7
The segment: primitive or derived?

JOHN J. OHALA

7.1 Introduction

The segmental or articulated character of speech has been one of the cornerstones of phonology since its beginnings some two-and-a-half millennia ago.* Even though segments were broken down into component features, the temporal coordination of these features was still regarded as a given. Other common characteristics of the segment, not always made explicit, are that they have a roughly steady-state character (or that most of them do), and that they are created out of the same relatively small set of features used in various combinations.

Autosegmental phonology deviates somewhat from this by positing an underlying representation of speech which includes autonomous features (autosegments) uncoordinated with respect to each other or to a CV core or "skeleton" which is characterized as "timing units."[1] These autonomous features can undergo a variety of phonological processes on their own. Ultimately, of course, the various features become associated with given Cs or Vs in the CV skeleton. These associations or linkages are supposed to be governed by general principles, e.g. left-to-right mapping (Goldsmith 1976), the obligatory contour principle (Leben 1978), the shared feature convention (Steriade 1982). These principles of association are "general" in the sense

*I thank Björn Lindblom, Nick Clements, Larry Hyman, John Local, Maria-Josep Solé, and an anonymous reviewer for helpful comments on earlier versions of this paper. The program which computed the formant frequencies of the vocal-tract shapes in figure 7.2 was written by Ray Weitzman, based on earlier programs constructed by Lloyd Rice and Peter Ladefoged. A grant from the University of California Committee on Research enabled me to attend and present this paper in Edinburgh.

[1] As far as I have been able to tell, the terms "timing unit" or "timing slot" are just arbitrary labels. There is no justification to impute a temporal character to these entities. Rather, they are just "place holders" for the site of linkage that the autosegments eventually receive.

that they do not tak
(Chomsky and Halle
autosegments were [:
preserves something
this (auto)segment a
temporal coordinatic
(except insofar as it i
vowels or consonant

Is the primitive or
necessary? I suggest
"no" and "yes": "no
case after speech be
mental grammars of
have articulated spe
temporally coordina
separately the tempo
and the use of a sr
characteristics do nc
"chunk" in the stre
correspond in all po

For the evolution
based primarily on tl
simulation; an actua
models has not be
simulated and explor
Also relevant is a cor
(1982) and summar
discussed below. In a
known phonetic prin
made different from

7.2 I

Imagine a prespeech
ear (including their
vocal tract and the
constraints that they

[1] A computational imple
but was not successful
define more carefully tl

that they do not take into account the "intrinsic content" of the features (Chomsky and Halle 1968: 400ff.); the linkage would be the same whether the autosegments were [± nasal] or [± strident]. Thus autosegmental phonology preserves something of the traditional notion of segment in the CV-tier but this (auto)segment at the underlying level is no longer determined by the temporal coordination of various features. Rather, it is an abstract entity (except insofar as it is predestined to receive linkages with features proper to vowels or consonants).

Is the primitive or even half-primitive nature of the segment justified or necessary? I suggest that the answer to this question is paradoxically both "no" and "yes": "no" from an evolutionary point of view, but "yes" in every case after speech became fully developed; this latter naturally includes the mental grammars of all current speakers. I will argue that it is impossible to have articulated speech, i.e., with "segments," without having linked, i.e. temporally coordinated, features. However, it will be necessary to justify separately the temporal linkage of features, the existence of steady-states, and the use of a small set of basic features; it will turn out that these characteristics do not all occur in precisely the same temporal domain or "chunk" in the stream of speech. Thus the "segment" derived will not correspond in all points to the traditional notion of segment.

For the evolutionary part of my story I am only able to offer arguments based primarily on the plausibility of the expected outcome of a "gedanken" simulation; an actual simulation of the evolution of speech using computer models has not been done yet.[2] However, Lindblom (1984, 1989) has simulated and explored in detail some aspects of the scenario presented here. Also relevant is a comparison of speech-sound sequences done by Kawasaki (1982) and summarized by Ohala and Kawasaki (1984). These will be discussed below. In any case, much of my argument consists of bringing well-known phonetic principles to bear on the issue of how speech sounds can be made different from each other – the essential function of the speech code.

7.2 Evolutionary development of the segment

7.2.1 Initial conditions

Imagine a prespeech state in which all that existed was the vocal tract and the ear (including their neurological and neuromuscular underpinnings). The vocal tract and the ear would have the same physical and psychophysical constraints that they have now (and which presumably can be attributed to

[2] A computational implementation of the scenario described was attempted by Michelle Caisse but was not successful due to the enormity of the computations required and the need to define more carefully the concept of "segmentality," the expected outcome.

natural physical and physiological principles and the constraints of the ecological niche occupied by humans). We then assign the vocal tract the task of creating a vocabulary of a few hundred different utterances (words) which have the following properties:

1 They must be inherently robust acoustically, that is, easily differentiable from the acoustic background and also sufficiently different from each other. I will refer to both these properties as "distinctness". Usable measures of acoustic distinctness exist which are applicable to all-voiced speech with no discontinuities in its formant tracks; these have been applied to tasks comparable to that specified here (Kawasaki 1982; Lindblom 1984; Ohala *et al.* 1984). Of course, speech involves acoustic modulations in more than just spectral pattern; there are also modulations in amplitude, degree of periodicity, rate of periodicity (fundamental frequency), and perhaps other parameters that characterize voice quality. Ultimately all such modulations have to be taken into account.

2 Errors in reception are inevitable, so it would be desirable to have some means of error correction or error reduction incorporated into the code.

3 The rate and magnitude of movements of the vocal tract must operate within its own physical constraints and within the constraints of the ear to detect acoustic modulations. What I have in mind here is, first, the observation that the speech organs, although having no constraint on how slowly they can move, definitely have a constraint on how rapidly they can move. Furthermore, as with any muscular system, there is a trade-off between amplitude of movement and the speed of movement; the movements of speech typically seem to operate at a speed faster than that which would permit maximal amplitude of movement but much slower than the maximal rate of movement (Ohala 1981b, 1989). (See McCroskey 1957; Lindblom 1983; Lindblom and Lubker 1985 on energy expenditure during speech.) On the auditory side, there are limits to the magnitude of an optimal acoustic modulation, i.e., any change in a sound. Thus, as we know from numerous psychophysical studies, very slow changes are hardly noticeable and very rapid changes present a largely indistinguishable "blur" to the ear. There is some optimal range of rates of change in between these extremes (see Licklider and Miller 1951; Bertsch *et al.* 1956). Similar constraints govern the rate of modulations detectable by other sense modalities and show up in, e.g., the use of flashing lights to attract attention.

4 The words should be as short as possible (and we might also establish an upper limit on the length of a word, say, 1 sec.). This is designed to prevent a vocabulary where one word is /bə/, another /bəbə/, another /bəbəbə/ etc.,

with the longest wor
of the vocabulary.

7.2.2.1 Initially: ran
What will happen w
vocabulary? Notice
any other units asic
happens between sil
sequences of constr
vocal tract, which
duration. At the enc
robustness and disti
create a second canc
and a different, poss
sions and again app
was achieved. Real
nonrobust modulati
not hear them, etc.,
be seen in the loss c
pronunciations of E
sūn), and *ooze* (from
modulation created
larly weak in that it
1982).

