Prosodic phonology and phonetics*

John J. Ohala
University of California, Berkeley

Haruko Kawasaki
M.I.T.

1 Introduction: a parable

Our colleague Charles Fillmore invented the following parable (personal communication) to characterise the two dominant ways linguists attempt to solve problems of language structure and behaviour.

There is a high-tech restaurant where one orders food by punching buttons on a keyboard-like menu and the food comes to the table by a conveyor belt which emerges from a little trap door in the wall. It is therefore not possible to look directly into the kitchen to see how the food is prepared nor is it possible to interrogate an employee about this. Open-face sandwiches are one of the restaurant's specialities. Thus, one can order plain bread, bread with peanut butter, bread with mayonnaise, and even bread with peanut butter with mayonnaise on top of that. However, when one tries to order bread with mayonnaise with peanut butter on top of that nothing emerges from the kitchen. How can one account for this pattern of possible sandwiches? There are two ways to answer this question. One, the 'formal' approach, attempts to construct a 'grammar' of open-face sandwiches. A grammar in this sense means a theory which will specify or generate all and only the possible sandwich types and none of the impossible sandwich types. Such a formal grammar might be something like the following (where '→' means 'is manifested as'):

\[
\begin{align*}
(a) & \quad \text{Sandwich} \rightarrow \text{Slice of bread} + (\text{spread}) \\
(b) & \quad \text{(optional) Spread} \rightarrow \text{Peanut butter} \\
(c) & \quad \text{(optional) Spread} \rightarrow \text{Mayonnaise}
\end{align*}
\]

The ordering of the rules is crucial, of course, and accounts for the fact that one cannot order spreads of peanut butter and/or mayonnaise without bread and why peanut butter cannot be applied after mayonnaise. The other approach, the 'substantive' one, looks for an answer to the question by examining the properties and inherent constraints of the ingredients of the sandwiches. By this approach it might be found that, among other things, substance A can be applied as a spread to substance B if A is more deformable than B and if the surface of B exhibits sufficient friction (or viscous drag) and both surfaces exhibit sufficient adhesion.
There is no question of one approach being 'right' and the other 'wrong'; both are right and both accomplish the principal aim of all scientific inquiry of making the universe a little bit less mysterious. Each approach has its advantages and disadvantages. A formal analysis of some sort is not that difficult to create, given a little ingenuity and imagination. Some formal accounts may be more elegant than others and, as we know, the competition between purely formal accounts of the same body of data is thus a continuous process. But in the end purely formal analyses are just systematisations of the data, not explanations of it. So, in fact, did Newton himself regard his own formalisation of the principles of free fall and the orbits of satellites. They represent a useful intermediate stage on the way towards our ultimate understanding of how the universe works. The substantive answer (which may also be expressed formally) allows a higher level of understanding and has a greater capacity for prediction and generalisation. For example, the substantive approach but not the formal approach would permit one to predict (after ascertaining the appropriate physical properties) that one also could not apply peanut butter to jelly, mashed bananas, or ice cream (unless it was very cold, but in that case the ice cream could not be spread on the bread). With some care, however, it might be possible to apply peanut butter or liver paste in either order (if one wished to do so). The substantive theory could also show that the same principles governing sandwich construction apply in other domains (e.g. the application of paint to walls, ink to paper, asphalt to roads, etc.).

There is, of course, no guarantee that an adequate substantive solution will always be discovered; this is scientific discovery and cannot be produced on command. Even lacking a full substantive theory it may at least be possible to point to some substantive facts which provide helpful clues which will lead to such a theory.¹

2 Prosodic phonology needs an empirical base

Our understanding of how speech sounds behave has been enriched by several studies produced within the past decade or so which account for important sound patterns by reference to structure above the phonological segment (e.g. Liberman 1975; Goldsmith 1976; Kahn 1976; Liberman & Prince 1977; McCarthy 1979, 1981; Halle & Vergnaud 1980; Selkirk 1980; Hayes 1981; Steriade 1982; Clements & Keyser 1983). Following Selkirk, we will refer to these collectively as PROSODIC PHONOLING. These tend to be, however, primarily formal accounts and one would hope that the facts which are so admirably encompassed by these models will eventually be explained by reference to substantive elements, whether these be phonetic or psychological, synchronic or diachronic. Substantive evidence is especially necessary for these models since, in general, the main argument in their favour at present is limited to the claim that they make it possible to represent phonological facts in a way judged to be more insightful than alternative ways, i.e. that they make diverse phonological facts 'hang together' better. There is fulness, like beauty, is in the eye of the beholder (Hudson 1982). Even a case that the criterion of making a sufficient test of a theory, namely, the ability to account for the sub-lunar universe, is not yet accounted for by phonology, virtually any work in phonology in this paper we offer prosodic phonology added the asymmetrical behaviour to the vowels, correlating with metrical 'treet', usually couched in terms of

