

## Alternatives to the Sonority Hierarchy for Explaining Segmental Sequential Constraints

John J. Ohala  
*University of California, Berkeley*

It is quite an old observation that certain sequences of speech sounds are favored and others disfavored when languages make up words and syllables. More specifically, that they are typically arranged in a hierarchical fashion such that in word or syllable-initial position, e.g., the sequence stop + liquid + vowel is possible or any subset that maintains the same order of elements. The reverse order is supposed to hold at word or syllable-final position. In most current phonological literature, this ordering or hierarchy of segment types, commonly called the "sonority hierarchy" (SH) or "strength hierarchy" (which is equivalent to the sonority hierarchy but reverses the LESS - MORE labels at the endpoints) is usually traced back to Sievers (1893), Jespersen (1904: 185ff)<sup>1</sup> or Saussure (1916). There was work which predates these efforts, however. (Awedyk 1975 gives a good review of early concepts of the syllable.) Whitney (1874)<sup>2</sup> in a chapter entitled "The relation of vowel and consonant" discusses in detail such a hierarchy based on the degree of openness of the vocal tract. He applied it to the form of syllables. The earliest mention of such a hierarchy that I have encountered is in de Brosches (or 'de Brosses') (1765) who in his Traité de la formation mécanique des langues, et de principes physiques de l'étymologie presents a three-element hierarchy (without any name) consisting roughly of stops + liquids & glides + vowels (p. 130-133), which ordering produced syllables which were 'doux'. (He also notes that 'les langues barbares'—in which he apparently included German and English—often used the reverse order at the end of syllables (p. 132).)

Such hierarchies have at the very least considerable statistical validity, i.e., they embody very common, if not exceptionless, patterns of segment sequencing: that segments are sequenced such that they show a monotonic increase in sonority up to the syllable peak and a monotonic decrease from that point to the end of the syllable. Nevertheless, in this paper I call attention to the following deficiencies of these hierarchies:

- A. As explanations for syllable shapes they are circular.
  - B. They neglect and are incapable of accommodating even in a descriptively adequate way other common phonotactic patterns.
-

C. They are not well integrated with other phonological and phonetic phenomena.

#### CIRCULARITY

The main problem with these hierarchies is that the parameters supposed to define them, sonority, strength, or openness, which the various segment types are said to possess in varying degrees and which thus determines their position in the hierarchy, have never been defined in any way that could be empirically verified, claims to the contrary notwithstanding (e.g., Hankamer & Aissen 1974). Thus terms such as sonority, etc., are just labels for the rank ordering of the segment types; they do not explain it. The situation would be no different if one called it the 'temperature' hierarchy and claimed that voiceless stops at one end are 'cold' and vowels at the other are 'hot'.

Another potential circularity involves the use of the notion of syllable in determining these hierarchies, since the syllable has not been defined empirically either. Consider, for example, someone who cited the word *scoundrel* as evidence reinforcing the sonority hierarchy, pointing out that the medial cluster was divided /n \$ dr/ in accord with the claim that syllable initially consonants were sequenced so as to maintain increasing sonority. If a skeptic challenges this, asking why the cluster isn't analyzed as /\$ndr/, thus violating the SH, the linguist is likely to respond, begging the question, that the /n/ has to be assigned to the preceding, not the following, syllable.

It should be noted, however, that such circularity would be avoided if the evidence for the hierarchy were limited to the sequences of segment types in word margins since words are in most cases easily identified and isolated. In fact many accounts of word-medial syllable boundaries insist that it is based on the kind of sequences that can appear at the margins of words (Awedyk 1975).

#### NEGLECTED PHONOTACTICS

There are certain sequential constraints which, though not as strong as those which motivate the SH, show greater-than-chance incidence in numerous unrelated languages and which therefore should be dealt with systematically. The SH as presently formulated is unable to accommodate them. I review below some of the patterns

---

documented by Kawasaki (1982) who also gives full references to the literature which describes the patterns in specific languages. In all cases, the patterns concern initial clusters.

(1) /w/ is disfavored in C2 position where C1 is a labial; secondary labialization is disfavored on labials.

The above generalization has been stated as two separate conditions but are obviously related and could probably be given a unified phonetic formulation (e.g., offglides with lowered F2 and F3 are disfavored after consonants with lowered F2 and F3). The disjoint formulation, however, is a reflection of the two separate forms of evidence obtainable from the phonological literature. In support of (1) Kawasaki cites the evidence in (2), (3), and (4).

(2) English, Korean, Ronga, Tarascan, Urhobo, Vietnamese, Zulu have some Cw sequences but not if C = [labial]. E.g., English *twin*, *dwarf*, *quick*, *Gwen*, *swill* (exceptions found only among obvious loanwords: *bwana*, *pueblo*, *moi*).

(3) Abkhaz, Gã, Mambila have a labialized series of consonants, but not on labials.

(4) Crothers et al. (1979) found that 40 languages out of 197 surveyed used labialization and of these only three have labialized labials, two have labialized alveolars and palatals, and the rest have labialized velars and uvulars.<sup>3</sup>

A related constraint is that in (5).

