Journal of Phonetics (1990) **18**, 375–394

Phonological and phonetic representation

Janet Pierrehumbert

Northwestern University, Department of Linguistics, 2016 Sheridan Road, Evanston, IL 60201, U.S.A.

1. Introduction

Both phonology and phonetics have the aim of describing and explaining the sound pattern of human language. Insofar as both are exact theories, both must provide exact representations of the sounds of language. But the work done by these representations is different. Phonological representation is responsible for describing the qualitative contrasts in sound which can be used to convey qualitatively different meanings in any given language, or in all languages. The entities it posits are attributed to the mind of the speaker/listener, since this is where the association between sound and meaning takes place. Phonetic representation is responsible for describing speech as a physical phenomenon. That is, it covers measurable properties of articulation, acoustics and audition. Intellectual imperialists have sometimes taken the view that either phonology or phonetics is the whole story with respect to language sound structure. When coming from the side of phonology, they feel that phonetics is a relatively uninteresting subfield of biology and physics, which sheds no light on those aspects of the human mind that distinguish humans from beasts. Therefore, they argue, phonetics has no place in linguistics proper. Intellectual imperialists coming from the side of phonetics argue that the mental entities posited in phonology are not subject to rigorous scientific investigation. Therefore, they feel that phonetics provides a scientific theory of sound structure while phonology is a relatively uninteresting subfield of the humanities.

Neither of these positions is coherent. Both phonology and phonetics are necessary to understand language as a means of communication between people. If phonology is not related to phonetics, it models the mind of a solipsistic isolate. If phonetics is not related to phonology, it models noises and gestures to which no meaning or category structure can be assigned. A theory encompassing phonology, phonetics and their relation to each other is needed as a foundation for a theory of language processing and language acquisition. It is also needed for a model of historical sound changes, which typically originate in subphonological aspects of pronunciation and which are only of interest as they become shared in the speech community.

Therefore, this essay will take the view that phonetic representation cannot be discussed in isolation but only in conjunction with phonological representation. Section 2 will discuss the respective tasks of phonological and phonetic representation and the general character of the principles relating these representations. Section 3 discusses some successes of the field from the point of the strengths and shortcomings of the representations employed. Section 4 takes up the delicate issues

0095–4470/90/030375 + 20 $03.00/0

of qualitative representations in phonetics and quantitative representations in the mind.

2. Phonological and phonetic representation

2.1. *Phonological representation*

The starting point for phonological representation is the phonological principle underlying lexical inventories. Human languages do not use some arbitrary collection of noises to convey word meanings. Instead, a large number of words is created by combining a small number of elements, which are themselves meaningless. This principle is classically expressed as "the phonemic principle" and is one of the central ideas of structuralist linguistics (see for example Swadesh, 1934; Hockett, 1960). The subsequent displacement of the phoneme by distinctive features and autosegments does not change the main idea, but might lead us to refer instead to "the phonological principle". Only familiarity can blind us to its striking robustness and universality. Any theory of speech should have something explicit to say about it.

Some of the hallmarks of phonological representation follow from the phonological principle. First, phonological representation is concerned with speakers' implicit knowledge, that is, with information in the mind. This is the case because the association between a word's form and its meaning must have a cognitive status for language to function as it does as a means of communication. Second, phonology concerns qualitative rather than gradient distinctions, since complex forms are built up from a small number of distinct elements. Third, the rules which specify what complex forms are well-formed are syntactic, in the sense of formal language theory. Much as sentence grammar can be described by rules like (1), phonological structures can be described by rules like (2).

(1) S → NP VP.
(2) Syllable → (Onset) Rhyme (Appendix).

Although the symbols in (1) and (2) mean different things to the reader, the formal character of the rules is the same.

Phonology also encompasses those aspects of phrase and sentence sound structure which share these broad characteristics. Consider prosody and intonation. Phrasing and phrasal stress are included in the phonology by virtue of the fact that they display qualitative distinctions which are related to qualitative distinctions in interpretation. For example, shifting the phrasal stress can change truth conditions, as in the following sentences discussed in detail in Rooth (1985).

(3) I only said that Carl likes HERRING.
(4) I only said that CARL likes herring.

(3) is false if I also said that Carl likes some other kind of fish; (4) is false if I also said that someone else likes herring. A phonological status for distinctions in melody is supported by experiments such as Cruz–Ferreira (1983), which demonstrate conventional language particular pragmatic meanings for different melodies, and by experiments such as Pierrehumbert & Steele (1987, 1990) and Kohler (1987*a*, *b*),

which demonstrate the existence of melodic categories by finding categorical effects in pitch accent production and perception. In contrast to pitch accent, aspects of melody which appear to be gradient have a problematic status with respect to phonology. By "gradient", we mean that a continuous scale of phonetic values is related to some continuous dimension of expression or interpretation. An example is overall pitch range, found to behave gradiently in Liberman & Pierrehumbert (1984), and experimentally related to a dimension of interpretation in Silverman (1987). The problematic character of such cases underlines our understanding that the core of phonology involves qualitative distinctions which are manipulated by syntactic rules.

2.2. *Phonetic representation*

The starting point for linguistic phonetic theory is the fact that language in its most basic, universal and productive form has sound as its physical medium. The sounds are produced with the vocal apparatus and perceived with the ears. Extreme structuralists such as Hjelmslev attempted to treat this fact as incidental, arguing that language could be as well transmitted by semaphore as by sound and that the body of theory relating to the physical medium is accordingly not part of linguistics proper. However, such efforts are now widely recognized to be misdirected. They ignore both evidence that phonological categories are founded on articulation and acoustics (Jakobson, Fant & Halle, 1952; Stevens, 1972) and evidence that speech production and perception involve specialized use of the articulatory apparatus and auditory systems (Nelson, Perkell & Westbury, 1984; papers reviewed in Liberman & Mattingly, 1985).

The hallmarks of phonetic representation follow from the fact that sounds, as well as articulatory gestures and events in peripheral auditory processing, are observables in the physical world. Speech sounds can, like any other sounds, be recorded with microphones and described by physics. Similarly, activation and motion of the articulators, movements of the basilar membrane, and responses in the auditory nerves can all be recorded and physically modeled.

Representation at this level is not cognitive, because it concerns events in the world rather than events in the mind. The non-cognitive status of phonetic representation cuts even deeper; like other extremely familiar events and automatic behaviors, the phonetic domain is not well accessed by introspection. This is shown by experimental results such as those reported in Werker & Tees (1984*b*) Poser (1984) and Pierrehumbert & Beckman (1988). Werker & Tees found the adult listeners were unable to perceive differences which were subphonemic in their own language (though phonemic in other languages) even when directed to listen to speech as nonspeech. Poser (1984) and Pierrehumbert & Beckman (1988) report that tonal sequences generally transcribed as level by Japanese phonologists regularly display a fall–rise of 30–60 Hz. Differences in the phonetic domain are gradient rather than qualitative, since the articulators have continuous trajectories which give rise to continuously variable acoustic and psychoacoustic phenomena. The formal apparatus for describing complex forms is calculus rather than formal syntax, because physics is mathematically founded in calculus.

Like other physical theories, phonetics makes use of many different representations. This point can be readily appreciated by considering first, the number of different things it is of interest to measure (the sound waveform itself, the motions of all the articulators, the excitation of the articulators, the events in the ear itself, the neural events involved in auditory processing, etc.) and second, the number of different ways we may express or summarize the measured information (as a digitized waveform, as a spectrum, as a set of extrema or otherwise crucial values, etc.). Another battery of representations is introduced by efforts to abstract low level phonetic information in directions relevant to the linguistic system. For example, estimates of F_0, points of glottal closure, or segmental durations may be computed as sources of information about the prosodic system. Some representations are totally interconvertible with others. Two examples are the interconvertibility of time-domain and spectral domain representations, and the interconvertibility of linear prediction reflection coefficients with poles. Others are complementary. For example, formant trajectories provide information about resonances but not periodicity, while F_0 tracks provide information about periodicity but not resonances. Consideration of such cases helps to make the point that many different phonetic representations can be equally correct.