3 Syllable onset vs. coda

An important point for pre is the fact that the phonological-initial position as Hooper 1976: 195ff; Foley 1979 have more distinct syllable onsets in some languages permit no this to the observation that diachronically in onsets the syllabification in languages sequence of the sort VVC stores may try to associate second syllable in order to first assigning any (Clements & Kent). Thus not *[ekst.ra], etc., even sequences in English. It has syllable onsets contribute to the stress or accent these asymmetries exist? A There is good reason to and summary, however, whether might be derivable from an A perceptual phenomena phonological facts is that centre) (Rapp 1971; All 

Marcus 1981). When subje
being 'right' and the other
ish the principal aim of all
be bit less mysterious. Each
A formal analysis of some
ingenious and imagination.
an others and, as we know,
streams of the same body of data
real formal analyses are just
it. So, in fact, did Newton
principles of free fall and the
mediate stage on the way
the universe works. The
ed formally) allows a higher
spatiality for prediction and
approach but not the formal
scattering the appropriate
paly peanut butter to jelly,
ry cold, but in that case the
With some care, however,
river paste in either order
ly could also show that the
ion apply in other domains
iper, asphalt to roads, etc.),
te substantive solution will
ry and cannot be produced.
 theory it may at least be
hich provide helpful clues

irical base
have been enriched by
de or so which account for
are above the phonological
Kahn 1976; Liberman &
rnaud 1980; Selkirk 1980;
1983). Following Selkirk,
phonology. These tend to
would hope that the facts
models will eventually be
whether these be phonetic
Substantive evidence is
genral, the main argument
that they make it possible
to be more insightful than
phonological facts 'hang
together' better. There is abundant evidence, however, that 'insight-
fulness', like beauty, is in the eye of the beholder (cf. McCarthy 1981
and
Hudson 1982). Even a casual glance at the history of science reveals to
us that the criterion of making facts 'hang together' is a necessary but
lled a sufficient test of a theory. This is evident from an
Greek theory - now regarded as 'quaint' - that the ultimate constituents
of the sub-lunar universe were earth, air, fire, and water, a theory which
nevertheless 'accounted' for an impressive variety of facts. In fact, we
don't have to go so far afield; we can learn the same lesson from reading
virtually any work in phonology published more than 15 years ago.

In this paper we offer some preliminary remarks which might give
prosodic phonology added empirical anchoring. We address three topics:
the asymmetrical behaviour of syllable onsets vs. codas, possible temporal
correlates of metrical 'trees', and the constraints on syllable formation
usually couched in terms of the 'sonority hierarchy'.

3 Syllable onset vs. coda
An important point for prosodic phonology (and earlier approaches, too)
is the fact that the phonological behaviour of segments is different in
syllable-initial position as opposed to syllable-final position (see also
have more distinct syllable onsets than they do syllable codas; indeed,
some languages permit no post-vocalic tautosyllabic consonants. Related
to this is the observation that segments (or contrasts) are better preserved
achronically in onsets than in codas. Further, it has been suggested that
yllabification in languages follows an 'onset first' principle, i.e. given a
sequence of the sort VCCCV, the syllable boundary will be determined
by first trying to associate as many of the medial C's as possible with the
second syllable in order to form a syllable onset permitted by the language,
and then assigning any C's left over to the preceding syllable (e.g.
Clements & Keyser). Thus extra would have the syllabification [ek.stra],
ot [ek.stra], etc., even though -Vks- and strV- are both permitted
sequences in English. It has long been known that syllabics but not
syllable onsets contribute to the 'weight' of a syllable (which, in turn,
determines stress or accent placement in polysyllabic words). Why should
these asymmetries exist? Are they related?

There is good reason to think that these facts are related and that there
is an identifiable phonetic basis for them. (In the following discussion we
will first use the term 'syllable' as an a priori concept. At the conclusion
and summary, however, we will suggest how the notion of the syllable
might be derivable from certain more basic facts.)