(5) /j/ is disfavored in C2 position where C1 is a dental, alveolar, or palatal, i.e., [acute] consonants; secondary palatalization is disfavored on these same [acute] consonants.

In support of (5) Kawasaki cites the evidence in (6) and (7).

(6) There is some disfavoring of /j/ following [acute] C, including dialectal dropping of /j/ in this environment, found in: English, Burmese, Korean, Lisu, Yay, Wukari Jukun, Fox, Paez, Sre, Armenian, Dagbani.

(7) Palatalization is disfavored on [acute] in Akha, Dagbani, Even, Gilyak, Wapishana.

She also notes as seeming counterexamples the frequent non-distinctive palatal offglides from the same class of consonants. However, the fact that such glides are phonetically predictable near [acute] consonants is not inconsistent with their systematic absence as distinctive elements in the same environment (see note 3).

In discussions of phonotactics it is often tacitly assumed that any CV combination is possible. In fact, many languages disfavor specific CV sequences.

(8) /w/ and labialized consonants are disfavored before back, rounded vowels.<sup>4</sup>

Kawasaki found some manifestation of (8) in many unrelated languages, as given in (9), (10), and (11).

(9) Japanese, Mazateco, Vietnamese, Guahibo, Loma, Mixtec, Tenango Otomi, Ainu, Huichol, Ignaciano Moxo, Kalinga, Totonaco, Yaqui, Yucuna, Capanahua, Chacobo, Orizaba Nahuatl, Cavineña, Tarascan, Belangao, Trique, Luo, Korean, Sinhalese, Suto-Chuana, Akɔ̃sɔ̃, Yao, Zulu, Wukari Jukun, Mambila, Parintintin, Yuma, Toura, Amharic, Bella Coola and others.

(10) Labialized and plain velars are neutralized before /o u/ in Chehalis.

(11) Some /w/ + /u o/ clusters in English have disappeared, e.g., "swoon" = dial. [su:n]; "ooze" < OE /wo:s/; "sword" = [sɔ:d].

Corresponding to (8) which involves sequences of [grave] consonant and vowel, there is (12) which involves [acute] segments.

(12) /j/ and palatalized consonants are disfavored before front vowels.

Kawasaki found evidence of a /j/ + front vowel gap in the languages listed in (13), among others.

(13) Tarascan, Guajajara, Nahuatl, Goajiro, Acoma, Mazateco, Digueño, Western Popoloca, Yucuna, Ainu, Capanahua, Guahibo, Cavineña, Huichol, Korean, Ignaciano Moxo, Totonaco, Tenango Otomi, Sinhalese, Japanese, Trique, Bulgarian, Mixtec, Dan, Yao.

---

Neutralization of palatalized vs. plain consonants before front vowels were found in the languages in (14).

(14) Northern Estonian, Karakatsan.

An example of (13) that is close to home is the fact that Mid-Atlantic & Southern English renders *yeast* and *east* as homophones, thus prompting the old riddle-joke in (15).

(15) Q: Why does the sun rise in the east?  
A: Yeast (= [ist]) makes anything rise.

Other examples could be given of sequential constraints that cannot be easily handled by the SH, e.g., the familiar constraint against initial sequences of apical + /l/.

#### FURTHER DIFFICULTY WITH THE SONORITY HIERARCHY

Commonly encountered clusters such as /st-, xt-, vd-/ violate the principle that syllable onsets consist of elements with increasing sonority. Clusters such as /-ts, -k/, -dz/ violate the principle that sonority decreases in syllable codas.

Another challenge to the SH is the often demonstrated fact that words such as "blow" [blo] can be transformed into a convincing version of "below" [bɛlo] by simply lengthening the duration of the /l/. (This was demonstrated during the oral presentation of this paper; see Price 1980 and references cited there.) The SH assumes that syllables are local maxima in the sonority of the segments which are concatenated in speech, each segment or segment type having its own inherent sonority. But simply lengthening the /l/ does not change any inherent quality. So where does the extra syllable and the perceived [ə] come from?

#### NON-SOLUTIONS FOR SEQUENTIAL CONSTRAINTS

Before presenting proposed alternative explanations for sequential constraints, it may be useful to emphasize how they cannot and should not be handled.

The SH as traditionally conceived cannot accommodate the above constraints. The SH recognizes that in initial clusters glides can follow those consonant types with lesser sonority and, indeed,

such sequences are abundant, e.g., English *cute* [kjut], *sweet*, *quick*, *puce* [pjus], *beaut* [bjut], etc. It is just particular combinations of consonant + glide that are disfavored. As for CV sequences, most applications of the SH do not recognize any constraints. But even if one wanted to proclaim that sequences of glide + high vowels were "marked", given the disfavoring of sequences like /ji/ and /wu/, it would not work because other glide + high vowel sequences are not disfavored, e.g., /ju/ and /wi/.

The Obligatory Contour Principle (OCP) has been invoked to account for various co-occurrence restrictions, including some similar to those discussed above (McCarthy 1989, Yip 1989). But this is equivalent to using the OCP as a diacritic: invoke it when co-occurrence restrictions exist and forget it when they don't. The OCP provides no way in principle of predicting (and thus explaining) those sequences commonly subject to co-occurrence restrictions and those which are not (e.g., sequences of [αvoice][αvoice] such as /ml- st-/).