Some of the major controversies in phonetics reinforce rather than challenge the idea that multiple phonetic representations are both correct and useful. In the experiments which motivate the motor theory of speech perception (reviewed in Liberman & Mattingly, 1985), spectral representations are used in the controlled speech synthesis needed for rigorous experimentation. In the work of Stevens (1972, 1981, 1989), a model of the vocal tract as an acoustic tube is used to derive the acoustic regularities which are argued to provide a basis for phoneme systems. Thus none of these researchers claims that there is a unique correct phonetic representation—either articulatory or acoustic. Neither could even contribute to the debate without using both acoustic and articulatory representations. The issue under discussion is rather the respective roles of the acoustic and articulatory representations in speech perception and in constraining phonological inventories. Since phonological objects figure in this debate as targets of perception or explanation, how they are represented also has a crucial part to play.

Some readers may feel that this discussion of low level phonetic representation is excessively literal minded, involving surreptitious substitution of "data" or "measurements" for what the editor asked the authors to discuss. Certainly, many issues about intermediate representation have not yet been addressed. However, it is important to realize that data and measurements only exist by virtue of decisions about what to represent and how to represent it. Decisions about the most concrete levels of representation can have consequences for our understanding of entire aspects of sound structure. Careful thought about representation at the lowest level clarifies the relationship between phonology and phonetics by clarifying the character of the rules and intermediate representations needed. The following section discusses the general character of this relation. Section 4 takes up in more detail the representations used in the acoustic theory of speech production and in models of intonation as it determines fundamental frequency.

2.3. *The relation between phonological and phonetic representation*

The relationship of phonology to phonetics is profoundly affected by the fact that it involves disparate representations. Phonological representations are qualitative, cognitive, and relatively accessible to introspection; phonetic representations are quantitative, non-cognitive, and relatively inaccessible to introspection. Consider, by way of contrast, principles of phonetics proper or phonology proper which draw relationships among broadly comparable representations or in some cases describe regularities strictly within a single representation. For example, the acoustic theory of speech production relates the shape of acoustic tubes to spectra; both the spectra and the tubes are physical observables describable by continuous mathematics. The acoustic theory also asserts relationships within the spectrum *per se* by the constraints it places on formant values and bandwidths. The theory of English word stress, clearly a part of the phonology, relates the segmental string via the syllable structure to a hierarchical structure (the metrical structure) which represents the stress pattern (Liberman & Prince, 1977; Hayes, 1980, 1982; Halle & Vergnaud, 1988). Both the segmental string and the metrical structure are described using a small inventory of elements, which are claimed to represent mental entities, and whose behavior is controlled by a well-defined syntax. Phonological rules which change one segment into another in the course of word derivation and inflection modify the segmental representation rather than creating a different one. In the theory of Lexical Phonology, the principle of structure preservation requires that morphophonemic rules maintain the same type of representation, in a precise technical sense (Kiparsky, 1985; Kaisse & Shaw, 1985).

Because phonetic representations are quantitative rather than qualitative, the principles relating them to phonological representations are not insightfully viewed as syntactic rules. It is of course possible to view them as syntactic, since any continuous variation can be approximated with arbitrary precision by a sufficiently large set of discrete elements. Indeed, a waveform digitized at 16 bits uses an inventory of 65 000 different discrete elements to describe the sound pressure level at each particular time. Thus in principle, the well-formed sequences of sound pressure levels could be specified by rewrite rules similar to those in (1) and (2), with elements designating aspects of the phonological description appearing as non-terminals and elements designating sequences of sound pressure levels appearing as terminals. The point is that such a description would be both uninstructive and prohibitively complex. By exploiting the resources of continuous mathematics (or its discrete analog) we are able to make use of concepts in calculus (not to mention algebra and simple arithmetic) which are not available in formal language theory. The same point can of course be made with respect to any quantized phonetic parameter, such as F_0, jaw position, or F_1, as it relates to phonological categories.

A related point is that the principles cannot be viewed purely in terms of mental computations, since part of their job is to model physiology and physics. Speech synthesis algorithms, which can be viewed as more or less systematic and motivated implementations of our understanding of phonetic realization, compute a digitized waveform as the terminal phonetic representation; it would be unrealistic to suppose that people have cognitive control and awareness of speech on the time scale of a 0.1 ms. Since phonetic realization is not merely a matter of mental computation, ideas of simplicity and elegance appropriate for theories of mental computation may not be applicable. For example, it is clear that articulation involves nonlinearities both at the most peripheral level (e.g. contact of articulators leads to saturation) and at the control level (Abbs, Gracco & Cole, 1984). The best theory of articulation is thus likely to use non-linear mathematics even if it strikes us as more difficult or less elegant than linear mathematics.

If the principles relating phonology to phonetics are not syntactic, then what are they? They are semantic. That is, they have the same general character as the principles relating ordinary nouns or adjectives to their meanings in the real world. Now, analytical philosophy has a rather undifferentiated idea of the word. Consideration of the word's internal structure is needed to render coherent the idea that phonology and phonetics are semantically related. Let us take "dog" to mean the word as a whole without a view to its internal structure; DOG to mean the concept of dogs; /dɔg/ to mean the phonological form associated with this concept. The claim, then, is that the relationship between DOG and actual dogs is the analogous to the relationship between /dɔg/ and the actual pronunciations of this phonological form. That is, just as the semantic theory defines a collection of animals as the extension of DOG, it could define a collection of articulations and sounds as the extension of /dɔg/.

The mental association between /dɔg/ and DOG is arbitrary (that is, it arises through the cumulative effect of historical accidents), but the relation of each of these to the real world is not. This point may be appreciating by considering how children learn to talk. Children learning words such as "dog", "animal" or "red" acquire a mental association between these words and particular experiences of the real world. The semantic associations thus acquired are determined by properties of the world (What animals exist?), by the lexical inventory of the language (Is there a word for "dog"?) and (it is presumed) by cognitive constraints on category structure. Similarly, children learning to talk acquire a mental association between phonemes and particular real world experiences, namely particular types of events in speaking and hearing speech. For example, they learn that /p/ "means" lip closure, raised velum, rising formant transitions, and so on. The system of phonological categories children construct is constrained by the physical possibilities of articulation and speech acoustics, by the phonological inventory of the language being acquired, and by cognitive constraints on category structure (which may in this case be specifically linguistic). We should not be surprised if the relation of phonological elements to phonetics appears complex and baffling, since children themselves take eleven or twelve years to acquire fully adult patterns of variation (Kent, 1976). From the point of view of our present understanding, the representations of phonology and phonetics appear far apart. However, since we are scientists rather than mystics, we must labor in the belief that the relationship will in the end prove amenable to exact characterization (see also Pierrehumbert & Pierrehumbert, 1990).

Some of the most vexing aspects of the relation of phonology to phonetics are due to its semantic character. Let me mention three such aspects here: the failure of syntactic formalism; the pervasiveness of contextual effects; and difficulties with the idea of "object".

Syntactic formalism is both intuitive and mathematically well understood. Researchers familiar with its application in generative linguistics (that is, in syntax and phonology) can fluently propose new schemes for representation and analysis as they meet up with new data. As a result, efforts have been made to push a syntactic approach into the phonetic domain. The rules generating *n*-ary feature values in Chomsky & Halle (1968) represent one such effort. Such efforts are, I feel, fundamentally flawed; they inevitably fall short of saying anything about the relation of phonological entities to facts of the physical world. At the same time, they tend to spawn representational schemes for phonetics which have a dubious ontological status. One such scheme, IPA fine transcription, is discussed at more length below. However, the failure of a syntactic approach to phonetics leaves researchers somewhat adrift. Specifically, they need to invent phonetic representations of a profoundly novel character, and (as we will see) they need to invent many of them. Successes in this area, such as the acoustic theory of speech production, combine physical and linguistic insight with sophisticated mathematics. Therefore, I think it is critical to view phonetic representation as a difficult, ongoing research effort. Existing representations should not be taken for granted, and we should not expect that superficial representational ideas will work out.

