetic component. They have characterized it as a continuous backdrop decline that unfolds over time, independent of the phonological tone categories. They have motivated the backdrop decline either as a physiological artifact of decaying subglottal pressure during a "breath group" (e.g. Lieberman 1967), or as a phonetic strategy for defining syntactic constituents within the temporal constraints of articulatory planning (e.g. Cooper and Sorenson 1981).

Each of these models is circumscribed by our notions about what research methods are appropriate to which linguistic subcomponent. If the observed downtrend in a language is to be in the province of phonological investigation, it must be audible as a categorical tone change or register difference, and its immediate cause must be something that can be discovered just by examining the paradigm of possible phonological environments. If the downtrend is to be in the province of phonetic investigation, on the other hand, it must be quantifiable as a response to some physically specifiable variable, either by correlating fundamental frequency point-by-point to subglottal air pressure or by relating fundamental frequency averages for syllables to their positions within phonologically unanalyzed utterances of varying length. Each sort of model accounts for only those features of downtrend which can be observed by the methods used. Suppose, however, that the downtrend observed in a given language is not a single homogeneous effect, or suppose that it crucially refers both to discrete phonological categories and to continuous phonetic functions. Then there will be essential features of the downtrend that cannot be accounted for in either model. Indeed these features could not even be observed, because the research strategy attributes downtrend *a priori* either to manipulations of phonological representations or to phonologically blind phonetic processes.

In recent examples of the hybrid methods of laboratory phonology, Pierrehumbert has argued with respect to English and Poser and others (Pierrehumbert and Beckman 1988) regarding Japanese that downtrend is just such a heterogeneous complex of different components, many of which are generated in the mapping between phonological and phonetic representations. In both English and Japanese, certain phrase-final tones trigger a gradual lowering and compression of the pitch range as a function of the distance in time from the phrase edge. This component of downtrend is like the phonologically-blind declination assumed in earlier phonetic models in that it seems to be a gradual backdrop decline. Yet it is unlike them in that it refers crucially to phonological

phrasing and phrase-final tone features. Also, in both languages, certain other, phrase-internal, tonal configurations trigger a compression of the overall pitch range, which drastically lowers all following fundamental frequency values within some intermediate level of phonological phrasing. This largest component of downtrend is like the intermediate tone levels or register shifts in earlier phonological models in that it is a step-like change triggered by a particular phonological event, the bitonal pitch accent. Yet it is unlike them in that it is implemented only in the phonetic representation, without changing the phonological specification of the affected tones. If these characterizations are accurate, then downtrend cannot be modeled just by reference to the phonological or the phonetic structure. Indeed neither of these two components of downtrend can even be observed without instrumental measurements of fundamental frequency values in experiments that control for phonological tone values and phrasal structures. The phenomenon of downtrend seems to require such hybrid methods.

We think, moreover, that the list of phenomena requiring such hybrid methods and models is much larger than hitherto supposed. We believe that the time has come to undo the assumed division of labor between phonologists and other speech scientists; we believe this division of labor creates a harmful illusion that we can compartmentalize phonological facts from phonetic facts. At the very least, we maintain that the endeavor of modeling the grammar and the physics of speech can only benefit from explicit argument on this point. In support of this thesis, we present to you the papers in this volume.

Most of these papers were first presented at a conference we held in early June of 1987 at the Ohio State University. To this conference we invited about 30 phonologists and phoneticians. The papers at the conference were of two sorts. We asked some of the participants to report on their own research or ideas about some phenomenon in this area between phonology and phonetics. We asked the other participants to present papers reacting to these reports, by showing how the research either did or did not consider relevant phonological structures or phonetic patterns, and by reminding us of other research that either supported or contradicted the results and models proposed. By structuring the conference in this way we hoped to accomplish two things. First, we wanted to show the value of doing research in this area between phonology and phonetics, and second, we wanted to provoke phonologists and phoneticians into talking to each other and into thinking about how the methods and aims of the two fields could be united in a hybrid laboratory discipline tuned specifically to doing this sort of research. After the conference, we commissioned both sets of participants to develop their presentations into the papers which we have grouped in this volume so that the commentary papers follow immediately upon the paper to which they are reacting.