My prediction is
necessary to its task,
out" as a matter
constraints. My reas

7.2.2.2 Temporal co
The first property th
tion between differer
or expansion) would
but the system woul
could create modul
distinct from other
A simple simulati
vowel space defined
this figure to the trad
male vowels report

with the longest word consisting of a sequence of n /bə/s where n = the size of the vocabulary.

7.2.2 Anticipated results

7.2.2.1 Initially: random constrictions and expansions

What will happen when such a system initially sets out to make the required vocabulary? Notice that no mention has been made of segments, syllables or any other units aside from "word" which in this case is simply whatever happens between silences. One might imagine that it would start by making sequences of constrictions and expansions randomly positioned along the vocal tract, which sequences had some initially chosen arbitrary time duration. At the end of this exercise it would apply its measure of accoustic robustness and distinctness and, if the result was unsatisfactory, proceed to create a second candidate vocabulary, now trying a different time duration and a different, possibly less random, sequence of constrictions and expansions and again apply its acoustic metric, and so on until the desired result was achieved. Realistically the evaluation of a vocabulary occurs when nonrobust modulations are weeded out because listeners confuse them, do not hear them, etc., and replace them by others. Something of this sort may be seen in the loss of [w] before back rounded vowels in the history of the pronunciations of English words *sword* (now [sɔɹd]), *swoon* (Middle English *sūn*), and *ooze* (from Old English *wōs*) (Dobson 1968: 979ff.). The acoustic modulation created when going from [w] to back rounded vowel is particularly weak in that it involves little change in acoustic parameters (Kawasaki 1982).

My prediction is that this system would "discover" that segments were necessary to its task, i.e. that segments, far from being primitives, would "fall out" as a matter of course, given the initial conditions and the task constraints. My reasons for this speculation are as follows.

7.2.2.2 Temporal coordination

The first property that would evolve is the necessity for temporal coordination between different articulators. A single articulatory gesture (constriction or expansion) would create an acoustic modulation of a certain magnitude, but the system would find that by coordinating two or more such gestures it could create modulations that were greater in magnitude and thus more distinct from other gestures.

A simple simulation will demonstrate this. Figure 7.1 shows a possible vowel space defined by the frequencies of the first two formants. To relate this figure to the traditonal vowel space, the peripheral vowels from the adult male vowels reported by Peterson and Barney (1952) are given as filled

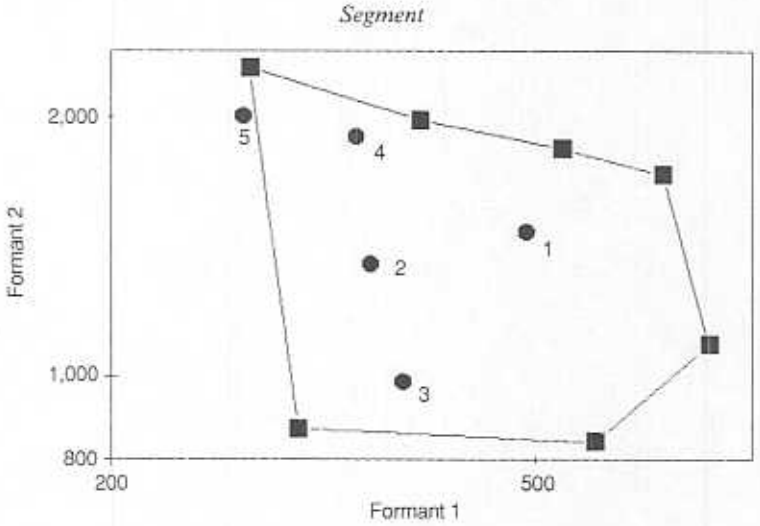

Figure 7.1 Vowel space with five hypothetical vowels corresponding to the vocal-tract configurations shown in figure 7.2. Abscissa: Formant 1; ordinate: Formant 2. For reference, the average peripheral vowels produced by adult male speakers, as reported by Peterson and Barney (1952) is shown by filled squares connected by solid lines.

squares connected by solid lines; hypothetical vowels produced by the shapes given in figure 7.2 are shown as filled circles. Note that the origin is in the lower left corner, thus placing high back vowels in the lower left, high front vowels in the upper left, and low vowels on the far right. Point 1 marks the formant frequencies produced by a uniform vocal tract of 17 cm length, i.e. with equal cross-dimensional area from glottis to lips. Such a tract is represented schematically as "1" in figure 7.2. A constriction at the lips (schematically represented as "2" in figure 7.2) would yield the vowel labelled 2. If vowel 1 is the "neutral" central vowel, then vowel 2 is somewhat higher and more back. Now if, simultaneous with the constriction in vowel 2, a second constriction were made one-third of the way up from the glottis, approximately in the uvular or upper pharyngeal region (shown schematically as "3" in figure 7.2) vowel 3 would result. This is considerably more back and higher than vowel 2. As is well known (Chiba and Kajiyama 1941; Fant 1960) we get this effect by placing constrictions at *both* of the nodes in the pressure standing wave (or equivalently, the antinodes in the velocity standing wave) for the second resonance of the vocal tract. (See also Ohala and Lorentz 1977; Ohala 1979a, 1985b.)

Consider another case. Vowel 4 results when a constriction is made in the palatal region. With respect to vowel 1, it is somewhat higher. But if a pharyngeal expansion is combined with the palatal constriction, as in vowel 5, we get a vowel that is, in fact, maximally front and high.

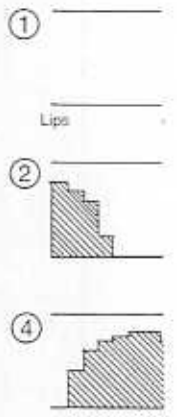

Figure 7.2 Five hypoth
positions in figure 7.1. Ve
tract length from glottis (

This system, I ma
necessary in order to
of acoustically distin
tions, but the same p
consonantal articula
cues for which are
signal) can reach e
Minimal amplitude
of the vocal cords b
arguments can be m

In fact, there is a g
distant and anator
together, presumably
For example, Riord
and larynx height, e
(1983) have discove
articulation: a voice
comparable voiceles
differing tension – n
between voiced and
obstruents have a gr

to the vocal-tract con-
ınt 2. For reference, the
by Peterson and Barney

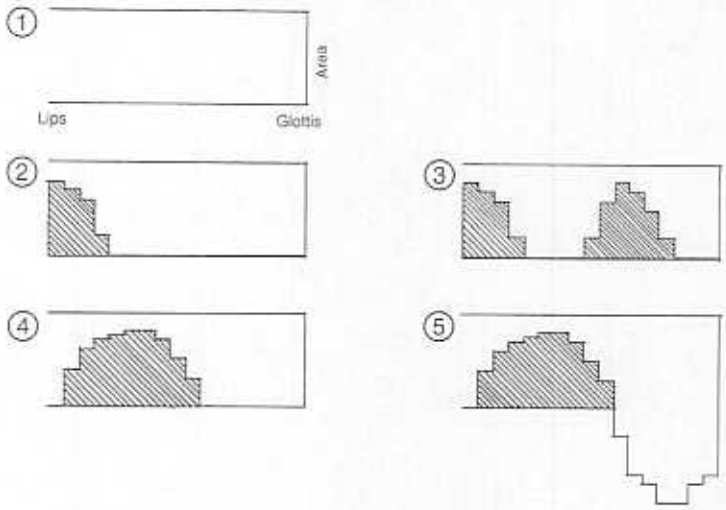

Figure 7.2 Five hypothetical vocal-tract shapes corresponding to the formant frequency positions in figure 7.1. Vertical axis: vocal-tract cross-dimensional area; horizontal axis: vocal-tract length from glottis (right) to lips (left). See text for further explanation

duced by the shapes
the origin is in the
wer left, high front
. Point 1 marks the
of 17 cm length, i.e.
s. Such a tract is
triction at the lips
l the vowel labelled
is somewhat higher
tion in vowel 2, a
o from the glottis,
(shown schemati-
considerably more
nd Kajiyama 1941;
oth of the nodes in
des in the velocity
ct. (See also Ohala

tion is made in the
t higher. But if a
iction, as in vowel
gh.