Recent research on semantics emphasizes pervasive contextual or situational effects on meaning. No hard threshold separates "tall" from "short"; although a tall six-year-old is taller than a short six-year-old, he is probably shorter than a short basketball player. As even small children are aware, a cup that is "big" for a doll may be "small" for a human (Gelman & Ebeling, in press). Even something as directly related to the world as color perception is found on detailed experimentation to exhibit large and pervasive contextual dependencies. Experiments summarized in Land (1977) showed that because of the sophisticated use of context made in estimating reflectances (that is, in estimating color as a property of surfaces), exactly the same light spectrum can appear to be grey, red, yellow, blue or green.

The phonetic interpretation of phonological categories is context dependent in just the same way. We are all aware that similar phenomena may be observed in speech with respect to variation across speakers and dialects. For example, the lowest low tone for a female or child may be higher than the high tones of a low-pitched male. The extent to which such phenomena can be related to linguistic context within the speech of a single person is less widely appreciated. In languages with downstep, repeated applications of the rule can lower the F_0 values of H tones late in a phrase below those of L tones earlier in the phrase (Schachter & Fromkin, 1968; Pierrehumbert & Beckman, 1988). Utterance final /z/s may be identical in both duration and spectral properties to phrase-medial /s/s, although /z/s tend to be shorter and more voiced than /s/s in the same context. Pierrehumbert & Talkin (in press) report that effects of phrasal prosody cause acoustic indices of vocal fold abduction for /h/ to overlap values for vowels, within the speech of single individuals. However, /h/s quite reliably show more abduction than vowels in the same prosodic context.

Such observations show that the phonetic "meanings" cannot be supplied for phonological objects taken in isolation. On the contrary, a theory of the phonetic meaning of phonology will have to include an inventory of contextual effects and an account of how they interact. In Section 3.2 we will discuss particularly the issue of how syntagmatic structure affects the manifestations of paradigmatic contrasts.

A third consequence of the semantic character of the relationship between phonology and phonetics is that it is difficult to identify linguistic objects within the phonetics. Well-known problems include segmentation problems (e.g. where is the boundary between /a/ and /l/ in "all"?); problems of overlap (e.g. velar lowering for a nasal overlaps the preceding vowel) and dual attribution (the length of a vocalic region provides information both the phonological character of the vowel and that of a following obstruent). Such problems suggest that discrete elements do not really exist in the phonetics. Rather, we associate the phonological elements with various phenomena and events in the phonetics. Or, put differently, we explain the phonetic record by attributing a phonological analysis to it. This state of affairs is not peculiar to phonology and phonetics. It obtains generally in the relation of mental categories to facts about the world. We regularly manipulate concepts whose boundaries in time and space are vague—such concepts include not only conspicuously poorly demarcated objects such as "the fall of Rome", but also concrete physical objects such as "knee" or "river". Objects can overlap. Smoke may be part of an explosion, but also extend beyond the area of the explosion. Multiple attributions are also pervasive. A boondoggle is both work and a vacation; a tree can be part of "a windbreak" and also part of "the landscaping". In short, the division of the world into objects is implicitly and fuzzily defined by our analysis of it. We return to this point in Section 4 with an attack on discrete representations in phonetics.

3. Some successes and their lessons

3.1. *The acoustic theory of speech production*

The best understood aspect of the "semantics" of phonological representations is the relation of phonemes, or their distinctive feature representations, to dimensions of articulatory control and acoustic variation. Because of the notable successes of this theory, it is worthwhile to think about what kinds of representations and principles it employs, and also about what it leaves undone.

The acoustic theory of vowel production as presented in Fant (1960) models the semantics of vowel features by relating phonological features of height, backness and rounding to an idealization of the vocal tract, and by using this idealization to compute the resonances which are known to be perceptually important. Central to this theory are three disparate representations: the phonological one and two physical ones (one articulatory and one acoustic). In work on speech acoustics, the relation of the phonological representation to the vocal tract model is often treated rather informally; the phonology is used merely to identify a class of acoustic tubes which merit discussion. However, in cases where a completely explicit relationship is drawn, as in an articulatory synthesizer, we can see that the rules relate discrete categories on one side to quantitative specifications on the other. Continuous mathematics is used to relate the physical representations to each other. Additional representations manipulated in this theory include the full spectral representation, which can can be computed when assumptions are made about the source, and a time domain representation of the speech. The treatments of nasalization (Maeda, 1983) and interaction effects at the glottis (Ananthapadmanabha, 1982; Ananthapadmanabha & Fant, 1982; Fant, 1983; Lin, 1990) maintain essentially the same representational schema while advancing the mathematical treatment of individual components. The exact treatment of fricatives requires the crucial addition of an aerodynamic description (Shadle, 1985).

It is important to note that the phonological representation is not the only abstract one in this theory; the central physical representations are also very abstract. It is obviously an idealization to describe the vocal tract as a hard-walled and straight uniform tube. Computing plane-wave propagation only, disregarding cross-modes, has the further consequence that all tubes with the same relative cross-dimensions are taken to be equivalent, regardless of absolute cross-dimensions. This idealization is based on empirical results and scientific judgment about what frequency range is linguistically relevant. It is what makes possible the treatment of the vocal tract as an electrical transmission line. Although the relation of the formant frequencies and bandwidths to the spectrum is well-established, it is not at all trivial. It depends on a branch of mathematics which was developed only in the last two centuries and which only a minority of phoneticians actually commands. A characterization of the source is also required, and when the source is unknown the relation is not in general invertible; this is why estimating formants from spectra is often problematic. Insofar as formants approximate perceptually salient features of vowels, the abstraction from the speech signal to formants is a step along the path from the real world into the mind. The abstraction of vocal tract configuration to an electrical transmission line does not appear to have the same cognitive import. Two lessons may be drawn concerning the space of representations manipulated in a complete theory of language sound structure. First, quantitative representations can involve a significant degree of abstraction. Second, the dimension of abstractness is not the same as the dimension leading from the world to the mind, so that both physical and mental representations of varying degrees of abstractness may be useful.

The acoustic theory of speech production is a theory of paradigmatic relations both at the phonological level and in the phonetics. At the phonological level, it deals with phonemes or their distinctive feature characterization. That is, it concerns an inventory of elements which could be substituted for each other in the same position. Syntagmatic (or organizational) features such as stress and phrasing are outside of its domain. The concern with paradigmatic relations is carried into the phonetics through a mathematical treatment that assumes quasi-stationarity. That is, the mathematical treatment is a treatment of contrasting physical states. The theory is most successful in the cases where the quasi-stationary approximation is most appropriate, that is in treating vowels and other sonorants. This means that it provides an incomplete semantics in cases where a paradigmatic phonological distinction has intrinsically non-stationary phonetic manifestations.

The treatment of paradigmatic relations is also incomplete precisely because syntagmatic relations are not covered. The theory relates dimensions of linguistic contrast to dimensions of articulatory and acoustic variation. It does not say what value along these phonetic dimensions will manifest the linguistic contrast in any particular case. These values are known to differ systematically depending on context, but no complete and explicit theory of such variation has yet been developed. For example, consonants in the syllable coda tend to be more weakly produced than those in the onset. Formant values for vowels are influenced by stress (Fry, 1965) and by overall phrasal voice level (Schulman, 1989). Pierrehumbert & Talkin (in press) report that post-nuclear /h/s are lenited and that voiceless stops have a longer VOT after an intonational boundary. Any reader with extensive experience with continuous speech will be aware of many such effects.