Regarding common exceptions to the prediction of the SH, e.g., the abundance of /st- ts/ clusters, one way is to make special stipulations for them, i.e., to argue that somehow they are different and should not be covered by exactly the same generalizations as other sequences. This is quite valid if such a position can be defended on empirical grounds. For example, it was quite appropriate and revolutionary for its time to explain that the disease pellagra, despite breaking out as apparent epidemics, was different from microbially transmitted infectious diseases like measles because it was actually due to a vitamin deficiency (niacin). On the other hand, the Ptolemaic astronomical system handled exceptions by freely adding any number of extra epicycles to planetary orbits. English clusters such as /st-/ have thus been accounted for by positing that the /s/ is not really part of the same syllable that the /t/ belongs to. Similarly non-occurring but SH-mandated clusters /pw-, mw-/ are eliminated by "filters" (Borowsky 1986, p. 174). But these are purely ad hoc strategies designed like the Ptolemaic epicycles "to save the appearances" of the data.

Diver (1975) presented a novel articulatory account for phonotactic constraints which involved the classification of consonants into the new categories of [mobile] (all stops and /r/) and [stable] (all fricatives and /l/). He claimed that dissimilar combinations, i.e., [stable] + [mobile] and vice-versa, were disfavored, if not absolutely, then statistically. E.g., clusters such as /pl-, tl-, kl-, sr-, fr-/ were disfavored, /pr-, tr-, kr-, sl-/ favored. In addition, clusters made at the same place of articulation such as /pw-/ are disfavored due to speakers avoiding the repetition of the same

gesture. I am unaware of any empirical support for the features [mobile] and [stable] and common clusters like /st, sl/ as well as affricates /ts, kx/ would seem to indicate that speakers do not avoid repetition of similar gestures. And unless elaborated in some way, this system would seem incapable of accounting for the disfavoring of CV sequences like /wu/ and /ji/.

I intend to present below an account of sequential constraints that encompasses all the usual patterns handled by the SH as well as those discussed above which are not. Moreover the account has the potential of being empirically evaluated and in one domain has been so tested. It differs from most other accounts by being situated primarily in the acoustic-auditory domain.

#### THE PROPOSAL

There are three essential elements to the alternative I offer to account for common cross-language sequential constraints:<sup>5</sup>

(1) Rather than posit a single parameter, sonority, which has never been identified empirically, we should focus our attention on several acoustic parameters which are well known and readily measured in the speech signal—at least these four—: amplitude, periodicity, spectral shape, and fundamental frequency (F0) (Dudley 1940). Spectral shape itself might best be analyzed into a number of different parameters.<sup>6</sup> This step allows us to integrate the study of universals of sequential constraints with other phonetic and phonological phenomena.

(2) Rather than focus on some alleged intrinsic value that individual speech sounds or sound types are supposed to have we should concentrate on the modulations in the relevant parameters created by concatenating one speech sound with another. Saporta (1955), Cutting (1975), and Steriade (1982) made similar proposals by emphasizing the importance of the difference(s) in the intrinsic parameters of successive segments. However, as a refinement to this work Kawasaki has argued that it is not practical to just take the differences between invariant feature values of adjacent segments since the degree of modulation is highly context dependent. For example, both the sequences /gi/ and /gu/ show much less modulation of formants than does /ga/. In part this is due to the fact that /g/ (and most other segments) exhibits much coarticulation with adjacent vowels. In addition, the features or parameters cannot simply be binary but have to be continua.

(3) Define the degree of "goodness" of one of these acoustic modulations as proportional to the length of the trajectory it makes through the acoustic "space" whose dimensions are the acoustic parameters listed above. A large trajectory involving many different parameters should be more easily detected than a small one involving only a few parameters.

All communication systems, whether writing, Morse code, semaphore, sign language, AM, FM, and PCM encodings of audio signals in radio and telephone involve modulations of some carrier signal and possess ciphers that are better detected the greater the magnitude of their modulations. There is no reason to expect speech to differ in this regard. Of course, as Saporta pointed out, the modulations may not always be as large as it is physically possible to make them because speakers may attempt to limit the effort they put into speech production. Lindblom (1983, 1989) suggests that speakers only expend as much energy on producing the signal as they estimate is necessary for listeners to understand them. Nevertheless, even if speakers attenuate the modulations somewhat it should still be the case that larger modulations have more survival value than lesser ones and so will, in the long run, persist in languages. The forces which tend to eliminate non-optimal sequences (listener misperceptions, presumably) should not be expected to act in an all-or-nothing fashion. Rather, over time, more of the less salient modulations would be expected to change or be lost than the better "fit" ones. Some non-optimal sound sequences may still remain. The acoustic modulations in the English words "woo" and "wool" would certainly be low on this "goodness" scale at least in comparison to those in words such as "bash" or "strips". The vast majority of words in languages and thus the phonotactic patterns that can be abstracted from them do seem to comply with the constraint of being sufficiently detectable by virtue of traversing a relatively large trajectory through the multi-dimensional acoustic space.