This system, I maintain, will discover that coordinated articulations are necessary in order to accomplish the task of making a vocabulary consisting of acoustically distinct modulations. This was illustrated with vowel articulations, but the same principle would apply even more obviously in the case of consonantal articulations. In general, manner distinctions (the most robust cues for which are modulations of the amplitude envelope of the speech signal) can reach extremes only by coordinating different articulators. Minimal amplitude during an oral constriction requires not only abduction of the vocal cords but also a firm seal at the velopharyngeal port. Similar arguments can be made for other classes of speech sounds.

In fact, there is a growing body of evidence that what might seem like quite distant and anatomically unlinked articulatory events actually work together, presumably in order to create an optimal acoustic–auditory signal. For example, Riordan (1977) discovered interactions between lip rounding and larynx height, especially for rounded vowels. Sawashima and Hirose (1983) have discovered different glottal states for different manners of articulation: a voiceless fricative apparently has a wider glottis than a comparable voiceless stop does. Löfqvist *et al.* (1989) find evidence of differing tension – not simply the degree of abduction – in the vocal cords between voiced and voiceless obstruents. It is well known, too, that voiceless obstruents have a greater closure duration than cognate voiced obstruents

(Lehiste 1970: 28); thus there is interaction between glottal state and the overall consonantal duration. The American English vowel [ɚ], which is characterized by the lowest third formant of any human vowel, has three constrictions: labial, mid-palatal, and pharyngeal (Uldall 1958; Delattre 1971; Ohala 1985b). These three locations are precisely the locations of the three antinodes of the third standing wave (the third resonance) of the vocal tract. In many languages the elevation of the soft palate in vowels is correlated with vowel height or, what is probably more to the point, inversely correlated with the first formant of the vowel (Lubker 1968; Fritzell 1969; Ohala 1975). There is much cross-linguistic evidence that [u]-like vowels are characterized not only by the obvious lip protrusion but also by a lowered larynx (*vis-à-vis* the larynx position for a low vowel like [ɑ]) (Ohala and Eukel 1987). Presumably, this lengthening of the vocal tract helps to keep the vowel resonances as low as possible and thus maximally distinct from other vowels.

As alluded to above, it is well known in sensory physiology that modulations of stimulus parameters elicit maximum response from the sensory receptor systems only if they occur at some optimal rate (in time or space, depending on the sense involved). A good *prima facie* case can be made that the speech events which come closest to satisfying this requirement for the auditory system are what are known as "transitions" or the boundaries between traditional segments, e.g. bursts, rapid changes in formants and amplitude, changes from silence to sound or from periodic to aperiodic excitation and vice versa. So all that has been argued for so far is that temporally coordinated gestures would evolve – including, perhaps, some acoustic events consisting of continuous trajectories through the vowel space, clockwise and counterclockwise loops, S-shaped loops, etc. These may not fully satisfy all of our requirements for the notion of "segment," so other factors, discussed below, must also come into play.

7.2.2.3 "Steady-state"

Regarding steady-state segments, several things need to be said. First of all, from an articulatory point of view there are few if any true steady-state postures adopted by the speech organs. However, due to the nonlinear mapping from articulation to aerodynamics and to acoustics there do exist near steady-states in these latter domains.[3] In most cases the reason for this

[3] Thus the claim, often encountered, that the speech signal is continuous, that is, shows few discontinuities and nothing approximating steady-states in between (e.g. Schane 1973: 3; Hyman 1975: 3), is exaggerated and misleading. The claim is largely true in the articulatory domain (though not in the aerodynamic domain). And it is true that in the perceptual domain the cues for separate segments or "phonemes" may overlap, but this by itself does not mean that the perceptual signal has no discontinuities. The claim is patently false in the acoustic domain as even a casual examination of spectrograms of speech reveals.

nonlinear relationship is not difficult to understand. Given the elasticity of the tissue and the inertia of the articulators, during a consonantal closing gesture the articulators continue to move even after complete closure is attained. Nevertheless, for as long as the complete closure lasts it effectively attenuates the output sound in a uniform way. Other parts of the vocal tract can be moving and still there will be little or no acoustic output to reveal it. Other nonlinearities govern the creation of steady-states or near-steady-states for other types of speech events (Stevens 1972, 1989).

But there may be another reason why steady-states would be included in the speech signal. Recall the task constraint that the code should include some means for error correction or error reduction. Benoit Mandelbrot (1954) has argued persuasively that any coded transmission subject to errors could effect error reduction or at least error limitation by having "breakpoints" in the transmission. Consider the consequences of the alternative, where everything transmitted in between silence constituted the individual cipher. An error affecting any part of that transmission would make the entire transmission erroneous. Imagine, for example, a Morse-code type of system which for each of the 16 million possible sentences that could be conveyed had a unique string of twenty-four dots and dashes. An error on even one of the dots and dashes would make the whole transmission fail. On the other hand if the transmission had breakpoints often enough, that is, places where what had been transmsitted so far could be decoded, then any error could be limited to that portion and it would not nullify the whole of the transmission. Checksums and other devices in digital communications are examples of this strategy. I think the steady-states that we find in speech, from 50 to 200 msec. or so in duration, constitute the necessary "dead" intervals or breakpoints that clearly demarcate the chunks with high information density. During these dead intervals the listener can decode these chunks and then get ready for the subsequent chunks. What I am calling "dead" intervals are, of course, not truly devoid of information but I would maintain that they transmit information at a demonstrably lower rate than that during the rapid acoustic modulations they separate. This, in fact, is the interpretation I give to the experimental results of Öhman (1966b) and Strange, Verbrugge, and Edman (1976).

It must be pointed out that if there is a high amount of redundancy in the code, which is certainly true of any human language's vocabulary, then the ability to localize an error of transmission allows error correction, too. Hearing "skrawberry" and knowing that there is no such word while there is a word *strawberry* allows us to correct a (probable) transmission error.

I believe that these chunks or bursts of high-density information flow are what we call "transitions" between phonemes. I would maintain that these are the kind of units required by the constraints of the communication task.

These are what the speaker is intending to produce when coordinating the movements of diverse articulators[4] and these are what the listener attends to.

Nevertheless, these are not equivalent to our traditional conception of the "segment." The units arrived at up to this point contain information on a sequential pair of traditional segments. Furthermore, the inventory of such units is larger than the inventory of traditional segments by an order of magnitude. Finally, what I have called the "dead interval" between these units is equivalent to the traditional segment (the precise boundaries may be somewhat ambiguous but that, in fact, corresponds to reality).