A perceptual phenomenon that is almost certainly related to these
phonological facts is that referred to as the 'P-centre' (or 'perceptual
centre') (Rapp 1971; Allen 1972; Morton et al. 1979; Fowler 1979;
Marcus 1981). When subjects are asked to synchronise clicks with syllables
(where either the timing of the clicks or of the speech is under the control of the subject) it turns out that the clicks are aligned at a point, called the P-CENTRE, which is close to the CV transitions of syllables. Subjects therefore seem to regard the P-centre as the moment at which the given syllable began. Although there is no clear acoustic ‘landmark’ correlated with the P-centre (e.g. in the case of voiceless aspirated stops, the P-centre is located not at stop release and not at voicing onset, but rather roughly in the middle of the aspiration), it seems likely that it correlates with some sort of weighted ‘average’, over time, of all the acoustic modulations (changes in amplitude, spectrum, periodicity, fundamental frequency (Fo)) that occur near the beginning of the syllable (but not completely independent of events at the end of the syllable, as Marcus showed).

On the face of it, this behaviour is somewhat puzzling. The articular gestures needed to produce a syllable must occur well before the P-centre — indeed, given the well-known fact of anticipatory coarticulation, the true ‘beginning’ of one syllable actually occurs in the middle of the preceding syllable. Furthermore, if one were searching for acoustic events which are closer to (if not precisely synchronised with) the articulations which precede syllable onset, better candidates exist than the P-centre (which, as indicated, often does not coincide with any clear acoustic landmark). For example, the onset of frication during a syllable-initial /str/ cluster would seem to be a better marker of syllable onset than where the P-centre actually falls: about two-thirds of the way into the sequence. Nevertheless, plausible speculations can be offered to suggest why, of all the various acoustic landmarks (discontinuities or rapid modulations of acoustic parameters) that occur in a syllable, it should be those that occur near the CV interface that listeners identify as the most useful timing mark for the beginning of a syllable.

In the following discussion we will need to refer to the salience of certain acoustic events in speech. This is, admittedly, not a very well documented concept as far as speech is concerned — although there is some evidence for its applicability to stimuli presented to other sensory modalities and to non-speech acoustic stimuli as processed by non-humans. We will assume that there is an intuitively plausible notion of salience of speech events which is a scalar property that determines a given event’s detectability — i.e. both the probability that it will be detected in the stream of speech (as opposed to being missed) and that its identity (the way in which it differs from other speech events) will also be perceived. Without being able to give a quantitative account of it (although see Kawasaki 1982 and Ohala et al. 1984 for a quantitative approximation to this parameter) we believe (with others) that the salience of an acoustic signal or a portion of it depends on the magnitude, rate, and the number of stimulus parameters varying simultaneously. All of these may operate within limits, e.g. second formant transitions exceeding 30 kHz/sec may not be perceptible (as formant changes). (It is not our purpose here to review the extensive literature relevant to this concept but see, for example, Pollack 1968; Nabelek & Hirsh 1969; Stevens 1971.)

First, although it is not always salient acoustic modulations that are large part this happens because frequency of occurrence in sp.

1960: Kucera & Monrose 1960 dynamic parameters of oral air flow at obstruent offset it takes place, it takes a relative is maximum level behind it obstruent and potentially the but much less time to return it is released — usually less than change in the transglottal pr. the point of constriction at co sudden changes in voice amg. fundamental frequency. It is absent at consonant onset. A salient landmark near obstute of the reason why contrasts codas.

A second reason for the gr junction may have to do with: two types of sequences. It (coarticulation) is predominl (i.e. regressive as opposed C,C,v,C,v, string. then, it fo. proper to v, will be coarticl those for v, etc. But since t during the production of the is greatly attenuated or who constriction effectively exerting shape behind the constriction evident auditorily than the c. Furthermore, the admiss the cues for consonants (w. modulation, bursts, and fric. distortion by the vowel-spe consonant-proper gestures d v is more noticeably colore a C is coloured by a following the degree of change in the a variety of combinations of V: CV. (It should also be mentio changes whereby the V pre function of the C so that the f of nasals — where a distinctiv where some sort of back gli
Prosodic phonology and phonetics

First, although it is not always true, it is generally the case that the most salient acoustic modulations in a syllable occur near the CV interface. In large part this happens because obstruents (which generally have higher frequency of occurrence in speech than non-obstruents: Wang & Crawford 1960; Kučera & Monrose 1968) cause more abrupt changes in the aerodynamic parameters of oral and subglottal air pressure and glottal and oral airflow at obstruent offset than onset. As soon as the obstruent closure takes place, it takes a relatively long time for air pressure to build up to its maximum level behind the constriction - 10 to 15 msec for voiceless obstruents and potentially the entire closure interval for voiced obstruents, but much less time to return to atmospheric pressure once the constriction is released - usually less than 10 msec. As a result there is a more abrupt change in the transglottal pressure drop and in the pressure drop across the point of constriction at consonant release than at onset. This produces sudden changes in voice amplitude and in the case of voiced obstruents, fundamental frequency. It also creates a noise burst which is of course absent at consonant onset. All of this gives rise to an acoustically more salient landmark near obstruent release than onset. This is probably part of the reason why contrasts are better preserved in syllable onset than codas.