#### A PHONETIC SIMULATION

In her dissertation Kawasaki explored these ideas with a quantitative phonetic implementation which measured the magnitude of acoustic modulations in various sound sequences, including C1 (+ C2) + V, which I will focus on here. The C1 was any of the voiced stops /b d g/, the optional C2 /l r w j/, and the V /i e a u/. One speaker of American English produced all these sequences and

---



the trajectories of the first three formants (F1, F2, F3) were tracked and measured. (Since the issue addressed is universal phonotactic constraints, it would have been ideal to use a speaker of universal phonetics but such are hard to find. Any study of this sort must, for starters, use speakers of particular languages. Needless to say, if the phonetic manifestation of any of the segment types differ in important ways for a given language, the results obtained could differ.) The hypothesis tested was that the magnitude of the trajectory in the F1 X F2 X F3 space would correlate with the degree to which the sequence was universally preferred in languages.

In general, her results supported the hypothesis, i.e., sequences like /dw/ were "better" than /bw/ in that they made longer trajectories in this acoustic space and thus made more salient modulations. Likewise, /bj/ and /gj/ were better than /dj/ and /ja/ and /ju/ better than /ji/, as predicted. Sequences such as /wi/ and /wa/ were better than /wu/. /bu/ /di/ /gi/ /gu/ were among the least salient CV combinations. The most important mismatch between the results and universal preferences was that /dl/ was in all cases better than /bl/ and in most cases better than /gl/.

One would expect the results to be even better if amplitude and the acoustic properties of stop bursts were taken into account.

#### QUALIFICATIONS AND HEDGES

There are many uncertainties about the proper way to measure acoustic salience, i.e., in a way that corresponds to auditory detectability. The various parameters mentioned need to be weighted properly. No doubt rate of modulation matters: rapid modulations should be better than slow (within limits).

In addition to the salience created by sound sequences, their degree of "predictability" should also be factored in. In some cases, this may be orthogonal to acoustic salience. For example, nasalization on vowels does not greatly change the salience of their transitions into or out of nasal consonants and yet such combinations are rarely distinctive (Kawasaki 1986). The most important cause of the pressure against such sequences may be that nasalization is predictable on vowels next to nasals. Listeners tend to ignore or factor out features of speech that are predictable; this is also the basis for my theory of dissimilation (Ohala 1981, 1986).

Kawasaki recognized that if two or more sound sequences created substantially similar trajectories, i.e., were acoustically and auditorily similar, they would be subject to confusion and thus

---

merger. Combine this with asymmetries in the direction of confusion (Ohala, 1983a, in press a) and one could find situations where a given segment sequence is salient enough but is disfavored due to its merger with another equally salient but acoustically similar sequence.

Although I have emphasized the acoustic-auditory factors which lead to favoring or disfavoring of sound sequences, it should be acknowledged that some sequences may be disfavored due to their being difficult to articulate. For example, in obstruent clusters, voicing may be difficult to maintain because of the buildup of air pressure behind the oral constriction and above the glottis, thus lowering the pressure drop across the glottis required for voicing (Catford 1977, p. 26ff; Ohala 1983b). One can also imagine that palatal and retroflex consonants might be incompatible with adjacent apical trills because the tongue configurations for the sequences are so different. Articulatory factors can also lead to the introduction of "extra" segments (if listeners misinterpret the signal), e.g., the epenthetic stops in nasal + obstruent or lateral + /s/ sequences: *warm[p]th, pul[t]se*.

Finally, in spite of having invoked the notion of "space", I wish to officially "distance" myself from this concept. I remain suspicious of the usefulness of the distance and spatial metaphors for the domain which speech sounds occupy due to the inescapable fact that auditory confusions between speech sounds are very often asymmetrical, i.e., X confused with Y more often than Y with X. A spatial concept is inherently unable to deal with such facts since if X is confused with Y by virtue of being "close" to it, it should be the case that X is as close to Y and thus confusable with it as Y is close to X, i.e., the confusion should be mutual and symmetrical. However, I acknowledge that for preliminary computations, spatial metaphors and the accompanying mathematics (e.g., for computing distance in a multi-dimensional space via the Pythagorean theorem) are very useful. I think a "cost" measure would be more appropriate to characterize the transformation of one sound into another (by the listener) (Ohala, 1985, in press a).

#### A GEDANKEN SIMULATION

As a "thought" simulation one can imagine generating sound sequences by stringing together various sounds randomly as done in (16a-d).

---

- (16) (a) sfs ptp iji uwu yyy fəf  
 (b) pwe dje dle  
 (c) ske ble gre sts  
 (d) ba sa tja

Given the discussion above, I would expect that although all of these sequences are possible (and probably attested in some language somewhere), they could be rank ordered according to the degree to which they created sufficient acoustic modulations. (16a) through (16d) is a possible ordering, where (a) represents the weakest modulations and (d) the strongest (without attempting to make fine distinctions in strength of modulations within each ranking). The predicted survival of a given sequence and thus its frequency of occurrence in languages of the world would be predicted to be proportional to the magnitude of the modulations it created.