I think that our traditional conception of the segment arises from the fact that adjacent pairs of novel segments, i.e. transitions, are generally correlated. For example, the transition found in the sequence /ab/ is almost invariably followed by one of a restricted set of transitions, those characteristic of /bi/, /be/, /bu/, etc., but not /gi/, /de/. As it happens, this correlation between adjacent pairs of transitions arises because it is not so easy for our vocal tract to produce uncorrelated transitions: the articulator that makes a closure is usually the same one that breaks the closure. The traditional segment, then, is an entity constructed by speakers–listeners; it has a psychological reality based on the correlations that necessarily occur between successive pairs of the units that emerge from the underlying articulatory constraints.

The relationship between the acoustic signal, the transitions which require the close temporal coordination between articulators, and the traditional segments is represented schematically in figure 7.3.

7.2.2.4 Features

If an acoustically salient gesture is "discovered" by combining labial closure, velic elevation, and glottal abduction, will the same velic elevation and glottal abduction be "discovered" to work well with apical and dorsal closures? Plausibly, the system should also be able to discover how to "recycle" features, especially in the case of modulations made distinct by the combination of different "valves" in the vocal tract. There are, after all, very few options in this respect: glottis, velum, lips, and various actions of the tongue (see also Fujimura 1989b). A further limitation exists in the options available for modulating and controlling spectral pattern by virtue of the fact that the standing wave patterns of the lowest resonances have nodes and

[4] The gestures which produce these acoustic modulations may require not only temporal coordination between articulators but also precision in the articulatory movements themselves. This may correspond to what Fujimura (1986) calls "icebergs": patterns of temporally localized invariant articulatory gestures separated by periods where the gestures are more variable.

(a)

(b) Bit rate →

(c)

T

Figure 7.3 Relationshi
information transmissio
the traditional segment

antinodes at discre
and Kajiyama 194
pharynx would ser
palatal constriction
labial and uvular c
antinode in the pre

Having said thi
phonologists often
considered to be the
same velic coupling
one to create [m] a
consonants that res

[5] Pharyngeal expansion
but if it had been it
Peterson and Barney s

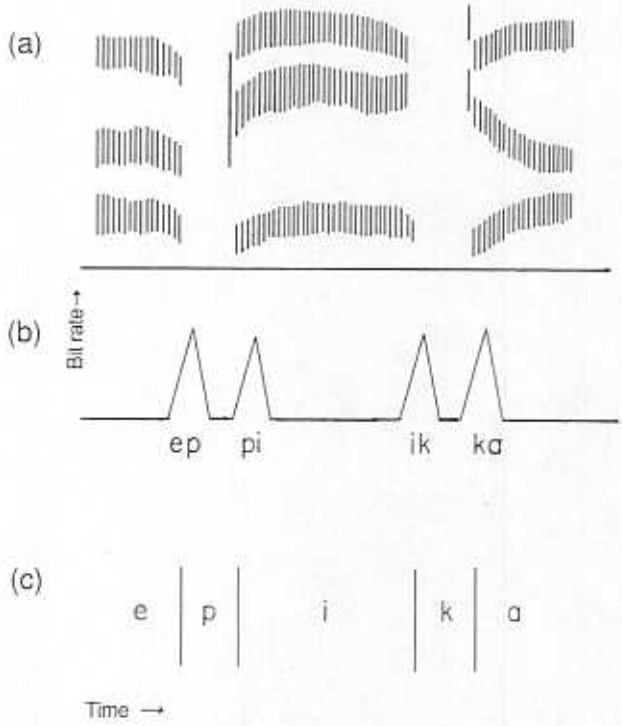

Figure 7.3 Relationship between acoustic speech signal (a), the units with high-rate-of-information transmission that require close temporal coordination between articulators (b), and the traditional segment (c)

antinodes at discrete and relatively few locations in the vocal tract (Chiba and Kajiyama 1941; Fant 1960; Stevens 1972, 1989): an expansion of the pharynx would serve to keep F_1 as low as possible when accompanying a palatal constriction (for an [i]) as well as when accompanying simultaneous labial and uvular constrictions (for an [u]) due to the presence there of an antinode in the pressure standing wave of the lowest resonance.[5]

Having said this, however, it would be well not to exaggerate (as phonologists often do) the similarity in state or function of what is considered to be the "same" feature when used with different segments. The same velic coupling will work about as well with a labial closure as an apical one to create [m] and [n] but as the closure gets further back the nasal consonants that result get progressively less consonantal. This is because an

[5] Pharyngeal expansion was not used in the implementation of the [u]-like vowel 3 in figure 7.1, but if it had been it would have approached more closely the corner vowel [u] from the Peterson and Barney study.

important element in the creation of a nasal consonant is the "cul-de-sac" resonating cavity branching off the pharyngeal–nasal resonating cavity. This "cul-de-sac" naturally gets shorter and acts less effectively as a separate cavity the further back the oral closure is (Fujimura 1962; Ohala 1975, 1979a, b; Ohala and Lorentz 1977). I believe this accounts for the lesser incidence or more restricted distribution of [ŋ] in the sound systems of the languages of the world[6]. Similarly, although a stop burst is generally a highly salient acoustic event, all stop bursts are not created equal. Velar and apical stop bursts have the advantage of a resonating cavity downstream which serves to reinforce their amplitude; this is missing in the case of labial stop bursts. Accordingly, among stops that rely heavily on bursts, i.e. voiceless stops (pulmonic or glottalic), the labial position is often unused, has a highly restricted distribution, or simply occurs less often in running speech (Wang and Crawford 1960; Gamkrelidze 1975; Maddieson 1984: ch. 2). The more one digs into such matters, the more differences are found in the "same" feature occurring in different segments: as mentioned above, Sawashima and Hirose have found differences in the character of glottal state during fricatives *vis-à-vis* cognate stops. The conclusion to draw from this is that what matters most in speech communication is making sounds which differ from each other; it is less important that these be made out of recombinations of the same gestures used in other segments. The orderly grid-like systems of oppositions among the sounds of a language which one finds especially in Prague School writings (Trubetzkoy 1939 [1969]) are quite elusive when examined phonetically. Instead, they usually exhibit subtle or occasionally not-so-subtle asymmetries. Whether one can make a case for symmetry phonologically is another matter but phonologists cannot simply assume that the symmetry is self-evident in the phonetic data.

7.2.2.5 Final comment on the preceding evolutionary scenario

I have offered plausibility arguments that some of the properties we commonly associate with the notion "segment," i.e. temporal coordination of articulators, steady-states, and use of a small set of combinable features, are derivable from physical and physiological constraints of the speaking and hearing mechanisms in combination with constraints of the task of forming a vocabulary. High bit-rate transitions separated by "dead" intervals are suggested to be the result of this effort. The traditional notion of the "segment" itself – which is associated with the intervals between the

[6] Given that [ŋ] is much less "consonantal" than other nasal consonants and given its long transitions (which it shares with any velar consonant), it is often more a nasalized velar glide or even a nasalized vowel. I think this is the reason it often shows up as an alternant of, or substitute for, nasalized vowels in coda position, e.g., in Japanese, Spanish, Vietnamese. See Ohala (1975).

transitions – is thought to be derived from the probabilities of cooccurrence of successive transitions. It is important to note that temporal coordination of articulators is a necessary property of the transitions not of the traditional segment.