A second reason for the greater salience of the CV as opposed to the VC junction may have to do with asymmetrical effects of coarticulation in these two types of sequences. It seems to be the case that assimilation (coarticulation) is predominantly anticipatory rather than perseveratory (i.e. regressive as opposed to progressive) (Javkin 1979). Given a CV, CV, CV, CV string, then, it follows that some of the articulatory gestures proper to V₁ will be coarticulated with those for C₁, gestures for C₂ with those for V₁, etc. But since the vowel-proper gestures will be anticipated during the production of the consonant when the amplitude of the signal is greatly attenuated or where the high impedance of the consonantal constriction effectively extinguishes the resonance effects of the vocal tract shape behind the constriction, these coarticulatory effects will be less evident auditorily than the consonantal gestures made during the vowel. Furthermore, the admixture of vowel-proper gestures does less to distort the cues for consonants (which rely on such robust cues as amplitude modulation, bursts, and friction noise, which are largely impervious to distortion by the vowel-specific coarticulation) than the admixture of consonant-proper gestures does to distort vowels. The result is that the V is more noticeably coloured by - is more similar to - a following C than a C is coloured by a following V. Since auditory salience is correlated with the degree of change in the acoustic parameters, it follows that, for a wide variety of combinations of V and C, a VC sequence will be less salient than CV. (It should also be mentioned that this situation sets the stage for sound changes whereby the V preceding a C may 'take over' the distinctive function of the C so that the C is eliminated. This is especially true in cases of nasals - where a distinctively nasalised vowel is substituted, laterals - where some sort of back glide is substituted, e.g. French autre < Latin
alterum 'other', and other sonorants. It is also not unknown in the case of post-vocalic obstruents, e.g. Lhasa Tibetan [pho] < earlier Bdef 'Tibet': see Michailovsky (1975). This is yet another mechanism by which contrasts in syllable codas get depleted.

Certain asymmetrical tendencies evident in sound changes may be interpreted as providing support for the claim that auditory cues present in CV's are more robust than those in VC's. Original VC; C;V sequences where (presumably) C; was not released often undergo a sound change whereby C; partially or completely assimilates to C; e.g. Latin applicare < ad + plicare. (Certainly the reverse direction of assimilation is also found but is decidedly less common.) Although this has usually been explained as due to 'ease of articulation', i.e. the speaker decides it is 'easier' to make one articulatory gesture than two, we believe that acoustic-auditory factors play at least as important a role in this process. There is experimental evidence that when place of articulation cues are different at VC and CV transitions, listeners tend to follow the CV cues (Wang 1959; Malecot 1962; Repp 1978; Fujimura et al. 1978; Streeper & Nigro 1979; Schouten & Polis 1983). Presumably, something of this sort is what happened in the kind of sound changes under discussion. Furthermore, the perception of consonant place of articulation has been shown under certain conditions to be poorer for unreleased final stops than for syllable-initial stops which are also utterance-initial and therefore lack auditorily clear onsets (Householder 1956). Accordingly, sound changes whereby post-vocalic consonants, especially voiceless stops, are lost or are substituted by [?] are far more common than similar changes involving initial consonants.

Second, the CV junction may provide a more logical timing mark at which to synchronise the prosodic and segmental articulatory streams. Given that some syllables have to be accentuated (or receive distinctive tone in tone languages), and given that one of the phonetic manifestations of these prosodic signals, e.g. some sort of fundamental frequency contour, requires time and sufficient amplitude of voicing to be implemented, it would make sense to start the accent or tone at a well-defined point (a) which was near the beginning of the vowel and (b) where voice amplitude was high. Neglecting (a) would mean that the contour might be initiated at a point where insufficient time for its full realisation remained; neglecting (b) would mean that the contour might not be audible. The P-centre would seem to satisfy these requirements or, at least, is better than any alternative point near the middle or end of the vowel or at the onset of a post-vocalic C. This is true even with sonorants such as nasals, laterals, and glides, since voice amplitude is more attenuated during their production than it is during vowels. There is some evidence that the Fo contours of the Swedish word accents may be timed to begin when the voicing of the vowel begins, e.g. the accent pattern for V1 of a V1C*V2 sequence shows greater delay with respect to that of V1 the longer the duration of the intervening consonants (Eriksson & Alstermark 1972; Eriksson 1973). (However, the data are more complicated when the full range of Swedish dialectal manifestations Bannert & Bredvad-Jense observation that syllable syllable, i.e. since syllable those segments in a syllable manifestations of accent & the rhyme.