At this point no mention has been made of syllables. I believe this correct. From the point of view of the requirements of the vocal-auditory communication system syllables do not seem to be the first priority; modulations of the carrier signal are. The segmental stream could be transmitted without being chunked syllabically. Similarly, typing simply requires depressing one typewriter key after another; grouping a certain number of keystrokes together is not required. As Mandelbrot (1954) has argued, some sort of discretization of the signal (or imposition of "breakpoints") is important for the sake of efficient communication but this requirement is met by the segments themselves (see Ohala, in press b). A larger unit, the syllable, which groups some segments together is not logically necessary. Now, if it happens that in both speaking and typing some sort of chunking of the "segments" occurs, and this apparently is the case, this is interesting and we should seek the purpose of it. But I think it is important to realize that syllables are logically subsequent, not antecedent, to constructing the optimal segment stream itself. In this way we avoid the circularity of taking syllables as given, finding the favored segment sequences in them and then restricting "legitimate" syllables only to such segment strings.

Stetson (1928) represents one of the few expressions of the view 'syllable first, then segments'. The reverse attitude which I have argued for above is for the most part the established view in

---

phonology today. Nevertheless, I believe it is important to try to reinforce its logical basis and not simply accept it uncritically.

#### WHAT DETERMINES SYLLABIC CHUNKING OF THE SEGMENTAL STREAM?

We are still left with the question of how syllables are imposed on or carved out of the segmental stream. I don't claim to know how this is done or indeed if it is always done on all stretches of speech or if it is done consistently by all speakers. I am also skeptical of many of the unqualified claims in the literature regarding how syllabifications are made since these seem to be based primarily on introspections of silently articulated or artificially slowed-down speech. In addition, such introspections are often done in the presence of an already formulated hypothesis about proper syllabification, thus raising the serious possibility of bias in the observations.

I offer below some speculations about the origin and purpose of syllables. I will also try to link some of the common conceptions regarding syllable structure to elements in the phonetic domain.

Ohala and Kawasaki (1984) speculated that syllables (or as they phrased it, certain "landmarks" within the syllable) may be necessary for the synchronization of the prosodic and segmental streams. If F<sub>0</sub> and other prosodic parameters are to be used to differentiate or demarcate words it might be beneficial that these prosodic modulations always occur synchronized with a fixed part of a word. Syllables would be the chunks on which these modulations are superimposed.

It is also possible that neuromotor constraints require the syllable for the sake of efficient speech production, i.e., that it is a grouping of segments or gestures that are more efficiently or more skillfully produced if linked together. We know next to nothing about the neuromotor level of speech production so it is scarcely possible to provide any details on this. However, to make the idea plausible, consider an analogy: The ideal theoretical requirements for propelling a bicycle might specify only that a constant torque be applied to the rear wheel. But due to practical constraints in implementation, torque is applied in "bursts" when either pedal is parallel to the ground. The neuromotor system that produces speech may have its own constraints, as does the bicycle and its rider, which motivates chunking of what might theoretically have been a continuous stream. That skill somehow plays a role in syllable production is suggested by the fact that speakers of English can

produce the velar nasal [ŋ] in non-initial position but have trouble with it in initial position (and many similar cases with speakers of other languages).

#### IN SOME CASES THE SYLLABLE IS A PERCEPTUAL CONSTRUCT

In the case cited above where the artificially lengthened /l/ in the word "blow" induced listeners to hear "below", we must conclude that syllabicity is a perceptual construct, i.e., created in the mind of the listener. The basis for this, I think, is the fact that when listeners hear the sonorant after the /b/ which is longer than expected for a normal /l/ they are induced to construct another percept which is more in accord with the events in the acoustic signal. The transition at the release of the /b/ is compatible with a vocalic release /ba/ and of course the /lo/ transition is unchanged. Listeners don't require an actual /e/ transition in order to conclude that there is a /e/ in between the /b/ and /l/ (perhaps because VC transitions are less important than CV transitions; see below). This is not just a perceptual curiosity created by artificial manipulations of the speech signal. Menendez-Pidal (1926, p. 217-218) notes the existence in Spanish, sometimes sporadically, of a 'vocal relajada' breaking up clusters of the sort Cl and Cr, e.g., *Inglaterra, coronica, egelesia, peredicto*. Similar examples can be found in many other languages. Something of the same sort probably underlies the variable phonetic and phonemic analysis of syllabic /l r n m ŋ/ in English as /l/ or /e/ l/, etc. Instrumental examination of words with these syllabic consonants as pronounced by most American English speakers usually gives little evidence of a separate vowel.

#### WHY IS THE CV STRUCTURE SO COMMON?

There are several indications that CV structures have a special status vis-a-vis VC. CV is much more common in languages of the world than VC; virtually all languages have CV but many do not have VC. The "onset first" principle dictates that intervocalic consonants are preferentially associated with the following rather than the preceding vowel. Even if one would be tempted to dismiss this as introspection subject to bias, the fact that so many different writers on syllabicity have reached the same conclusion independently makes it difficult to dismiss (Pulgram 1970, Awedyk 1975). There are, perhaps, some common sense reasons why the

---

modulations in CV and VC would not be symmetrical and these factors may be related to the special phonological behavior of CV.