Since modern speech synthesis algorithms do map phonological elements into actual values along phonetic dimensions, they have an implicit theory of syntagmatic effects. However, this theory is neither accurate nor comprehensive. First, present day systems permit syntagmatic structure above the word to affect pitch, amplitude and duration, but not other phonetic parameters. Formant values may be indirectly affected via smoothing algorithms; such effects are atheoretical, and other types of lenition are not modeled. Second, small scale effects are stored rather than modeled. That is, small scale effects are handled by storing syllables, demi-syllables, dyads or tabulated transitions.

The essential incompleteness of the paradigmatic theory was plain to its founders. For example, Jakobson, Fant & Halle (1952) emphasize the relational character of distinctive features in their discussion of the features' phonetic correlates. That is, the work identifies phonetic *dimensions* related to *dimensions* of linguistic contrast, and states that positive values for features do not designate absolute values along the phonetic dimensions, but rather contrastively greater values. The same point is made again recently in Fant's discussion of "relational invariance" (Fant, 1987). Thus we cannot blame the fathers of the paradigmatic theory for the fact that our understanding of what absolute values occur in what contexts is neither comprehensive nor exact. On the contrary, one might suggest that the implications of their work for the design of subsequent research were not sufficiently appreciated.

3.2. *Models of tone and intonation*

Models of tone and intonation and their relation to fundamental frequency provide an integrated treatment of paradigmatic and syntagmatic relations, though only in a subpart of the theory. This is the area in which the meaning of syntagmatic relations has been most fully worked out. It is therefore worthwhile to discuss the representations used in such models and their successes and shortcomings.

The intonation of languages like English has a paradigmatic component, because the speaker can choose for any phrase from an inventory of qualitatively different patterns, which convey different pragmatic meanings. As argued in Pierrehumbert (1980) and Pierrehumbert & Hirschberg (1990), the choices are actually more localized than the whole phrase. At each stressed syllable, the speaker can place one of the six English pitch accents, and furthermore intonational boundaries of two levels of strength can be marked with a boundary tone. The pitch accents and boundary tones can be viewed as intonational morphemes, if Pierrehumbert & Hirschberg are correct about their identifiable contributions to meaning. The "phonemes" of the system are the two tones, L and H, which occur singly and in pairs in the pitch accents and boundary tones. This is a very trivial inventory and can be represented with a single distinctive feature, ±Hightone. A two-tone inventory is also sufficient to describe the accent and intonation systems of Japanese (Poser, 1984; Pierrehumbert & Beckman, 1988) and Swedish (Bruce, 1977). Yip (1980) argues that the more complex tonal system of Mandarin is most insightfully described using two distinctive features, one representing register and one representing relative level within the register. In any case, all known tone languages use a rather small inventory of paradigmatic contrasts.

Syntagmatic relations are important both phonologically and phonetically. In all languages that have been studied in detail, the syntagmatic feature of phrasing controls the number and location of boundary tones. In languages which have stess (another syntagmatic feature), it too controls intonational elements. Furthermore, the phonetic realizations of these phonological elements depend on their positions. For example, a phonetic realization rule of English ("upstep") raises the F_0 value for an H tone which follows another H tone and is the boundary tone of an intonation phrase (Pierrehumbert, 1980). As a result, the F_0 values found at the end of yes/no questions are higher than those found anywhere else. The H tone in a Japanese pitch accent is realized at a higher F_0 value than a H tone marking the onset of an accentual phrase (Poser, 1984; Pierrehumbert & Beckman, 1988). Structural effects on tonal realization are also reported for Swedish in Bruce (1977) and for Chinese in Gårding (1985).

Fundamental frequency is taken to be the primary phonetic dimension expressing tonal contrasts. Algorithms for synthesizing F_0 from a phonological description compute four levels of phonetic description. First, the elements of the phonological description are mapped into critical events in a schematic F_0 contour, on the basis of their phonological type and their tonal and structural environment. In Bruce (1977), Pierrehumbert (1981), and Pierrehumbert & Beckman (1988), the critical F_0 events are F_0 target levels. In 't Hart & Cohen's (1973) model of Dutch, they are F_0 transitions with defined starting and ending points. The models of Fujisaki and colleagues (Fujisaki *et al.*, 1979; Fujisaki & Hirose, 1984) and of Anderson, Pierrehumbert & Liberman (1984) assign abstract shapes combining target and transitional specifications. In all cases, an abstract system of contrasts (whether tone levels or abstract transitions) is related by rule to specific F_0 values at specific points in time.[1]

The schematic F_0 contour is then used to compute an idealized contour by providing transitions between critical elements and in some models by applying a smoother. This converts the schematic contour into a continuously varying function, which might be viewed as the outcome for continuously voiced speech with no segmental effects. The algorithms then convert the ideal contour into an actual F_0 contour by adding some segmental effects. Minimally, voicing is suppressed during unvoiced obstruents. In Silverman (1987), extensive experimental results on segmental effects are also incorporated. Lastly, the F_0 contour is used to define a source function—that is, a train of impulses or more naturally shaped glottal pulses which can serve as input to the waveform synthesizer.

Such models provide a semantics for tone by describing the relationship between linguistic contrasts in tone and a quantitative phonetic representation. Several of the models have sufficient expressive power to synthesize a fairly accurate match for any naturally occurring F_0 pattern. Furthermore, the character of the mapping rules needed in the models has led to insights about relationship between phonology and phonetics in general. Pierrehumbert (1980), Liberman & Pierrehumbert (1984), and Pierrehumbert & Beckman (1988) propose constraints on how phonological structure can be accessed by phonetic realization rules. These constraints exploit the idea of locality which first motivated metrical and autosegmental phonology (Liberman & Prince, 1977; Goldsmith, 1979) and which has proved important in creating a theory of rule application (Ito, 1988). That is, the metrical and autosegmental aspects of the phonological representation define as adjacent a few things which would otherwise appear to be far apart, and thus support a limited

[1] In this discussion, I have singled out models which are fairly explicit about the inventory of contrasts in the language under consideration. Some F_0 models are described in such a procedural fashion that a phonological level of description cannot be discerned.

set of non-local interactions. The fact that phonetic realization rules appear to respect the same principles of locality means that they provide an important window into phonological structure.

The models have shortcomings on the phonetic side, particularly related to the lack of explicit articulatory representation. As discussed in the last section, the acoustic theory of speech production uses two phonetic representations, one articulatory and one acoustic. In contrast, intonation synthesis maps phonological categories directly into an acoustic parameter, F_0. This representation has an implicit articulatory interpretation, in terms of oscillations of the vocal folds. However when considered in detail, the relation of fundamental frequency to vocal fold oscillations is somewhat problematic. Furthermore, the medium scale shape of the F_0 contour (that is, shape at roughly the size of the metrical foot or the word) arises indirectly through the computation of a schematic contour followed by interpolation and smoothing procedures. No doubt these shapes in fact reflect the nature of laryngeal gestures, so that an articulatory treatment would be less *ad hoc*. Let us take up each of these points in more detail.

3.2.1. *Fundamental frequency*

Fundamental frequency is an idealized property of the speech signal whose relation to linguistic expression is rather indirect. Strictly speaking, F_0 is defined only for signals of infinite duration which show a pattern of exact periodic repetion. As discussed in Anderson (1986), this strict idea can be rigorously extended to a notion of quasi-fundamental-frequency which is related to quasi-periodicity. This mathematical construct is of course independent of the type of physical system responsible for the signal.