Some evidence also exists to create temporally more CV as opposed to VC inte synchronisations between are better correlated for t the VC junctions. Greate The reason for this is that an acoustic change of a of two or more articulators to an acoustic modulation principle may derive from the clear temporal anchors of segmental and suprasegmental reasons given above he kn by making pre-vocalic seg Our speculations may be stream of acoustic events on speech production that auditory modulations that occur at CV boundaries due to the eroding influence of more contrasts in positions or, conversely, v last than the former. Be they constitute an ideal is the segmental stream and the frequency component). Its realisation point may be near the l where voicing has greate duration of the segments constraints on accent place preceding this point. Fina of the events at or near vow of the prosodic and segm precisely. It is this differ events as 'before the voice notion that the stream of s
diallelal manifestations of the word accents are taken into account; Bannert & Bredvad-Jensen 1975, 1977.) If so, this would account for the observation that syllable onsets don't contribute to the 'weight' of a syllable, i.e. since syllable weight corresponds roughly to the duration of those segments in a syllable which can bear accent and (as we suggest) the manifestations of accent start near the CV interface and thus include only the rhyme.

Some evidence also exists which suggests that the speaker actively tries to create temporally more well defined, more precise, articulations near the CV as opposed to VC interface. Tuller et al. (1982) found that the timing (synchronisations) between the various articulations of a CVCVC utterance are better correlated for those articulations associated with the CV than the VC junctions. Greater precision should give rise to greater salience. The reason for this is that if a movement of a single articulator gives rise to an acoustic change of a certain degree, then simultaneous movements of two or more articulators should, under some circumstances, give rise to an acoustic modulation of an even greater degree. The 'onset first principle' may derive from an attempt by the speaker to create maximally clear temporal anchors which will make his synchronisation of the segmental and suprasegmental streams obvious to the listener. For the reasons given above he knows tacitly that there is greater pay-off in salience by making pre-vocalic segments precise than post-vocalic.

Our speculations may be summarised and clarified as follows: given the stream of acoustic events in speech, it is due to the physical constraints on speech production that some of these events will create larger acoustically-auditory modulations than others. These more salient modulations tend to occur at CV boundaries more than at VC boundaries. Over time, then, due to the eroding influence of sound change, languages will tend to preserve more contrasts in these pre-vocalic positions than in post-vocalic positions or, conversely, will show more co-occurrence constraints in the latter than the former. Because these pre-vocalic events are more salient they constitute an ideal timing mark for the synchronisation of the segmental stream and the prosodic (here limited largely to the fundamental frequency component). It is also logically necessary that this synchronisation point be near the beginning of these portions of the speech stream where voicing has greatest amplitude. For this latter reason it is the duration of the segments from vowel onset onward that plays a role in constraints on accent placement in languages, not the duration of segments preceding this point. Finally, speakers may actively enhance the salience of the events at or near vowel onset (in order to make their synchronisation of the prosodic and segmental streams clearer) by articulating them more precisely. It is this differential grouping or organisation of the speech events as 'before the vowel' vs. 'after the vowel' which may underlie our notion that the stream of speech is divided up into syllables.
4 Phonetic processes within the metrical grid

One of the most important aspects of prosodic phonology is the construction for each utterance of a metrical tree consisting of a series of binary branches labeled strong or weak, such that terminal nodes are assigned to each syllable in the utterance (except at certain stages of derivation for so-called 'extrametrical' syllables). The rules which assign degrees of prominence and which determine certain other phonological processes, e.g., the flapping of alveolar stops in English, make reference to these trees. Sample tree structures are given in (1) and (2):

(1) divinity

\[ \begin{array}{c}
\text{w} \\
\text{s} \\
\text{w} \\
\text{s} \\
\end{array} \]

(2) law degree requirement changes

\[ \begin{array}{c}
\text{s} \\
\text{w} \\
\text{w} \\
\text{s} \\
\end{array} \]