The specific topics that these groups of papers address fall into several large categories. First are papers which focus on suprasegmental phenomena in

language. One of these is the representation of tone and intonation. In this group, Inkelas and Leben examine tone patterns in Hausa, a language which is unlike English or Japanese in having a dense lexical assignment of tone. They ask what intonational patterns can exist in such a language and how such patterns should be represented. They show representative F_0 contours and mean F_0 values in a lexically and syntactically various set of Hausa sentences to argue for several intonational modifications of lexical tone patterns, including downdrift (a downstepping at each phrase-internal LH sequence) and a boosting of tones in the final phrase of a question. They propose that all of these modifications, including downdrift, are generated directly in the phonology by inserting tones on an autosegmental register-tone tier. These register tones are indirectly associated to tones on a primary tone tier by being linked to the same tonal nodes for minimal tone-bearing units, and they are interpreted in the phonetics as contextual modifications to the values for the corresponding primary tones. The boosting of tones in questions is represented as the effect of H register tones that link to tonal nodes in the final phrase, and downdrift is represented as the effect of a L that is inserted at each register slot corresponding to a primary L and then spread to following nodes for primary H tones. Inkelas and Leben argue that only this sort of directly phonological representation can explain the distribution of intonational effects in Hausa. For example, downdrift seems to be blocked in the last phrase of a question, just as would be expected if the posited register H here blocked the insertion of the downdrift-causing register L. Pierrehumbert and Beckman (1988), on the other hand, have suggested that some of these seemingly categorical dependencies might be artifacts of interactions among the effects of phonetic rules. For example, if the boosting of tones in the last phrase of a question were a gradual increase like the final raising attested in Tokyo Japanese, it could obliterate the effects of a phonetic downdrift rule without actually blocking its application. The experiments necessary to distinguish between these two accounts promise to enrich the phonological theory by the constraints on surface phonological representations that they might indicate.

Ladd's paper also proposes a more directly phonological account of downtrend, in this case of downstepping intonations in English. In his model of shifting phrasal pitch registers, downstep is not a phonetic rule triggered by bitonal pitch accents, as proposed by Pierrehumbert (see Pierrehumbert 1980; Liberman and Pierrehumbert 1984; Beckman and Pierrehumbert 1986), but rather is the tonal interpretation of strength relationships among nodes in a prosodic constituent tree. In other words, the model formulates downstep as a global effect of tonal prominence, the abstract phonetic property that highlights one pitch-accented syllable over another. Ladd further argues that tonal prominence is generated and represented in the grammar as a pattern of relationships among nodes in a constituent tree. Thus, he rejects Pierrehumbert's (1980) argument against direct arboreal representations of downtrend; he proposes that downstep and other

prominence relationships are instead generated in the grammar as the tonal counterpart of the utterance's stress pattern, and that a direct phonological representation of downstep is therefore not redundant to the phonetic mapping rules. By thus claiming that downstep is a phonological prominence relationship, Ladd's model predicts that there is no distinction between an accent that is merely downstepped relative to the preceding accent and one that is both downstepped and subordinated in focus. As Pierrehumbert and Beckman (1988) point out, this distinction is essential to the correct prediction of Japanese intonation patterns. It will be interesting to see whether the relevant production experiments show such a fundamental difference between English and Japanese. Thus, Ladd's proposal provides experimentally falsifiable predictions, a defining feature of research in laboratory phonology. It also motivates a new series of experiments on the perception and production of prominence relationships in English, and a new point of comparison between stress-accent and non-stress-accent patterns.

Clements argues that there are devices other than a supraordinate phonological representation of register which may account as well for effects that Inkelas and Leben or Ladd attribute to their register tiers or trees. Clements also reminds us of the importance of striving for completeness and explicitness in models of phonetic implementation for tone structures.