In phonetics, fundamental frequency has an interpretation which goes beyond its mathematical definition. That is, it is interpreted as a property of the source function. Two kinds of examples bring home this interpretion. First, failures of quasi-periodicity often arise from discontinuities in the transfer function, for example at the implosion of stops and nasals. In such cases quasi-F_0 is not well-defined mathematically, but phoneticians nonetheless complain that their pitch trackers have made errors. That is, they view lack of periodicity which arises from the transfer function as artifactual. Second, if the transfer function varied in a quasi-periodic fashion, as might happen during an alveolar or uvular trill, this frequency is not the one the phonetician wants. He wants a measure of periodicity related to voicing, even if it is less strong from a quantitative point of view. Such examples indicate that the "phonetic" ideal of fundamental frequency is strongly dependent on the appropriateness and success of source-filter decomposition. F_0 thus shares the abstractness of the source-filter characterization of speech in general. Indeed, some pitch tracking algorithms (such as cepstral and LPC residual methods) incorporate some effort to emphasize source properties and deemphasize filter properties in evaluating the periodicity of the signal, though they are not claimed to recover the source function exactly.

The interpretation of F_0 as a characteristic of the source function is also seen in speech synthesis. The computed F_0 contour does not control the periodicity of the speech signal directly, but is rather used to compute a source function which is filtered. It would be entirely possible for a mathematically optimal F_0 track of the output to fail to correspond at certain points to the input synthetic F_0 trajectory, if changes in the transfer function significantly affected the periodicity of the output. The most superficial phonetic representation of the intonation synthesis models, the impulse train or series of pulses which feeds the waveform synthesis, is in some ways the representation whose status is least shaky. This is the case because in actual speech production, a discontinuity in the source function is created when the vocal folds close. This discontinuity is the source of most of the high frequency energy which excites the formants. Thus, Anderson (1986) suggests that an abstraction of the source waveform as a sequence of critical discontinuities appears to be the most interpretable of the measures related to F_0.

However, a strong articulatory interpretation of F_0 in terms of regularly occurring points of glottal closure quickly comes up against cases in which the source function itself is not quasi-periodic. First, rapid transitions can give rise to poor correspondance from one glottal pulse to the next, simply because the glottal cycles are of such different durations. Second, breathy or glottalized voice qualities can produce such irregular glottal waveforms that a quasi-periodic interpretation is not appropriate. It is important not to disregard such source patterns as pathological; on the contrary, they arise as the limiting cases of generally used patterns of laryngeal control, and have a well-formed linguistic interpretation. In particular, glottalized voicing appears to be widely used and interpreted as if it represented an extra-low F_0 value, when in fact its F_0 appears not be well defined. Third, dissimilar glottal cycles can be found at the onset of voicing when non-initial glottal cycles show carry-over effects lacking on the first (Ananthapadmanabha, 1982). In all of these cases, however, the events at the glottis and the laryngeal states which give rise to them can be fruitfully discussed.

Such examples cast into doubt our sense that the idealized F_0 contour (which varies continuously but disregards segmental effects) is a well-defined level of phonetic representation. On the contrary, it appears to be a rough and handy, seat-of-the-pants, type of affair whose scientific status is hardly better than that of fine phonetic transcription, discussed below. Future work on phonetic representation for intonation needs to examine closely the dimensions of phonetic variation related to tonal contrasts and how they should be represented. Work on laryngeal modeling (Ishizaka & Flanagan, 1972; Titze & Talkin, 1979) makes important progress in this direction by defining more detailed representations and by separating the glottal control parameters, the physical mechanism, and the acoustic outcome. Such work needs to be informed by linguistic studies of what dimensions of variation are used in communication.

3.2.2. *Local shape*

Apart from the work of Fujisaki *et al.* intonation synthesis models have no pretensions about their treatment of laryngeal gestures.[2] In fact, the treatment of local shape is quite *ad hoc* due to the application of successive procedures in arriving

[2] Fujisaki *et al.* (1979) make much of the articulatory motivation for the second order critically damped filters they employ in smoothing. I do not feel this work is definitive, for several reasons. First, Tokyo Japanese does not present all of the descriptive problems of English since it apparently fails to use the active pitch lowering used in Osaka dialect and in English (Pierrehumbert & Beckman, 1988, discussing data from Kori, 1987). Second, even if this type of filter is important in the theory of motor control, it is unclear whether F_0 is an appropriate level of description for its application. Indeed probably not, because laryngeal models indicate that F_0 is a complex byproduct of the laryngeal control parameters. Third, this class of filters is very general and has considerable expressive power, so the approach is underconstrained.

at the F_0 contour. More specifically, local shapes arise as a consequence of (1) the time and F_0 location of events in the schematic contour, (2) how these events are connected to each other, and (3) what type of smoothing is employed. Furthermore, the theory of F_0 scaling rules (that is, rules defining what F_0 value is achieved for any given tone) is treated separately from local shape.

This approach has been surprisingly successful for H tones but has not been successfully extended to L tones. Specifically, no simple idea of L tone scaling has been found to work in synthesis, and the local shape in the vicinity of L tones has some peculiarities compared with H tones (see Pierrehumbert & Steele, 1990; Beckman & Pierrehumbert, in press). The success of a simple-minded treatment of H tones may be related to the fact that a single mechanism, crico-thyroid contraction, appears to have the main responsibility for H tone production. The mechanism for active F_0 lowering, such as is found for L pitch accents as well as at the end of declaratives, is complex and poorly understood. EMG studies suggest that a number of different muscles may be involved, and their respective roles is still unclear. Such observations suggest the need for an integrated treatment of F_0 scaling and local shape, which is founded on an understanding of laryngeal gestures.

4. Qualitative *vs*. Quantitative representation

4.1. *Discrete representations in phonetics*

Historically, there have been two types of effort to import the discreteness of phonology into phonetics. One is the fine phonetic transcription, as represented by IPA fine transcription. The other is the effort to identify acoustic segments. In my opinion, both of these efforts have involved significant confusion of representational issues and therefore require reappraisal.

4.1.1. *Phonetic transcription*

The fine phonetic transcription expands the symbolic inventory provided by the phonology in an effort to capture details of pronunciation which are systematic but not distinctive. For example, aspiration of voiceless stops in English would be represented in a phonetic transcription but not in a phonological one. Its nature is circumscribed by the following two claims, as expressed in Chomsky & Halle (1968). First, the inventory of phonetic symbols is universal. It specifies what distinctions in sound are linguistically relevant, or potentially contrastive. Universal extra-linguistic principles specify the meaning of the symbols in terms of articulations and sounds. Second, since the fine phonetic transcription is a symbolic representation, the rules which compute it from a more abstract phonological transcription are of the same general character as phonological rules. That is, they are syntactic rules which manipulate feature values; their power is only weakly augmented by the ability to assign *n*-ary instead of binary values.

A number of problems with fine transcription suggest that the representation it is claimed to provide is not a coherent one. First, practical transcription work often

comes up against ca
/l/ in "solider" go
without voicing in t
vowel deleted, with
questionable cases
difference drawn in
itative phonetic rep
can be viewed as th
Second, detailed
systematic and lang
& Beckman, 1988)
universal inventory
physiological comp
Third, there is
structures with bin
the evidence favors
quantitative values.
on F_0 values in do
be well modeled b
and to actual F_0 ta
phrase, computing
preceding F_0 targ
phonological chara
outcome in the im
sentation cannot se
modification of this
a model manipula
concerning interme
A similar point i
pattern of lenition
fold abduction sug
levels of prosodic s
phonetic transcripti
A fourth proble
There is no eviden
same kind of co
phonologically dist
contrasts which can
/b/; the experime
phonemic distincti
unnoticed. Werker
which is treated a
infants acquiring a
that the elements
representation in
elements in a discr
striking lack of d
amplified below.

comes up against cases where there is no way of deciding on the right answer. Is the /l/ in "solider" gone or merely extremely reduced? If "Topeka" is pronounced without voicing in the first syllable, should a devoiced vowel be recorded? Or is the vowel deleted, with the breathy region attributable to aspiration of the /t/? Such questionable cases especially tend to come up in connection with the categorical difference drawn in fine transcription between "something" and "nothing". Quantitative phonetic representations do not have the same problem, because "nothing" can be viewed as the limiting case for "less and less".