The evolutionary scenario presented above is admittedly speculative. But the arguments regarding the necessity for temporal coordination of articulators in order to build a vocabulary exhibiting sufficient contrast are based on well-known phonetic principles and have already been demonstrated in numerous efforts at articulatory-based synthesis.

7.3 Interpretation

7.3.1 Segment is primitive now

If the outcome of this "gedanken" simulation is accepted, then it must also be accepted that spoken languages incorporate, indeed, are based on, the segment. To paraphrase Voltaire on God's existence: if segments did not exist, we would have invented them (and perhaps we did). Though not a primitive in the prespeech stage, it is a primitive now, that is, from the point of view of anyone having to learn to speak and to build a vocabulary. All of the arguments given above for why temporal coordination between articulators was necessary to *create* an optimal vocabulary would still apply for the *maintenance* of that vocabulary. I suggest, then, that autosegmental phonology's desegmentalization of speech, especially traditional segmental sounds (as opposed to traditional suprasegmentals) is misguided.

Attempts to link up the features or autosegments to the "time slots" in the CV tier by purely formal means (left-to-right association, etc.) are missing an important function of the segment. The linkages or coordination between features are there for an important purpose: to create contrasts, which contrasts exploit the capabilities of the speech apparatus by coordinating gestures at different locations within the vocal tract. Rather than being linked by purely formal means that take no account of the "intrinsic content" of the features, the linkages are determined by physical principles. *If features are linked of necessity then they are not autonomous.* (See also Kingston [1990]; Ohala [1990b].)

It is true that the anatomy and physiology of the vocal tract permit the coordination between articulators to be loose or tight. A relatively loose link is found especially between the laryngeal gestures which control the fundamental frequency (F_0) of voice and the other articulators. Not coincidentally, it was tone and intonation that were the first to be autosegmentalized (by the Greeks, perhaps; see below). Nevertheless, even F_0 modulations have to be coordinated with the other speech events. Phonologically, there are cases

where tone spreading is blocked by consonants known to perturb F_0 in certain ways (Ohala 1982 and references cited there). Phonetically, it has been demonstrated that the F_0 contours characteristic of word accent in Swedish are tailored in various ways so that they accommodate to the voiced portions of the syllables they appear with (Erikson and Alstermark 1972; Erikson 1973). Evidence of a related sort for the F_0 contours signaling stress in English has been provided by Steele and Liberman (1987).

Vowel harmony and nasal prosodies, frequently given autosegmental treatment, also do not show themselves to be completely independent of other articulations occurring in the vocal tract. Vowel harmony shows exceptions which depend on the phonetic character of particular vowels and consonants involved (Zimmer 1969; L. Anderson 1980). Vowel harmony that is presumably still purely phonetic (i.e., which has not yet become phonologized) is observable in various languages (Öhman 1966a; Yaeger 1975) but these vowel-on-vowel effects are (a) modulated by adjacent consonants and (b) are generally highly localized in their domain, being manifested just on that fraction of one vowel which is closest to another conditioning vowel. In this latter respect, vowel–vowel coarticulation shows the same temporal limitation characteristic of most assimilations: it does not spread over unlimited domains. (I discuss below how assimilations can enlarge their temporal domain through sound change, i.e. the phonologization of these short-span phonetic assimilations.) Assimilatory nasalization, another process that can develop into a trans-syllabic operation, is sensitive to whether the segments it passes through are continuant or not and, if noncontinuant, whether they have a constriction further forward of the uvula (Ohala 1983).

All of this, I maintain, gives evidence for the temporal linkage of features.[7]

7.3.2 Possible counterarguments

Let me anticipate some counterarguments to this position.

7.3.2.1 Feature geometry

It might be said that some (all?) of the interactions between features will be taken care of by so-called "feature geometry" (Clements 1985; McCarthy 1989) which purports to capture the corelatedness between features through a hierarchical structure of dependency relationships. These dependencies are said to be based on phonetic considerations. I am not optimistic that the interdependencies among features can be adequately represented by *any* network that posits a simple, asymmetric, transitive, dependency relationship

[7] Though less common, there are also cases where spreading nasalization is blocked by certain continuants, too; see Ohala (1974, 1975).

between features. Th
physical relationship
basis has been consi
spatial anatomical re
relations, and feature
latter domains link a
many that could be ci
cord vibration; a low
oral obstruents (if ar
influences the F_1 (heig
(such as /s/, /p/), if as
mimics nasalization a
tion; labial–velar segr
when they influence
involved, but they fr
consonants assimilat
secondary articulatio
have similar effects o
little effect on [u]) (O
Goldstein 1986; Wrig

I challenge the adv
crossing and occasion
transitive, relationshi
of other dependencie
subject to the funda
relabeling of what *d*
principle – that is, wi
happen. For example,
in vowels and not th
inhibit [+voice] inste

7.3.2.2 Grammar, no

Also, it might legitim
far have been from t
tations have been po
domain.[8] The autos
gathering and evalua

[8] There is actually consid
physical or psychologica
to both domains or neith
some events that are pr
forthcoming). But even
psychological domain ev

between features. The problem is that there exist many different types of physical relationships between the various features. Insofar as a phonetic basis has been considered in feature geometry, it is primarily only that of spatial anatomical relations. But there are also aerodynamic and acoustic relations, and feature geometry, as currently proposed, ignores these. These latter domains link anatomically distant structures. Some examples (among many that could be cited): simultaneous oral and velic closures inhibit vocal-cord vibration; a lowered soft palate not only inhibits frication and trills in oral obstruents (if articulated at or further forward of the uvula) but also influences the F_1 (height) of vowels; the glottal state of high airflow segments (such as /s/, /p/), if assimilated onto adjacent vowels, creates a condition that mimics nasalization and is apparently reinterpreted by listeners as nasalization; labial–velar segments like [w, kp] pattern with plain labials ([+anterior]) when they influence vowel quality or when frication or noise bursts are involved, but they frequently pattern like velars ([−anterior]) when nasal consonants assimilate to them; such articulatorily distant and disjoint secondary articulations as labialization, retroflexion, and pharyngealization have similar effects on high vowels (they centralize [i] maximally and have little effect on [u]) (Ohala 1976, 1978, 1983, 1985a, b; Beddor, Krakow, and Goldstein 1986; Wright 1986).

I challenge the advocates of "feature geometry" to represent such criss-crossing and occasionally bidirectional dependencies in terms of asymmetric, transitive, relationships. In any case, the attempt to explain these and a host of other dependencies other than by reference to phonetic principles will be subject to the fundamental criticism: even if one can devise a formal relabeling of what *does* happen in speech, one will not be able to show in principle – that is, without *ad hoc* stipulations – why certain patterns do *not* happen. For example, why should [+nasal] affect primarily the feature [high] in vowels and not the feature [back]? Why should [−continuant] [−nasal] inhibit [+voice] instead of [−voice]?

7.3.2.2 Grammar, not physics

Also, it might legitimately be objected that the arguments I have offered so far have been from the physical domain, whereas autosegmental representations have been posited for speakers' grammars, i.e. in the psychological domain.[8] The autosegmental literature has not expended much effort gathering and evaluating evidence on the psychological status of autoseg-

[8] There is actually considerable ambiguity in current phonological literature as to whether physical or psychological claims are being made or, indeed, whether the claims should apply to both domains or neither. I have argued in several papers that phonologists currently assign some events that are properly phonetic to the psychological domain (Ohala 1974, 1985b, forthcoming). But even this is not quite so damaging as assigning to the synchronic psychological domain events which properly belong to a language's history.

ments, but there is at least some anecdotal and experimental evidence that can be cited and it is not all absolutely inconsistent with the autosegmental position (though, I would maintain, it does not unambiguously support it either). Systematic investigation of the issues is necessary, though, before any confident conclusions may be drawn.