First, in the case of obstruents there will be a higher pressure built up behind the constriction at their offset than at their onset (because it takes time for the air to accumulate behind the constriction and thus raise oral pressure). Thus stops have bursts offset but obviously not at onset. In general, stop bursts are highly salient and possess important cues for place and manner of articulation. In the case of fricatives, since it takes more time to build up pressure than to release it, there may be a slower rise-time of the fricative noise at onset but a more rapid fall-time at offset (as well as more rapid rise-time in the amplitude of the following vowel). Assuming that a more rapid modulation of the periodicity parameter is more salient than a slow one, fricative offsets may have stronger modulations than onsets.

Second, there are asymmetries in the direction of coarticulation or assimilation which may serve to make CV modulations stronger than those at VC. The reasoning behind this claim is a bit complex. First, it is taken as a given that anticipatory (regressive) assimilation is stronger (i.e., longer in time) than perseveratory (progressive) assimilation (Javkin 1979). The reasons for this are unknown but presumably have something to do with how the speech "motor program" is put together in the brain. Thus this means that coarticulation of a segment sequence XY will show a longer and more gradual interval of admixture of (some of) Y's features during production of X than there would be of X's features during Y. A longer and slower approach to Y during X would mean for a less strong modulation. Translating this to a VCV sequence would get the situation depicted schematically in Figure 1, where at the top of the figure the time course of the sequence runs from left to right and the "target" configuration for a consonant or a vowel is represented at the top of the vertical axis and non-target configuration at the bottom. The solid line represents the configuration of the consonant and the dashed line that of the vowel. The greater anticipatory assimilation is reflected by the less steep slope of the parameter when approaching the target; the lesser perseveratory assimilation, by the steeper slope at offset from the target. The latter corresponds to a stronger modulation. Now as represented at the top of the figure, the VC and CV transitions would be equal in that both would be the locus of slow anticipatory assimilation and rapid perseveratory assimilation and so there is asymmetry in them. However, the assimilations during the consonantal interval occur typically while substantial portions of t

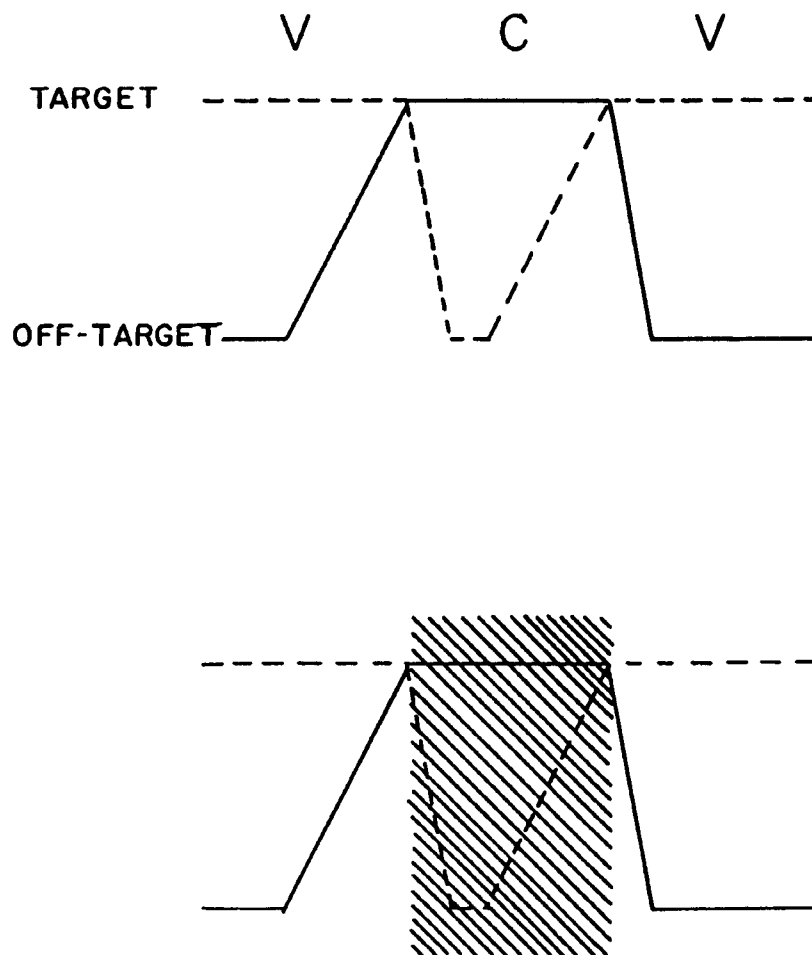


Figure 1. Schematic representation of consequences of asymmetry in anticipatory and perseveratory assimilation during a VCV utterance. TOP: dashed line represents vowel gestures, solid line, consonant gestures; horizontal axis is time, vertical axis represents target configuration at top and off-target at bottom. Longer anticipatory assimilation symbolized as less steep slope of lines leading up a target configuration; shorter perseveratory assimilation, as the steeper slope of lines leading away from target. BOTTOM: same as upper figure but with cross-hatching symbolizing the acoustic attenuation of the effects of vowel assimilation during the consonantal constriction. The result is that the steeper, more salient acoustic modulations will be heard at the CV not the VC juncture.

vocal tract are occluded, i.e., where the acoustic evidence for the assimilatory influence of the vowel is severely attenuated. This is represented symbolically in the lower part of the figure where the consonantal interval is obscured with cross-hatching. In other words, the articulatory assimilations which blunt or sharpen the transitions between vowel and consonant may be substantially equal but listeners can't hear that which occurs during the consonant. What they can hear is the slower transitions at VC and the more rapid and more salient transitions at CV.