Second, detailed quantitative properties of speech have been found to exhibit systematic and language specific characteristics (e.g. Keating, 1985; Pierrehumbert & Beckman, 1988). Such results indicate that it will not be possible to identify a universal inventory of phonetic segments, which can serve as input to a universal physiological component.

Third, there is no evidence that the phonetic interpretation of phonological structures with binary feature values has a syntactic component. On the contrary, the evidence favors direct mapping of rather abstract phonological descriptions into quantitative values. This point is made in Pierrehumbert (1980) on the basis of data on F_0 values in downstepped intonation patterns in English. Downstep is found to be well modeled by a rule which refers to abstract phonological elements (tones) and to actual F_0 targets. The rule operates from the beginning to the end of the phrase, computing the pronunciation of each new tone as a function of the preceding F_0 target value. That is, the outcome is found to depend on the phonological characterization of the string and on the quantitative character of the outcome in the immediate past. It is clear that an intermediate symbolic representation cannot sensibly be posited for such a rule of pronunciation. A subsequent modification of this model in Poser (1984) and Pierrehumbert & Beckman (1988) to a model manipulating pitch registers does not affect the force of the arguments concerning intermediate symbolic representations.

A similar point is made in Pierrehumbert & Talkin (in press) with regard to the pattern of lenition for /h/ in English. Data on acoustic measures related to vocal fold abduction suggest that the degree of abduction is directly influenced by several levels of prosodic structure in a way which is not well described by a traditional fine phonetic transcription.

A fourth problem with fine phonetic transcription concerns its cognitive status. There is no evidence that the discrete elements posited in fine transcription have the same kind of cognitive status that has been experimentally established for phonologically distinctive elements. The classics of categorical perception involve contrasts which can distinguish between words, such as the contrast between /p/ and /b/; the experimental methodology indeed relies on the assumption that subphonemic distinctions, such as subphonemic contrasts in voice onset time, will pass unnoticed. Werker & Tees (1984*a*) have found that a potentially phonemic contrast which is treated as categorical by infants quickly loses its categorical status for infants acquiring a language in which it is not distinctive. Thus, there is no evidence that the elements of fine transcription can be viewed as elements of a discrete representation in the mind. However, it is also problematic to view them as elements in a discrete representation of events in the physical word, because of the striking lack of discreteness in physical measurements of speech. This point is amplified below.

In general, I would suggest that the fine phonetic transcription represents in many cases an attempt to record systematic consequences of prosody and intonation for details of pronunciation which are more properly treated quantitatively. Although fine transcription is a convenience for the researcher attempting a rough organization of his observations, it has been supplanted from a scientific point of view by the combination of improved representations of prosody and improved technology for handling quantitative measurements of speech.

4.1.2. *Acoustic segments*

The idea of the acoustic segment received a serious blow in the 1950s when speech synthesized by concatentation of phoneme-sized pieces was found to be totally unintelligible. Problems in defining acoustic segments arise for a number of reasons. The articulators move continuously, so that no segmental boundaries are apparent in sequences where no acoustic non-linearities create them. Such sequences can be as long as a whole sentence—e.g. "We were away a year ago". Coarticulation causes the gestures for neighboring phonemes to overlap, with consequent overlaps in acoustic evidence. Such overlaps can span several phonemes, as in the cases of lip rounding and nasalization. In addition, the apparent segment count is a function of the syntagmatic structure. For example, a vocalic sequence that is heard as a single syllable utterance finally can sound like two syllables (with a vowel–vowel juncture) utterance medially. This can happen because the listener's strong expections about phrase-final lengthening can cause him to disregard acoustic irregularities which he seizes on in a position where syllables are shorter.

It may be noted that the phonological position of the phoneme has also been eroded by autosegmental treatments of length and assimilation (see McCarthy, 1986; Hayes, 1986; Ito, 1988). Even with this approach, however, the phonology shows a far more segmental character than the acoustic signal, due to the quantization of abstract time provided by the *X* or CV tier. I would suggest, in agreement with Ohala (in press), that this quantization is not a physical property of the speech signal, but rather a mental construct encouraged by its weakly segmental character. That is, the acoustic evidence for any particular distinctive element tends to be localized (rather than occurring at arbitrary distances). Furthermore, acoustic non-linearities in many cases create psychologically salient demarcations in the signal, as discussed in recent papers by Stevens (1981, 1989).

4.2. *Quantitative representations in the mind*

The case for quantitative representations in the mind appears much stronger than the case for discrete representations of the speech signal. Progress needs to be made on formulating such representations and understanding their relationship to the qualitative representations of current phonological theories.

Some dimensions of phonetic variation are not treated categorically but are still systematically controlled to convey information. Probably the clearest such case is overall pitch range or voice level. Liberman & Pierrehumbert (1984) found that subjects could successfully follow instructions to speak in 10 different voice levels in randomized orders; no tendency to cluster productions in a small number of preferred voice levels was found, indicating that overall range or level does not exhibit the categorical distinctions found in phoneme production. This interpretation of the data is strengthened by the subsequent discovery of categorical effects in another area of the intonation system, namely the pitch accents (Pierrehumbert & Steele, 1987, 1990; Kohler, 1987*a*, *b*). However, range or level is used systematically to mark discourse structure, as reported in Silverman (1987) and Hirschberg & Pierrehumbert (1986). It also relates regularly to some types of nonlinguistic information, such as the distance between speaker and listener and the speaker's emotional state. Other stylistic aspects of speech, such as voice quality and overall precision of articulation are candidates for a similar status.

Liberman & Pierrehumbert (1984) and Pierrehumbert & Beckman (1988) represent systematic control of pitch range by quantitative decorations on the hierarchical phonological structure. For example, the pitch range of an intonation phrase can be represented by permitting the intonation phrase node in the metrical tree to carry a number which specifies its pitch range. This idea could be extended to cover other gradient phrasal characteristics by extending the list of quantitative attributes that nodes in the metrical tree can have. The proposal carries the prediction that the manipulation of such gradient parameters is synchronized with the organization imposed by the phonology. This prediction needs to be experimentally verified for other gradient parameters besides pitch range.

The discussion of fine phonetic transcription already mentioned findings that languages differ systematically in the phonetic interpretation of phonological categories. Well-known examples include variation in the degree of frontness for the front high unrounded vowel; the amount of vowel lengthening before voiced consonants; and the extent and timing of the F_0 perturbation induced by voiceless obstruents. Downstep, viewed as a phonetic interpretation rule for H tones, is a particularly interesting case. Some languages have the rule whereas others lack it. In some languages it lowers H in all H L H sequences whereas in others it only affects Hs with a particular status in the pitch accent system. Some languages lower H to the F_0 value of a preceding L (total downstep) whereas in others a downstepped H remains higher than a preceding L.

Such observations indicate that speakers of different languages have learned different rules for what their phonological categories mean, in terms of articulation and acoustics. That is, they have learned principles concerning the relationship of phonology to phonetics. These principles, as Section 2 explained, involve disparate representations and are semantic rather than syntactic in character. Section 2 treated them from the point of view of an external observer, the scientist seeking a theory of the phonological system in the speaker's mind and its relation to the physically observable phonetics. However, the existence of language particular phonetic interpretation rules in a sense puts every speaker in the position of the linguist, and implies that such rules are in some fashion represented in his mind. This implies that there is a mental representation of at least some quantitative aspects of phonetics.