Even outside of linguistics analyzers of song and poetry have for millennia extracted metrical and prosodic structures from songs and poems. An elaborate vocabulary exists to describe these extracted prosodies, e.g. in the Western tradition the Greeks gave us terms and concepts such as *iamb, trochee, anapest*, etc. Although worth further study, it is not clear what implication this has for the psychological reality of autosegments. Linguistically naive (as well as linguistically sophisticated) speakers are liable to the reification fallacy. Like Plato, they are prone to regard abstract concepts as real entities. Fertility, war, learning, youth, and death are among the many fundamental abstract concepts that people have often hypostatized, sometimes in the form of specific deities. Yet, as we all know, these concepts only manifest themselves when linked with specific concrete people or objects. They cannot "float" as independent entities from one object to another. Though more prosaic than these (so to speak), is *iamb* any different? Are autosegments any different?

But even if we admit that ordinary speakers are able to form concepts of prosodic categories paralleling those in autosegmental phonology, no culture, to my knowledge, has shown an awareness of comparable concepts involving, say, nasal (to consider one feature often treated autosegmentally). That is, there is no vocabulary and no concept comparable to *iamb* and *trochee* for the opposite patterns of values for [nasal] in words like *dam* vs. *mid* or *mountain* vs. *damp*. The concepts and vocabulary that do exist in this domain concerning the manipulation of nonprosodic entities are things like *rhyme, alliteration*, and *assonance*, all of which involve the repetition of whole segments.

Somewhat more to the point, psychologically, is evidence from speech errors, word games and things like "tip of the tongue" (TOT) recall. Errors of stress placement and intonation contour do occur (Fromkin 1976; Cutler 1980), but they are often somewhat difficult to interpret. Is the error of *ambiguty* for the target *ambiguity* a grafting of the stress pattern from the morphologically related word *ambiguous* (which would mean that stress is an entity separable from the segments it sits on) or has the stem of this latter word itself intruded? Regarding the shifting of other features, including [nasal] and those for places of articulation, there is some controversy. Fromkin (1971) claimed there was evidence of feature interchange, but Shattuck-Hufnagel and Klatt (1979) say this is rare – usually whole bundles of features, i.e. phonemes, are what shift. Hombert (1986) has demonstrated

using word games that the tone and vowel length of words can, in some cases (but not all), be stripped off the segments they are normally realized on and materialized in new places. In general, though, word games show that it is almost invariably whole segments that are manipulated, not features. TOT recall (recall of some aspects of the pronunciation of a word without full retrieval of the word) frequently exhibits awareness of the prosodic character of the target word (including the number of syllables, as it happens; Brown and McNeill 1966; Browman 1978). Such evidence is suggestive but unfortunately, even when knowledge of features is demonstrated, it does not provide crucial evidence for differentiating between the psychological reality of traditional (mutually linked) features and autonomous features, i.e., autosegments. There is as yet no hitching post for autosegmental theory in this data.

7.3.2.3 Features migrate across segment boundary lines

If, as I maintain, there is temporal coordination between features in order to create contrasts, how do I account for the fact that features are observed to spill over onto adjacent segments as in assimilation? My answer to this is first to reemphasize that the temporal coordination occurs on the transitions, not necessarily during the traditional segment. Second, "coordination" does not imply that the participating articulators all change state simultaneously at segment boundaries, rather, that at the moment the rapid acoustic modulation, i.e. the transition, is to occur, e.g., onset of a postvocalic [ʒ], the various articulators have to be in specified *states*. These states can span two or more traditional segments. For a [ʒ] the soft palate must be elevated and the tongue must be elevated and bunched in the palatal region. Given the inertia of these articulators, these actions will necessarily have to start during the preceding vowel (if these postures had not already been attained). Typically, although many of these preparatory or perseveratory gestures leave some trace in the speech signal, listeners learn to discount them except at the moment when they contribute to a powerful acoustic modulation. The speech-perception literature provides many examples showing that listeners factor out such predictable details unless they are presented out of their normal context (i.e. where the conditioning environment has been deleted) (Ohala 1981b; Beddor, Krakow, and Goldstein 1986). My own pronunciation of "measure," with what I regard as the "monophthong" [ɛ] in the first syllable, has a very noticeable palatal glide at the end of that vowel which makes it resemble the diphthong [ej] in a word like "made." I do not perceive that vowel as [ej], presumably because when I "parse" [9] the signal I assign the palatal on-glide to the [ʒ], not the vowel. Other listeners may parse this glide with the vowel and thus one finds dialectally mergers of /ej/ and /ɛ/

[9] I use "parse" in the sense introduced by Fowler (1986).

before palato-alveolars, e.g., *spatial* and *special* become homophones. (See also Kawasaki [1986] regarding the perceptual "invisibility" of nasalization near nasal consonants.)

Thus, from a phonetic point of view such spill-over of articulatory gestures is well known (at least since the early physiological records of speech using the kymograph) and it is a constant and universal feature of speech, even before any sound change occurs which catches the attention of the linguist. Many features thus come "prespread," so to speak; they do not start unspread and then migrate to other segments. Such spill-over only affects the phonological interpretation of neighboring elements if a *sound change* occurs. I have presented evidence that sound change is a misapprehension or reinterpretation on the part of the listener (Ohala 1974, 1975, 1981b, 1985a, 1987, 1989). Along with this reinterpretation there may be some exaggeration of aspects of the original pronunciation, e.g. the slight nasalization on a vowel may now be heavy and longer. Under this view of sound change, no person, neither the speaker nor the listener, has implemented a change in the sense of having in their mental grammar a rule that states something like /ɛ/ → /ej/ /_/ʒ/; rather, the listener parses the signal in a way that differs from the way the speaker parses it. Similarly, if a reader misinterprets a carelessly handwritten "n" as the letter "u," we would not attribute to the writer or the reader the psychological act or intention characterized by the rule "n" → "u." Such a rule would just be a description of the event from the vantage point of an observer (a linguist?) outside the speaker's and listener's domains. In the case of sound patterns of language, however, we are now able to go beyond such external, "telescoped," descriptions of events and provide realistic, detailed, accounts in terms of the underlying mechanisms.

The migration of features is therefore not evidence for autosegmental representations and is not evidence capable of countering the claim that features are nonautonomous. There is no mental representation requiring unlinked or temporally uncoordinated features.

7.4 Conclusions

I have argued that features are so bound together due to physical principles and task constraints that if we started out with uncoordinated features they would have linked themselves of their own accord.[10] Claims that features can be unlinked have not been made with any evident awareness of the full phonetic complexity of speech, including not only the anatomical but also the aerodynamic and the acoustic–auditory principles governing it. Thus, more than twenty years after the defect was first pointed out, phonological

[10] Similar arguments are made under the heading of "feature enhancement" by Stevens, Keyser, and Kawasaki (1986) and Stevens and Keyser (1989).

representations stil
and Halle 1968: 40
kind of diachronic
features, one of th

What has been
possible to repres
eventually become
been shown that it
have been represe
autosegments. But
speech (indeed, we
doubt see several
apparent solar and
assumption of an
history to see that
mena is not justifi
than the lack of a c
the concept of aut
phenomena which i
of which was discu
does *not* occur in
monly. On the ot
reference to the fu
supported by expe
1990a).