Another pressing question in the area of intonation is how to predict the alignment between a tone and its associated tone-bearing unit. This is especially problematic in languages such as English, where the minimal tone-bearing units (syllables) have highly variable lengths due to the many different contextual prosodic features that affect segmental rhythms, including syllable weight, stress-foot organization, and phrasal position. Silverman and Pierrehumbert investigate this question for English prenuclear single-tone (H*) pitch accents. They review the possible phonological and phonetic mechanisms that might govern the alignment of the accent's tone to the syllable, and derive a set of predictions concerning the precise timing of a related F_0 event that can differentiate among the various mechanisms. They then present experimental evidence suggesting that the relationship cannot be accounted for as an artifact of aligning a tonal gesture that is invariant at a given rate to the beginning of a syllable whose duration varies with prosodic context. Nor can it be modeled as the consequence of a directly rhythmic phonological manipulation that adds discrete increments to the syllable's duration after the tone's alignment point. Instead, the timing of the F_0 event seems to be a combination of two phonetic rules: a rule of timing on the tonal tier that shifts the prenuclear accent tone in a stress clash leftward so as to ensure a minimum distance between it and the following nuclear accent tones, and a rule of coordination between tones and segments that computes when an accent's dominant tone will occur by referring to the time course of the sonority contour of the associated accented syllable.

Bruce's commentary compares Silverman and Pierrehumbert's findings to

similar data in Swedish which support their conclusion that coordination between tones and their associated syllables is described more accurately in the phonetic component than by building a direct phonological representation of the relevant rhythmic variation. He also points out that data such as these are relevant to the issue of tonal composition. The notation implies that the F_0 event corresponding to the starred tone of the pitch accent occurs within the accented syllable; yet the F_0 peak corresponding to the H* accent in Silverman and Pierrehumbert's experiment often occurred after the end of the stressed syllable. If the relevant F_0 event does not occur within the stressed syllable, should we revise our tonal analysis of the F_0 event? Precisely how can we apply such phonetic data to the determination of phonological form? Bruce concludes by suggesting that in stress-timed languages such as Swedish and English, coordination between tonal and segmental features is critical only at the boundaries of certain prosodic groupings larger than the syllable.

Kohler's paper addresses a closely related issue in the relationship between stress and accent, namely the differentiation of accent placement and accent composition in their contribution to the F_0 pattern. This is especially problematic in languages such as English and German, where the tonal composition of the pitch accent is not invariant (as in Danish), or lexically determined (as in Swedish), but rather is selected from an inventory of intonational shapes. In both languages, the combination of lexically contrastive accent placement with the different possible tonal shapes for the accent results in the possibility of different intonation patterns having superficially identical F_0 patterns. For example, when an intonation pattern with a peak aligned toward the end of the accented syllable is produced on a word that has primary stress on its first syllable, it may yield an F_0 contour that is virtually indistinguishable from that which occurs when another intonation pattern with a fall from a peak aligned toward the beginning of the accented syllable is produced on a word that has primary stress on its second syllable. Kohler asks how listeners can interpret such seemingly ambiguous structures; what cues do they use? He addresses this question with a series of perception tests involving hybrid resynthesis whereby the F_0 pattern from one intonation is combined with the segment durations and spectra of the opposing accent pattern. Kohler then turns his attention to the interaction of the more global determinants of F_0 contours with the more microscopic influences on F_0 of segmental features; specifically the fortis/lenis contrast in obstruents. He proposes that these microprosodic effects are not discernible in all intonational contexts and that listeners therefore cannot always employ them as cues to this contrast. As Silverman observes in his commentary on this paper, the results provide strong support for a hierarchical organization in which less prominent syllables at the lowest levels are marked absolutely as unstressed by their rhythmic and spectral characteristics, independent of the intonation pattern. Thus, the experiments provide new evidence for the phonetic reality of levels of stress unrelated to accent,

as maintained by phonologists. Regarding Kohler's data on the perception of segmental influences on F_0, Silverman presents contrary evidence arguing against the notion that segmental microprosodies are obliterated by the intonation contour in which the segments are embedded.

Beckman and Edwards's paper also addresses the relationship between intonation and stress, but from the point of view of prosodic constituency below the intonation phrase in English. They ask whether constituents within intonation phrases have overtly marked boundaries in addition to their clearly marked heads. They relate this question to the interpretation of two seemingly contradictory durational effects: the lengthening of syllables at the ends of phrases (indicative of edge-based prosodic constituency) and the shortening of stressed syllables in polysyllabic feet (indicative of head-based prosodic constituency). They present experiments designed to control for the second effect so as to examine the first effect alone. After demonstrating that there is a final lengthening internal to the intonation phrase, they attempt to determine its prosodic domain. Although the results provide no conclusive answer, they do suggest that prosodic constituency internal to the intonation phrase will not be adequately represented by a metrical grid.