The character of this mental representation of phonetics is not at all understood and it is important not to jump to mistaken conclusions about it. First, there are unlikely to be unified mental analogues of the phonetic interpretation rules that scientists now write. Rules mapping tones to F_0 target levels, or vowels to formant values, conflate mental computations with physiology and physics, and therefore represent a summary of numerous biological or psychological factors. Second, learned behavior is not the same as conscious behavior. Just because aspects of phonetics are learned does not mean they are accessible to introspection. Indeed highly practice motor activities such as speaking are typically inaccessible to introspection. Motor activities which are consciously controlled lack the same fluency and precision.

Historical changes in progress provide a third kind of evidence for quantitative cognitive representation of sound structure. Labov's studies (Labov, 1986) show that the shifts originate as subphonemic modifications of pronunciation which the speakers are not aware of consciously and may be unable to perceive. To model such changes from a psychological point of view, it is necessary to posit both a mental representation of subphonemic aspects of pronunciation, and a mechanism by which more frequently encountered variants can displace less frequently encountered ones. Note that this does not necessarily entail a direct representation of probabilities. Rather, the probabilistic effects could arise from incremental modification of the phonetic representations.

5. Conclusion

Phonetic representation is one of the most difficult problems in linguistics. We cannot arrive at a full understanding of language sound structure with the representations we have; it will be necessary to develop new ones all the time as we address different issues. Consideration of the best work in phonetics leads one to expect that successful representations will combine mathematical sophistication with physical and linguistic insight.

The nature of phonetic representations is circumscribed by their semantic relationship to phonology. That is, phonetics describes what phonological entities mean in terms of events in the world. Expectations about phonetics which ignore this relationship have not been fulfilled. Such failed hopes include the hope of finding discrete and easily identified phonetic objects; the hope of identifying phonological categories with absolute values of phonetic parameters; and the hope that phonetic representations can be computed by rules in the style of generative phonology.

On the other hand, much progress has been made in understanding what phonological entities mean phonetically. Some of the lessons of past successes should guide future work. Past work shows the value of multiple and well defined quantitative representations, and especially of parallel development of articulatory and acoustic representations. It also shows that explicit treatment of context and its effects is necessary to supply the actual phonetic values corresponding to phonological contrasts.

References

Abbs, J. H., Gracco, V. L. & Cole, K. J. (1984) Control of multimovement coordination: sensorimotor mechanisms in speech programming, *Journal of Motor Behaviour*, **16**, 195–231.

Ananthapadmanabha, T. V. (1982) Truncation and superposition, *Quarterly Status Report, Speech Transmission Laboratory*, 2–3, 1–17.

Ananthapadmanabha, T. V. & Fant, G. (1982) Calculation of true glottal flow and its components, *Speech Communication* **1**, 167–184.

Anderson, M. D. (1986) Pitch determination of speech signals. M.S. thesis, MIT.

Anderson, M. D., Pierrehumbert, J. & Liberman, M. Y. (1984) Synthesis by Rule of English intonation patterns. In *Proceedings of the IEEE International Congress on Acoustics, Speech and Signal Processing*, pp. 2.8.2–2.8.4.

Beckman, M. & Pierrehumbert, J. (in press) Tactics and strategies for thinking about F_0 variation. In

Papers in Laboratory
University Press.
Bruce, G. (1977) *Swedis*
Lund). CWK Gleeru
Chomsky, N. & Halle, M
Cruz-Ferreira, M. (1983
Ph.D. dissertation, U
Fant, G. (1960) *Acousti*
Fant, G. (1983) Prelimi
Report, Speech Trans
Fant, G. (1987) Interact
congress of phonetic s
Fry, D. B. (1965) The d
5th international cong
Fujisaki, H. & Hirose, I
sentences of Japanese
Fujisaki, H., Hirose, H.
declarative sentences
Phoniatrics, Universit
Gårding, E. (1985) Con
Phonetics, University
Gelman, S. A. & Ebelir
functional size, *Child*
Goldsmith, J. (1979) *Au*
Halle, M. & Vergnaud,
Hayes, B. (1980) A met
University Linguistics
Hayes, B. (1982) Extrar
Hayes, B. (1986) Inalter
Hirschberg, J. & Pierrel
24th meeting of the As
Computational Lingui
Hockett (1960) Logical
communication, (W.
Institute of Biological
Ishizaka, K. & Flanagar
cords, *Bell Systems T*
Ito, J. (1988) *Syllable th*
Jakobson, R., Fant, G.
Kaisse, E. & Shaw, P. (
Keating, P. (1985) Univ
Fromkin, editor pp. 1
Kent, R. D. (1976) Ana
acoustic studies, *J. Sp*
Kiparsky, P. (1985) Som
Kohler, K. J. (1987*a*) C
sciences, Talliin.
Kohler, K. J. (1987*b*) T
phonetic sciences, Tal
Kori, S. (1987) The ton
Laboratory, pp. 31–6
Columbus.
Labov, W. (1986) Sourc
& D. Klatt, editors).
Land, E. (1977) The ret
Liberman, A. & Mattin
Liberman, M. Y. & Pie
length, In *Language s*
Oehrle, R. T., editor
Liberman, M. Y. & Prir
Lin, Q. (1990) Speech p
Technology, Stockhol
Maeda, S. (1983) Corre
National d'Etudes de

Papers in Laboratory Phonology II (G. Docherty & D. R. Ladd, editors). London: Cambridge University Press.

Bruce, G. (1977) *Swedish word accents in sentence perspective* (*Travaux de l'institut de linguistique de Lund*). CWK Gleerup.

Chomsky, N. & Halle, M. (1968) *The sound pattern of English*. New York: Harper and Row.

Cruz-Ferreira, M. (1983) Nonnative comprehension patterns of intonation in Portuguese and English. Ph.D. dissertation, University of Manchester.

Fant, G. (1960) *Acoustic theory of speech production*. The Hague: Mouton.

Fant, G. (1983) Preliminaries to analysis of the human voice source. *Quarterly Progress and Status Report, Speech Transmission Laboratory*, 4, pp. 1–27.

Fant, G. (1987) Interactive phenomena in speech production. In *Proceedings of 11th international congress of phonetic sciences*, pp. 376–381.

Fry, D. B. (1965) The dependence of stress judgments on vowel formant structure. In *Proceedings of the 5th international congress of phonetic sciences*, pp. 306–311.

Fujisaki, H. & Hirose, H. (1984) Analysis of voice fundamental frequency contours for declarative sentences of Japanese, *Journal of the Acoustical Society of Japan*, **5**, 233–242.

Fujisaki, H., Hirose, H. & Ohta, K. (1979) Acoustic features of the fundamental frequency contours of declarative sentences in Japanese. In *Annual bulletin 13, Research Institute of Logopedics and Phoniatrics, University of Tokyo*, pp. 163–173.

Gårding, E. (1985) Constancy and variation in standard Chinese tonal patterns. Working Papers in Phonetics, University of Lund, pp. 19–51.

Gelman, S. A. & Ebeling, K. S. (in press) Children's use of nonegocentric standards in judgments of functional size, *Child Development*.

Goldsmith, J. (1979) *Autosegmental phonology*. New York: Garland.

Halle, M. & Vergnaud, J.-R. (1988) *An essay on stress*. Cambridge, MA: MIT Press.

Hayes, B. (1980) A metrical theory of stress rules. Ph.D. Dissertation, MIT. Distributed by Indiana University Linguistics Club, Bloomington.

Hayes, B. (1982) Extrametricality and English stress, *Linguistic Inquiry*, **13**, 227–276.

Hayes, B. (1986) Inalterability in CV phonology, *Language*, **62**, 321–351.

Hirschberg, J. & Pierrehumbert, J. (1986), Intonational structuring of discourse. In *Proceedings of the 24th meeting of the Association for Computational Linguistics*, pp. 136–144. Association for Computational Linguistics.