Unfortunately, perceptual evidence for the greater salience of CV than VC transitions is mixed. Sharf and Beiter (1974) found VC transitions gave rise to fewer errors in consonant identification than CV transitions, thus indicating that VC transitions are more informative. Öhman (1966) found no difference in the two. However, in Ohala (in press c) I presented original data and reviewed previous evidence that in intervocalic heterogeneous clusters of the sort /abg/ (created by cross-splicing the first and second halves respectively of the utterances /aba/ and /aga/) listeners attend selectively to cues at the CV transition in order to determine the place of articulation of the single stop perceived. Tuller, Kelso, and Harris (1982) found interarticulator synchronizations tighter at CV than VC junctions. This latter evidence is not from the perceptual domain but plausibly reflects special attention to articulation for the sake of maintaining the acoustic salience of the CV transition.

None of the above by itself suggests how syllables are carved out of the segmental stream but it does indicate that asymmetries exist which might form the basis or "bias" toward cutting up the stream in one way as opposed to another.

#### CONCLUSIONS

"Sonority" and its cousin "strength" do not exist and should be abandoned for the sake of explaining universal sequential constraints. (Of course they do have value for the history of phonology as does the concept of "phlogiston" in the history of chemistry.) It should be replaced by a measure of the degree of modulation in several acoustic parameters (amplitude, periodicity, spectral shape, F0) and the notion that survivability of a given segmental sequence is proportional to the strength of this modulation. This by itself would only predict which sequences should be found in languages. This still leaves unanswered how and why the segmental stream is chunked into syllables. I speculate that

---



syllable chunking may be done for the sake of synchronizing suprasegmental and segmental events or to accommodate neuromotor constraints. The principles of this chunking, however, may in part depend on the degree of salience of modulations created by segmental transitions.

#### NOTES

<sup>1</sup> Jespersen 1904 is a German translation of part of his earlier Danish work published in three parts in 1897-1899. His work on the syllable appeared in the third part and so the reference to Jespersen's thoughts on the syllable should be to 1899 (p. 521ff).

<sup>2</sup> Actually a few years earlier.

<sup>3</sup> Ohala and Lorentz (1977) claimed that labialization is especially associated with back velars (and uvulars) and labials. In fact the incidence of labialization on consonants they cite, based on the 706 languages surveyed in Ruhlen (1976) is very similar to that of Crothers et al. (whose languages are for the most part a subset of Ruhlen's). Phonetically it is quite true that labials and back velars produce offglides that resemble secondary labialization and these offglides can be rephonologized as distinctive labialization. Nevertheless, Kawasaki's observation is correct that labialized labials are proportionately underrepresented in the segment inventories of languages utilizing distinctive labialization.

<sup>4</sup> Kawasaki's original claim was that labial consonants like /p, b, m/ etc. were also disfavored before vowels /u, o/ etc. However, Janson's (1986) statistical study of CV sequences in five languages suggests that this is not the case. The most conservative claim, then, is that there is a disfavoring of the glides /w/ and /j/ and labialized and palatalized consonants before back rounded vowels and front unrounded vowels, respectively.

<sup>5</sup> See previous discussion of this in Ohala 1980, Kawasaki 1982, and Ohala and Kawasaki 1984.

<sup>6</sup> Bell and Hooper (1978, p. 12) raised the question of whether sonority might be more than one parameter (see also Price 1980). Saporta (1955) and Cutting (1975) essentially replaced sonority by several traditional phonological features in their treatment of universals of sequential constraints.