In the first of three commentaries on this paper, Selkirk reviews the phonological literature on the relationship between syntactic constituency and prosodic edges, and contrasts it to Beckman and Edwards's "prominence-based" theory of prosodic constituency. She points out that this literature yields cross-linguistic generalizations about how syntactic structure is mapped onto prosodic structure. The languages reviewed include some which are unlike English in not having its phonologically-dictated culminative prominence of lexical stress. Selkirk presents experimental evidence of her own for isomorphism in the syntax-to-prosody mapping in one of these – namely, Tokyo Japanese. Beckman and Edwards fail to characterize the domain of final lengthening in English, she says, because they have stated the question entirely in terms of phonological prominences rather than in terms of the edges defined in the mapping from syntax. She reinterprets their results as evidence for syntactically motivated constituents such as the Prosodic Word and the Phonological Phrase proposed in her own earlier tree-based work (e.g. Selkirk 1981, 1984).

Fowler offers a different but related criticism of Beckman and Edwards. She points out that in distinguishing intonation phrase boundary effects from final lengthening internal to the phrase, Beckman and Edwards fail to control for syntactic constituency. She also reviews aspects of the literature on polysyllabic stress-foot shortening and stress clash that Beckman and Edwards neglect to consider, and in doing so indirectly raises another potential criticism of the experiments – namely, that if stress feet are bounded by word boundaries, as assumed in many versions of metrical phonology, then Beckman and Edwards have not really controlled for foot size. Finally she suggests an important

alternative interpretation of final lengthening that is skirted completely in the discussion of prominence-based constituents versus syntactic constituents. What if final lengthening is not a grammatical marking of linguistic constituents at all, but rather is a mere phonetic reflection of inertial braking at the boundaries of production units? Should we then expect these production units to be necessarily phonological constituents? More generally, how are we to know whether observed regularities indicate evidence of grammatical processes or of physical processes?

Cutler's commentary deals primarily with this last question and suggests possible grounds for answering it in comparisons of regularities in production and perception. Her concern addresses two levels at once: first, whether the phonological constituents proposed by Beckman and Edwards, if they exist at all, arise out of constraints on production or perception, and second, whether strong or weak psychological reality should be ascribed to them. If prosodic constituents can be shown to be actively involved in psychological processes of production (e.g. as a demonstrably necessary unit of planning), then they are psychologically real in the strong sense, but if the duration patterns evident in Beckman and Edwards's data are not necessary in planning or salient in perception, then the prosodic constituents are psychologically real only in the weak sense of being accurate generalizations of the effects.

The next group of papers addresses the question of the relationship between phonological representations and phonetic structures more generally. Hertz addresses the problem primarily as the practical question of appropriate tools. Can a computational framework be provided that allows phonologists to implement phonological representations and rules in computer programs that can easily be linked to the synthesis of corresponding phonetic structures? Her paper describes Delta, a programming language and synthesis rule-writing system that she has been developing for the last five years. She illustrates Delta by applying it to an autosegmental analysis of Bambara tone and a multilinear targets-and-transitions analysis of English formant patterns. While her paper emphasizes the practical advantages of Delta as a research tool, the system can be interpreted as making more far-reaching theoretical claims. The Delta structure makes no distinction between phonetic patterns and phonological patterns; it represents phonetic time functions as streams parallel to and essentially no different from the autosegmental tiers or organizational levels of the phonological representation for an utterance. Similarly, the system's rule framework makes no distinction among processes that relate one phonological level to another, processes that coordinate one phonetic time function to another, and processes that build phonetic structures on the basis of the phonological representation. A literal interpretation of these features of the Delta programming language amounts to the claim that there is an underlying unity between phonological and phonetic patterns which is more essential to the representation than any attested differences and incompatibilities between the two different types of linguistic representation. Directly synchronizing phonological