Hayes (1981: 16) raises the question of whether these trees have any phonetic reality and quite candidly admits that they have none – that the trees 'depict [only] a mental representation of the relative prominence of syllables and words in an utterance', although phonetic facts allow one to infer their existence indirectly. Hayes may be right, but there is a well-documented phonetic effect that exhibits such an uncanny similarity to one aspect of such trees that it bears further examination in this light. Briefly, the duration of syllables has been shown to vary, specifically, to shorten, in a quite regular way as a function of how many syllables follow (usually those in the same 'phonological phrase'). Lindblom (1975), Lindblom & Rapp (1973), and Lindblom et al. (1976) present original data and review previous work which shows that the duration, \( D \), of a segment is a function of the canonical duration, \( T \), and the number of following syllables, \( n \), according to an equation of the form in (3):

\[ D = T/(n+1)^a \]

\( a \) is a constant that depends on certain phonetic properties inherent to the segment; \( a \) values of around 0.4 are typical, so that with two syllables following, a given syllable or syllable part would have a duration 64 per cent of its canonical duration.

McCarthy suggests that the closer the syllables are joined, the closer they are. But the formalism offers the probability of a stop becoming, which, in turn, gives syllable containing the /t/.

We believe the congruous examination, especially as it which prosodic phonology

5 Sonority hierarchy

A prerequisite for the consistency division of the phonemic's syllabification can be done for segments, for example, as
trical grid

Phonology is the construction of a series of binary branches, and grammar is the construction of a series of binary branching trees, for the purposes of derivation, for so-called grammatical processes, e.g., the flapping of trees. Sample tree:

\[
\begin{array}{c}
\text{repetitive} \\
\sigma \sigma \sigma \\
\end{array}
\]

whether these trees have any phonetic properties inherent to them, or none - that the relative prominence of phonetic facts allow one to be right, but there is such an uncanny similarity in how many syllables follow, specifically, to how many syllables follow the word "happy." Lindblom (1975), (1979) present original data, which are used to determine the number of following segments in (3):

\[
\text{McCarthy suggests that the probability of flapping occurring is a function of how close the syllables are as determined by the height of the node which joins them: the closer they are, the more likely it is that flapping may occur.}
\]

But the formalism offered by Lindblom will also account for this pattern: the probability of a stop becoming a flap is a function of how short the stop becomes, which, in turn, is a function of how many syllables follow the syllable containing the /t/.

We believe the congruence of these two formalisms deserves closer examination, especially as it may provide some of the substantial evidence which prosodic phonology currently lacks.

5 Sonority hierarchies

A prerequisite for the construction of the appropriate metrical trees is the division of the phonemic string into syllables. It is generally believed that syllabification can be done by reference to the intrinsic 'sonority' of segments, for example, as given in (5).
The 'Sonority' Hierarchy (arranged from least to most sonorous)

(5) The 'Sonority' Hierarchy (arranged from least to most sonorous)

stops
fricatives
nasals
liquids
glides
vowels

Syllables are claimed to begin with segments which are ordered so as to have ascending sonority and to terminate with segments which have descending sonority. (Of course, to actually determine syllable boundaries in intervocalic position, this rule will have to be supplemented by language-specific constraints as to permissible initial and final sequences and, possibly, the 'onset first' principle.)

There are a number of problems with this method of determining permissible syllable shapes, however. First, no one has yet come up with any way of measuring 'sonority', claims to the contrary (Hankamer & Aissen 1974) notwithstanding. This is true even though hierarchies of this sort have been around at least 110 years (Whitney 1874) and numerous attempts have been made to find some phonetic correlate of it (for a brief review, see Allen 1973). Lacking such independent definition, the sonority hierarchy remains just a label - a restatement of the originally observed facts of syllable formation ('segment types order themselves in this way when making syllables') - but offering no principled explanation for the observations. The second problem is that there is an important class of constraints on syllable formation which any theory of the syllable should handle but which the sonority hierarchy doesn't. Although these are not absolute constraints, there is a strong tendency among languages to avoid sequences of the sort in (6) (to focus just on syllable onsets; for a fuller treatment including documentation, see Kawasaki 1882):

(6) alveolar stops + [l]
labial(ised) consonants + [w] or high back rounded vowels
apical or palatal(ised) consonants + [j] or high front vowels