In his paper "The s
calls a "plausibilit
phonological repre
other in a strict on
have called attentic
production, Ohala
requires an alterna
and steady-states. F

*Research for this pape
Science Foundation.

representations still fail to reflect the "intrinsic content" of speech (Chomsky and Halle 1968: 400ff.). They also suffer from a failure to consider fully the kind of diachronic scenario which could give rise to apparent "spreading" of features, one of the principal motivations for unlinked features.

What has been demonstrated in the autosegmental literature is that it is *possible* to represent speech-sound behavior using autosegments which eventually become associated with the slots in the CV skeleton. It has not been shown that it is *necessary* to do so. The same phonological phenomena have been represented adequately (though still not explanatorily) without autosegments. But there must be an infinity of possible ways to represent speech (indeed, we have seen several in the past twenty-five years and will no doubt see several more in the future); equally, it was *possible* to represent apparent solar and planetary motion with the Ptolemaic epicycles and the assumption of an earth-centered universe. But we do not have to relive history to see that simply being able to "save the appearances" of phenomena is not justification in itself for a theory. However, even more damaging than the lack of a compelling motivation for the use of autosegments, is that the concept of autosegments cannot explain the full range of phonological phenomena which involve interactions between features, a very small sample of which was discussed above. This includes the failure to account for what does *not* occur in phonological processes or which occurs much less commonly. On the other hand, I think phonological accounts which make reference to the full range of articulatory, acoustic, and auditory factors, supported by experiments, have a good track record in this regard (Ohala 1990a).

Comments on chapter 7

G. N. CLEMENTS

In his paper "The segment: primitive or derived?" Ohala constructs what he calls a "plausibility argument" for the view that there is no level of phonological representation in which features are not coordinated with each other in a strict one-to-one fashion.* In contrast to many phoneticians who have called attention to the high degree of overlap and slippage in speech production, Ohala argues that the optimal condition for speech perception requires an alternating sequence of precisely coordinated rapid transitions and steady-states. From this observation, he concludes that the phonological

*Research for this paper was supported in part by grant no. INT-8807437 from the National Science Foundation.

representations belonging to the mental grammars of speakers must consist of segments defined by sets of coordinated features. Ohala claims to find no phonetic or phonological motivation for theories (such as autosegmental phonology) that allow segments to be decomposed into sets of unordered features.

We can agree that there is something right about the view that precise coordination of certain articulatory events is conducive to the optimal transmission of speech. This view is expressed, in autosegmental phonology, in the association conventions whose primary function is to align features in surface representation that are not aligned in underlying representation. In fact, autosegmental phonology goes a step further than this and claims that linear feature alignment is a default condition on underlying representation as well. It has been argued that a phonological system is more highly valued to the extent that its underlying representations consist of uniformly linear sequences in which segments and features are aligned in a linear, one-to-one fashion (Clements and Goldsmith 1984). Departures from this type of regularity come at a cost, in the sense that the learner must identify each type of nonlinearity and enter it as a special statement in the grammar. Far from advocating a thoroughgoing "desegmentalization of speech," the position of autosegmental phonology has been a conservative one in this respect, compared to other nonlinear theories such as Firthian prosodic analysis, which have taken desegmentalization a good deal further.[11]

While we can agree that the coordination of features represents a default condition on phonological representation perhaps at all levels, this is not the whole story. The study of phonological systems shows clearly that features do not always align themselves into segments, and it is exactly for this reason that nonlinear representational systems such as autosegmental phonology have been developed. The principle empirical claim of autosegmental phonology is that phonological rules treat certain features and feature sets independently of others, in ways suggesting that features and segments are not always aligned in one-to-one fashion. The assignment of such features to independent tiers represents a direct and nonarbitrary way of expressing this functional independence. Crucial evidence for such feature autonomy comes from the operation of rules which map one structure into another: it is only when phonological features are "in motion," so to speak, that we can determine which features act together as units.

Ohala responds to this claim by arguing that the rules which have been proposed in support of such functional feature independence belong to physics or history, not grammar. In his view, natural languages do not have

[11] See, for example, Local (this volume). Recently, Archangeli (1988) has argued for full desegmentalization of underlying representations within a version of underspecification theory.

synchronic assimilation rules, though constraints on segment sequences may reflect the reinterpretation (or "misanalysis") of phonetic processes operating at earlier historical periods. If true, this argument seriously undermines the theory at its basis. But is it true? Do we have any criteria for determining when a detectable linguistic generalization is a synchronic rule?

This issue has been raised elsewhere by Ohala himself. He has frequently expressed the view that a grammar which aims at proposing a model of speaker competence must distinguish between regularities which the speaker is "aware" of, in the sense that they are used productively, and those that are present only for historical reasons, and which do not form part of the speaker's grammar (see e.g. Ohala 1974; Ohala and Jaeger 1986). In this view, the mere existence of a detectable regularity is not by itself evidence that it is incorporated into the mental grammar as a synchronic rule. If a regularity has the status of a rule, we expect it to meet what we might call the productivity standard: the rule should apply to new forms which the speaker has not previously encountered, or which cannot have been plausibly memorized.

Since its beginnings, autosegmental phonology has been based on the study of productive rules in just this sense, and has based its major theoretical findings on such rules. Thus, for example, in some of the earliest work, Leben (1973) showed that when Bambara words combine into larger phrases, their surface tone melody varies in regular ways. This result has been confirmed and extended in more recent work showing that the surface tone pattern of any word depends on tonal and grammatical information contributed by the sentence as a whole (Rialland and Badjimé 1989). Unless all such phrases are memorized, we must assume that unlinked "tone melodies" constitute an autonomous functional unit in the phonological composition of Bambara words.

Many other studies in autosegmental phonology are based on productive rules of this sort. The rules involved in the Igbo and Kikuyu tone systems, for example, apply across word boundaries and, in the case of Kikuyu, can affect multiple sequences of words at once. Unless we are willing to believe that entire sentences are listed in the lexicon, we are forced to conclude that the rules are productive, and part of the synchronic grammar.[12] Such evidence is not restricted to tonal phenomena. In Icelandic, preaspirated stops are created by the deletion of the supralaryngeal features of the first member of a geminate unaspirated stop and of the laryngeal features of the second. In his study of this phenomenon, Thráinsson (1978) shows at considerable length that it satisfies a variety of productivity criteria, and must be part of a synchronic grammar. In Luganda, the rules whose operation brings to light

[12] For Igbo see Goldsmith (1976), Clark (1990); for Kikuyu see Clements (1984) and references therein.

the striking phenomenon of "mora stability" apply not only within morphologically complex words but also across word boundaries. The independence of the CV skeleton in this language is confirmed not only by regular alternations, but also by the children's play language called Ludikya, in which the segmental content of syllables is reversed while length and tone remain constant (Clements 1986). In English, the rule of intrusive stop formation which inserts a brief [t] in words like *prince* provides evidence for treating the features characterizing oral occlusion as an autosegmental node in hierarchical feature representation (see Clements 1987 for discussion); the productivity of this rule has never been questioned, and is experimentally demonstrated in Ohala (1981a). In sum, the argumentation upon which autosegmental phonology is based has regularly met the productivity standard as Ohala and others have characterized it. We are a long way from the days when to show that a regularity represented a synchronic rule, it was considered sufficient just to note that it existed.