Hockett (1960) Logical considerations in the study of animal communication. In *Animal sounds and communication*, (*W. E. Lanyon & W. N. Tavolja, editors pp.* 392–430. *Washington DC: American Institute of Biological Sciences.*

Ishizaka, K. & Flanagan, J. L. (1972) Synthesis of voiced sounds from a two-mass model of the vocal cords, *Bell Systems Technical Journal* **51**, 1233–1268.

Ito, J. (1988) *Syllable theory in prosodic phonology*. New York: Garland.

Jakobson, R., Fant, G. & Halle, M. (1952) *Preliminaries to speech analysis*. Cambridge, MA: MIT Press.

Kaisse, E. & Shaw, P. (1985) On the theory of Lexical Phonology, *Phonology Yearbook 2*, 1–30.

Keating, P. (1985) Universal phonetics and the organization of grammars. In *Phonetic linguistics* (V. A. Fromkin, editor pp. 115–132. New York: Academic Press.

Kent, R. D. (1976) Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies, *J. Speech and Hearing Research* **19**, 421–447.

Kiparsky, P. (1985) Some consequences of lexical phonology, *Phonology Year Book*, **2**, 85–138.

Kohler, K. J. (1987*a*) Categorical pitch perception. In *Proceedings 11th international congress on phonetic sciences*, Talliin.

Kohler, K. J. (1987*b*) The linguistic functions of F_0 peaks. In *Proceedings 11th international congress on phonetic sciences*, Talliin.

Kori, S. (1987) The tonal behaviour of Osaka Japanese: an interim report. In *Papers from the Linguistics Laboratory*, pp. 31–61 (working papers 36). Department of Linguistics, Ohio State University, Columbus.

Labov, W. (1986) Sources of inherent variation, In *Invariance and variability in speech process* (J. Perkell & D. Klatt, editors). Hillsdale, NJ: Erlbaum.

Land, E. (1977) The retinex theory of color vision, *Scientific American* **237**(6), 108–128.

Liberman, A. & Mattingly, I. (1985) Motor theory of speech perception revised, *Cognition*, **21**, 1–36.

Liberman, M. Y. & Pierrehumbert, J. (1984) Intonational invariance under changes in ptich range and length, In *Language sound structure: studies in phonology presented to Morris Halle* (Aronoff, M. & Oehrle, R. T., editors), pp. 157–233. Cambridge, MA: MIT Press.

Liberman, M. Y. & Prince, A. (1977) On stress and linguistic rhythm, *Linguistic Inquiry*, **8**, 249–336.

Lin, Q. (1990) Speech production theory and articulatory synthesis. Ph.D. dissertation, Royal Institute of Technology, Stockholm.

Maeda, S. (1983) Correlats acoustiques de la nasalization des voyelles: Unie etude de simulation. Centre National d'Etudes des Telecommunications, Centre de Recherches de Lannion, Rept. VII.

McCarthy, J. (1986) OCP Effects: gemination and antigemination, *Linguistic Inquiry*, **17**, 207–263.
Nelson, W., Perkell, J. & Westbury, J. (1984) Mandible movements during increasingly rapid articulations of single syllables: Preliminary observations, *Journal of the Acoustical Society of America*, **75**, 945–51.
Ohala, J. (in press) The segment: primitive or derived? In *Papers in Laboratory Phonology II* (G. Docherty & D. R. Ladd, editors) Cambridge: Cambridge University Press.
Pierrehumbert, J. (1980) The phonology and phonetics of English intonation. Ph.D. dissertation, MIT. Available from Indiana University Linguistics Club, Bloomington, Indiana.
Pierrehumbert, J. (1981) Synthesizing intonation, *Journal of the Acoustical Society of America*, **70**(4), 985–995.
Pierrehumbert, J. (1990) A preliminary study of consequences of intonation for the voice source, *Quarterly Progress and Status Report, Speech Transmission Laboratory*.
Pierrehumbert, J. & Beckman, M. (1988) *Japanese Tone Structure*. Linguistic Inquiry Monograph Series 15, Cambridge, MA: MIT Press.
Pierrehumbert, J. & Hirschberg, J. (1990) The meaning of intonation contours in the interpretation of discourse. In *Intentions in Communication* (P. R. Cohen, J. Morgan & M. E. Pollack, editors), SDF benchmark series in computational linguistics, Cambridge, MA: MIT Press.
Pierrehumbert, J. B. & Pierrehumbert, R. T. (1990) On attributing grammars to dynamical systems, *Journal of Phonetics*, **18**, 465–477.
Pierrehumbert, J. & Steele, S. (1987) How many rise-fall-rise contours? In *Proceedings of the 11th international congress of phonetic sciences*, Tallinn.
Pierrehumbert, J. & Steele, S. (1990) Categories of tonal alignment in English, *Phonetica*, 181–196.
Pierrehumbert, J. & Talkin, D. (in press) Lenition of /h/ and glottal stop. In *Papers in Laboratory Phonology II*, (G. Docherty & D. R. Ladd, editors) London: Cambridge University Press.
Poser, W. (1984) The phonetics and phonology of tone and intonation in Japanese. Ph.D. dissertation, MIT.
Rooth, M. (1985) Association with focus. Ph.D. dissertation, Universities Massachussets Amherst.
Schachter, P. & Fromkin, V. (1968) A phonology of Akan: Akuapem, Asante and Fante. UCLA Working Papers in Phonetics No. 9.
Schulman, R. (1989) Articulatory dynamics of loud and normal speech, *Journal of the Acoustical Society of America*, **85**, 295–312.
Shadle, C. (1985) The acoustics of fricative consonants. Ph.D. dissertation, MIT.
Silverman, K. (1987) The structure and processing of fundamental frequency contours. Ph.D. dissertation, Cambridge University.
Stevens, K. N. (1972) The quantal nature of speech: Evidence from articulatory-acoustic data. In *Human communication: a unified view*. (E. E. David & P. B. Denes, editors) pp. 51–66. New York: McGraw-Hill.
Stevens, K. N. (1981) Constraints imposed by the auditory system on the properties used to classify speech sounds: Data from phonology, acoustics, and psychophysics, In *The cognitive representation of speech* (T. Myers, J. Laver & J. Anderson, editors). Amsterdam: North Holland.
Stevens, K. N. (1989) On the quantal nature of speech. *Journal of Phonetics*, **17**, 3–45.
Swadesh, M. (1934) The phonemic principle, *Language*, **10**, 117–129.
't Hart, J. H. & Cohen, A. (1973) Intonation by rule: a perceptual quest, *Journal of Phonetics*, **1**, 309–327.
Titze, I. R. & Talkin, D. T. (1979) A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation, *Journal of the Acoustical Society of American* **66**, 60–74.
Werker, J. S. & Tees, R. C. (1984*a*) Cross language speech perception: evidence for perceptual reorganization during the first year of life, *Infant Behavioural Development*, **7**, 49–63.
Werker, J. S. & Tees, R. C. (1984*b*) Phonemic and phonetic factors in adult cross language speech perception, *Journal of the Acoustical Society of America*, **75**, 1866–1878.
Yip, M. (1980) The tonal phonology of Chinese. Ph.D. dissertation, MIT.

Journal of Phonetics (1990

What is pho

Jørgen Rischel

Institute of Phonetics,

This paper addresse
here as the *specific*
arrangements of sym
questions can be a
epistemological statu
sound (production, a
at what level(s) of
phonological repres
highlighting the t
representations.

The present paper
on the said issues.
presents a strictly p
debate of past and p
theory about phon
segmental transcript
1986, 1987*a*), I wis
linguistic structure
phonological repres

I wish to start t
phonetic ("experin
phonological) appro
perhaps even a cari
phonologists have b
and means of scienc
research paradigms
interlock the findin
other. (I suppose t
phoneticians and p
developing a schola

Phoneticians hav
for centuries, but
component of "pra
This difficulty reflec
very young science
greater than ever,

0095–4470/90/030395 +