Thus although initial sequences of the sort [bl, gl, cw, gw, gu, g"u, g'i, wu, ji] are common enough, [dl, bw, bu, g"u, g'i, wu, ji] are systematically excluded in a great number of languages. Following Steriade (1982: 218ff), one might first think of supplementing the 'increasing sonority' constraint with a constraint that required that there be some minimal difference in sonority between successive segments in onsets, the magnitude of this difference varying from one language to the other. (The notion that what matters is the difference in some parameters between the abutting segments is not unlike the principles mentioned above for the constraints on the sequencing of sandwich spreads.) Thus [ji] and [wu] could be excluded in Ignaciano Moxo (as happens to be the case) since that language would require a sonority difference greater than 1 step in this part of initial segment sequences. Unfortunately, this would also rule out sequences of
which are ordered so as to th segments which have rhyme syllable boundaries to be supplemented by initial and final sequences.

method of determining one has yet come up with e contrary (Hanksmer & though hierarchies of this ney 1974) and numerous correlate of it (for a brief definition, the sonority of the originally observed er themselves in this way applied explanation for the is an important class of ory of the syllable should t. Although these are not among languages to avoid syllable onsets; for a fuller aki 1982):

rounded vowels
high front vowels
[gi, dw, gw, gu, g*a, g*a] [wu, ji] are systematically ing Steriade(1982: 218f), easing sonority' constraint some minimal difference in ts, the magnitude of this. (The notion that what even the abutting segments for the constraints on the [wu] could be excluded in incase that language would tep in this part of initial also rule out sequences of

the sort [ju] and [wi], which are quite acceptable. Similar difficulties arise if the 'minimal difference in sonority' criterion is applied to rule out [dl, g*u], etc. Yet Steriade's notion is intuitively attractive: it must be the fact that [j] and [i] are not different enough (as opposed to [j] and [u], for example) that is responsible for the common exclusion of [ji]. The problem is not with the 'minimal difference' criterion but with the sonority hierarchy: it is one-dimensional where it should have as many dimensions as there are acoustic-auditory parameters that can be used to form lexical contrasts. The essence of any communication channel is the production of modulations, i.e., differences, in some carrier signal. This is true whether semaphore, sign language, or speech is involved. The more difference created in the stimulus parameters in passing from one cipher in the code to the next, the better; the smaller the difference, the more likely it is that the sequence will not be detected (accurately or at all). Thus, the problem with sequences such as [wu] and [ji] is that they create minimal modulations in amplitude, periodicity, and spectrum; with [wi] and [ju], on the other hand, at least the second and third formants show sufficient variation. Likewise, what makes sequences such as [sa], [sla], [mla] acceptable as segmental sequences in many languages is that each creates a sufficiently robust modulation of several acoustic-auditory parameters. But, it may be objected, so do the sequences [zmx], [spx], [pst], [sts], etc., and these are not very commonly found in languages (although they do exist). The answer to this is that languages will first utilise those segment sequences which create maximal modulations, and will only resort to sequences that produce lesser modulations after the better ones have been fully exploited (see Stevens 1980 for a similar view). Thus many languages permit in initial position only the optimal stop+vowel and, at the lower end of this scale, many languages exclude the non-optimal [ji] and [wu].

This account differs from that given in terms of the sonority hierarchy in the following ways. First, it makes reference to empirically measurable properties of the sounds involved, e.g., amplitude, periodicity, spectral structure, fundamental frequency. Thus it is not simply a re-labelling of the original observations. Second, what is valued in sound sequences is any sufficient modulation, i.e., any trajectory through the multi-dimensional space defined by the parameters. In this sense, it is blind as to whether the modulation occurs in syllable-initial or syllable-final position. This is not a defect of the hypothesis; there is good reason to believe (as suggested above) that the division of sound sequences into syllables is done for other purposes, e.g., for the sake of synchronising the segmental and supra-segmental articulations. This model therefore simply creates the acceptable sound sequences which can subsequently be broken up into syllables.

A version of this hypothesis was first explored (to our knowledge) by Saporta (1955), and, using binary distinctive features, by Cutting (1975). Cutting attempted to use the difference between the featural representation of segments to predict the frequency of occurrence of initial cluster types. We attempted a further test of the hypothesis by using (continuous) acoustic measures derived from actual speech to predict favoured and
disfavoured cluster types (these latter tendencies derived from an examination of the sequential constraints of approximately 200 languages; see Kawasaki & Ohala 1981; Kawasaki 1982).