But even if we agree that autosegmental rules constitute a synchronic reality, a fundamental question still remains: if, as Ohala argues, linear coordination of features represents the optimal condition for speech perception, why do we find feature asynchrony at all? The reasons for this lie, at least in part, in the fact that phonological structure involves not only perceptually motivated constraints, but also articulatorily motivated constraints, as well as higher-order grammatical considerations. Phonology (in the large sense, including much of what is traditionally viewed as phonetics) is concerned with the mapping between abstract lexicosyntactic representations and their primary medium of expression, articulated speech. At one end of this mapping we find linguistic structures whose formal organization is hierarchical rather than linear, and at the other end we find complex interactions of articulators involving various degrees of neural and muscular synergy and inertia. Neither type of structure lends itself readily or insightfully to expression in terms of linear sequences of primitives (segments) or stacks of primitives (feature bundles).

In many cases, we find that feature asynchrony is regularly characteristic of phonological systems in which features and feature sets larger and smaller than the segment have a grammatical or morphological function. For instance, in languages where tone typically serves a grammatical function (as in Bantu languages, or many West African languages), we find a greater mismatch between underlying and surface representations than in languages where its function is largely lexical (as in Chinese). In Bambara, to take an example cited earlier, the floating low tone represents the definite article and the floating high tone is a phrasal-boundary marker; while in Igbo, the floating tone is the associative-construction marker. The well-known nonlinearities found in Semitic verb morphology are due to the fact that

consonants, vowels, and templates all play a separate grammatical role in the make-up of the word (McCarthy 1981). In many further cases, autosegmentalized features have the status of morpheme-level features, rather than segment-level features (to use terminology first suggested by Robert Vago). Thus in many vowel-harmony systems, the harmonic feature (palatality, ATR (Advanced Tongue Root), etc.) commutes at the level of the root or morpheme, not the segment. In Japanese, as Pierrehumbert and Beckman point out (1988), some tones characterize the morpheme while others characterize the phrase, a fact which these authors represent by linking each tone to the node it characterizes. What these and other examples suggest is that nonlinearities tend to arise in a system to the extent that certain subsets of features have a morphological or syntactic role independent of others. Other types of asynchronies between features appear to have articulatory motivation, reflecting the relative sluggishness of some articulators with respect to others (cf. intrusive stop formation, nasal harmonies, etc.), while others may have functional or perceptual motivation (the Icelandic preaspiration rule preserves the distinction between underlying aspirated and unaspirated geminate stops from the effects of a potentially neutralizing deaspiration rule, but it translates this distinction into one between preaspirated and unaspirated geminates). If all such asynchronies represent departures from the optimal or "default" state and in this way add to the formal complexity of a phonological representation, then many of the rules and principles of autosegmental phonology can be viewed as motivated by the general, overriding principle: reduce complexity.

Ohala argues that "feature geometry," a model which uses evidence from phonological rules to specify a hierarchical organization among features (Clements 1985; Sagey 1986a; McCarthy 1988), does not capture all observable phonetic dependencies among features, and is therefore incomplete. However, feature geometry captures a number of significant cross-linguistic generalizations that could not be captured in less structured feature systems, such as the fact that the features defining place of articulation commonly function as a unit in assimilation rules. True, it does not and cannot express certain further dependencies, such as the fact that labiality combines less optimally with stop production (as in [p]) than do apical or velar closure (as in [t] or [k]). But not all such generalizations form part of phonological grammars. Thus, phonologists have not discovered any tendency for rules to refer to the set of all stops except [p] as a natural class. On the contrary, in spite of its less efficient exploitation of vocal-tract mechanics, [p] consistently patterns with [t] and [k] in rules referring to oral stops, reflecting the general tendency of phonological systems to impose a symmetrical classification upon speech sounds sharing linguistically significant properties. If feature geometry attempted to derive all phonetic as well as phonological dependen-

cies from its formalism, it would fail to make correct predictions about cross-linguistically favored rule types, in this and many other cases.

Ohala rejects autosegmental phonology (and indeed all formal approaches to phonology) on the grounds that its formal principles may have an ultimate explanation in physics and psychoacoustics, and should therefore be superfluous. But this argument overlooks the fact that physics and psychology are extremely complex sciences, which are subject to multiple (and often conflicting) interpretations of phenomena in almost every area. Physics and psychology in their present state can shed light on some aspects of phonological systems, but, taken together, they are far from being able to offer the hard, falsifiable predictions that Ohala's reductionist program requires if it is to acquire the status of a predictive empirical theory. In particular, it is often difficult to determine on *a priori* grounds whether articulatory, aerodynamic, acoustic, or perceptual considerations play the predominant role in any given case, and these different perspectives often lead to conflicting expectations. It is just here that the necessity for formal models becomes apparent. The advantage of using formal models in linguistics (and other sciences) is that they can help us to formulate and test hypotheses within the domain of study even when we do not yet know what their ultimate explanation might be. If we abandoned our models on the grounds that we cannot yet explain and interpret them in terms of higher-level principles, as Ohala's program requires, we would make many types of discovery impossible. To take one familiar example: Newton could not explain the Law of Gravity to the satisfaction of his contemporaries, but he could give a mathematical statement of it – and this statement proved to have considerable explanatory power.

It is quite possible that, ultimately, all linguistic (and other cognitive) phenomena will be shown to be grounded in physical, biological, and psychological principles in the largest sense, and that what is specific to the language faculty may itself, as some have argued, have an evolutionary explanation. But this possibility does not commit us to a reductionist philosophy of linguistics. Indeed, it is only by constructing explicit, predictive formal or mathematical models that we can identify generalizations whose relations to language-external phenomena (if they exist) may one day become clear. This is the procedure of modern science, which has been described as follows by one theoretical physicist (Hawking 1988: 10): "A theory is a good theory if it satisfies two requirements: It must ultimately describe a large class of observations on the basis of a model that contains only a few arbitrary elements, and it must make definite predictions about the results of future observations." This view is just as applicable to linguistics and phonetics as it is to physics.

Ohala's paper contains many challenging and useful ideas, but it over-

states its case by a
coordination plays a
without concluding tl
tal. On the contrary
number of language
segmental features de
cal systems tolerate
representation. This
phonetic motivation
different methodolog
eses about the nature
help to illuminate t
providing a complet
structure and the bic

states its case by a considerable margin. We can agree that temporal coordination plays an important role in speech production and perception without concluding that phonological representations are uniformly segmental. On the contrary, the evidence from a wide and typologically diverse number of languages involving both "suprasegmental" and traditionally segmental features demonstrates massively and convincingly that phonological systems tolerate asynchronic relations among features at all levels of representation. This fact of phonological structure has both linguistic and phonetic motivation. Although phonetics and phonology follow partly different methodologies and may (as in this case) generate different hypotheses about the nature of phonological structure, the results of each approach help to illuminate the other, and take us further towards our goal of providing a complete theory of the relationship between discrete linguistic structure and the biophysical continuum which serves as its medium.