For this preliminary test, only the length of the trajectory of the first three formants was measured (averages obtained in a semi-automatic way from natural utterances spoken by an adult male speaker of American English) in sequences of the sort $C_3(C_2)V$, where $C_1 = \{b, d, g\}$, $C_2 = \{j, w, r, l\}$ and $V = \{i, e, a, u\}$. We predicted that the longer the trajectory of these three formants for the given sequence types through the normalised $F_1-F_2-F_3$ space, the more frequent would those sequences be cross-linguistically. (It might be objected that it is unreasonable to expect to be able to make predictions that have cross-linguistic validity based on the pronunciation of one speaker of one language. But any test of this sort involving measurements of actual acoustic parameters will have to employ some small number of ‘real’ speakers of one or some small number of ‘real’ languages; a speaker of ‘universal phonetics’ doesn’t exist. If warranted, the test could always be repeated with as many other speakers of other languages as was deemed necessary.)

When comparisons were made within any given set of $C_3C_2V$ stimuli where $C_3$ was constant (including null), there was generally a good correlation with the phonological data. For example, the following expected relations held (where ‘$>$’ means ‘more salient than’): [da] $>$ [di], [gwi] $>$ [gwu], [bja] $>$ [bij] (to mention a few). Counter to expectations, [dIV] sequences were not always less salient than [bIV] or [gIV]. This ‘failure’ was largely erased by the results of a test of a second hypothesis: that if two or more segment sequences are auditorily similar, i.e. have trajectories through the multi-dimensional space that are nearly parallel, only one of these will survive (because they will be confused and one will be substituted for the other). This measure showed that [dI] sequences were very close to [gI] sequences (thus accounting for why only one of these is usually found in most languages) and satisfactorily accounted for many other commonly observed mergers as well, e.g. [z$\theta$] and [b$\theta$].

Obviously further work is necessary to test this hypothesis fully, especially to try it with different languages, different segment types, and more acoustic parameters which are properly weighted. Nevertheless, we believe the initial results are promising enough to suggest that the basic idea has merit. It seems capable of accomplishing what was attempted using the ‘sonority hierarchy’, but does it in an empirically verifiable way and it accomplishes what the sonority hierarchy was not able to achieve, a principled account of the low incidence of sound sequences of the sort [ii], [wu], [g$\theta$u], etc.

6 Conclusion

It must be admitted that much of the phonetic data cited above is of a quite preliminary nature in that they apply to a very limited corpus of speech and a restricted sample of languages. Our purpose, however, is to point
Prosodic phonology and phonetics

... possible fruitful connections between the phonological and the phonetic literature. We therefore suggest that phonetic research can (a) shed light on the cause of asymmetries in the behaviour of syllable onsets vs. syllable codas, (b) give a possible physical verification for the existence of metrical trees, and (c) provide a more explanatory account of syllable formation.

NOTES

1. We thank Manjari Ohala and Donca Steriade for helpful comments on an earlier draft of this paper.

2. For example, a useful clue which eventually led to the discovery of the structure of DNA was the substantive fact that adenine was found in it in the same proportion as thymine, and guanine in the same proportion as cytosine.

3. A possible exception to this pattern is that many languages have more contrasts among nasals in syllable codas than onsets; see Ohala (1973).

4. There is no clear evidence for Morton et al.'s and Fowler's contention that the P-centre is the moment when subjects regard the syllable to have occurred. Even linguistically naive listeners are aware that a syllable takes some time to utter and that it cannot be produced instantaneously, i.e. at a single moment.

5. As far as we know there is no explanation for this at present. It will have to be accepted as a given. We believe the ultimate explanation for it will be found in details of how speech is encoded neurologically.

6. The model used in these studies is not capable of predicting the direction of these mergers; however, see Ohala (1983, 1984).

REFERENCES


John J. Ohala and Haruko Kawasaki


Indiana University Linguistics

phonology, Journal of Linguistic

by, In A. Bruck, R. Fox, & M. il Phonology, Chicago Linguistic

a University Linguistics Club, native phonology, New York.

american English. In M. Halle, et al., For Roman Jakobson. The


ange, (Report of the Phonology

phonology, Indiana University

straints on sound sequence, PhD

dversal constraints on phoneme

itive phonology of Russian,

PhD dissertation, MIT.

istic rhythm, LI 8. 333-336.

oken Swedish. In G. Fant &

ception of speech, LI 12.

cities of spoken Swedish, Papers

urational patterns of Swedish

cesses, Stockholm: ms. 7

ative morphology, LI 12.

ification, LI 38. 574-590.

yllables: evidence from cross

re in American English, LI 36.

central (P-center) location.

d changes, Proceedings, Annual

al centers (P-centers). Psycho-

al frequency transitions, JASA

d patterns. In C. A. Ferguson, an a symposium on nasals and