## REFERENCES

- Awedyk, W. 1975. The syllable theory and Old English phonology. [Polska Akademia Nauk. Comitet Neofilologiczny] Wrocław: Polskiej Akademii Nauk.
- Bell, A. & Hooper, J. B. (eds.) 1978. Syllables and segments. Amsterdam: North Holland.
- Borowsky, T. J. 1986. Topics in the lexical phonology of English. University of Massachusetts (Amherst) doctoral dissertation.
- de Brosches, Ch. 1765. *Traité de la formation mécanique des langues, et de principes physiques de l'étymologie*. 2 volumes Paris: Chez Saillant, Vincent, Desaint.
- Catford, J. C. 1977. *Fundamental problems in phonetics*. Bloomington: Indiana University Press.
- Crothers, J., Lorentz, J., Sherman, D., & Vihman, M. 1979. *Handbook of phonological data from a sample of the world's languages: a report of the Stanford Phonology Archive*. Stanford.
- Cutting, J. 1975. Predicting initial cluster frequencies by phonemic difference. Haskins Laboratories, Status Report on Speech Research. SR-42/43.233-239.
- Diver, W. 1975. Phonology as human behavior. *Columbia University Working Papers in Linguistics* 2.27-57.
- Dudley, H. 1940. The carrier nature of speech. *The Bell System Technical Journal* 19.495-515.
- Hankamer, J. & Aissen, J. 1974. The sonority hierarchy. *Papers from the Parasession on Natural Phonology*. Chicago: Chicago Linguistic Society. 131-145.
- Janson, T. 1986. Cross-linguistic trends in the frequency of CV sequences. *Phonology Yearbook* 3. 179-195.
- Javkin, H. R. 1979. *Phonetic universals and phonological change*. Report of the Phonology Laboratory, Berkeley, No. 4.
- Jespersen, O. 1897-1899. *Fonetik en systematisk fremstilling af laeren om sproglyd*. Copenhagen: Det Schubotheske Forlag.
- Jespersen, Otto. 1904. *Phonetische Grundfragen*. Leipzig und Berlin. Teubner.
- Kawasaki, Haruko. 1982. *An acoustical basis for universal constraints on sound sequences*. University of California, Berkeley, doctoral dissertation.
- Kawasaki, H. 1986. *Phonetic explanation for phonological universals: The case of distinctive vowel nasalization*. In J. Ohala & J. J. Jaeger (eds.), *Experimental Phonology*. Orlando, FL: Academic Press. 81-103.

- Lindblom, B. 1983. Economy of speech gestures. In P. F. MacNeilage (ed.), *The production of speech*. New York: Springer-Verlag. 217-245.
- Lindblom, B. 1989. Phonetic invariance and the adaptive nature of speech. In B. A. G. Elsendoorn & H. Bouma (eds.), *Working models of human perception*. London: Academic Press. 139-173.
- Mandelbrot, B. 1954. Structure formelle des textes et communication. *Word* 10.1-27.
- McCarthy, J. J. 1989. Feature geometry and dependency: A review. *Phonetica* 45.84-108.
- Menendez-Pidal, R. 1926. *Orígenes del español*. Madrid: Hernando.
- Ohala, J. J. 1980. The application of phonological universals in speech pathology. In: N. J. Lass (ed.), *Speech and language. Advances in basic research and practice*. Vol. 3. New York: Academic Press. 75-97.
- Ohala, J. J. 1981. The listener as a source of sound change. In: C. S. Masek, R. A. Hendrick, & M. F. Miller (eds.), *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Linguistic Society. 178-203.
- Ohala, J. J. 1983a. The phonological end justifies any means. In: S. Hattori & K. Inoue (eds.), *Proceedings of the XIIIth International Congress of Linguists, Tokyo, 29 Aug.-4 Sept. 1982*. Tokyo. [Distributed by Sanseido Shoten.] 232-243.
- Ohala, J. J. 1983b. The origin of sound patterns in vocal tract constraints. In: P. F. MacNeilage (ed.), *The production of speech*. New York: Springer-Verlag. 189-216.
- Ohala, J. J. 1985. Linguistics and automatic speech processing. In: R. De Mori & C.-Y. Suen (eds.), *New systems and architectures for automatic speech recognition and synthesis*. [NATO ASI Series, Series F: Computer and System Sciences, Vol. 16] Berlin: Springer-Verlag. 447-475.
- Ohala, J. J. 1986. Phonological evidence for top-down processing in speech perception. In: J. S. Perkell & D. H. Klatt (eds.), *Invariance and Variability in Speech Processes*. Hillsdale, NJ: Lawrence Erlbaum. 386-397.
- Ohala, J. J. In press a. Comparison of speech sounds: Distance vs. cost metrics. *Festschrift for Osamu Fujimura*. Tokyo: Ohmsha.
- Ohala, J. J. In press b. The segment: primitive or derived? To appear in R. Ladd & G. Docherty (eds.), *Laboratory Phonology II*. Cambridge University Press.

- Ohala, J. J. In press c. The phonetics and phonology of aspects of assimilation. To appear in M. Beckman & J. Kingston (eds.), *Papers in Laboratory Phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press.
- Ohala, J. J. & Kawasaki, H. 1984. Phonetics and prosodic phonology. *Phonology Yearbook* 1.113-127.
- Ohala, J. J. & Lorentz, J. 1977. The story of [w]: an exercise in the phonetic explanation for sound patterns. *Berkeley Linguistic Society, Proceedings, Annual Meeting* 3.577-599.
- Öhman, S. E. G. 1966. Perception of segments of VCCV utterances. *Journal of the Acoustical Society of America* 40.979-988.
- Pulgram, E. 1970. *Syllable, word, nexus, cursus*. The Hague: Mouton.
- Price, P. J. 1980. Sonority and syllabicity: Acoustic correlates of perception. *Phonetica* 37.327-343.
- Ruhlen, M. 1976. *A guide to the languages of the world*. Stanford.
- Saporta, S. 1955. Frequency of consonant clusters. *Language* 31.25-30.
- de Saussure, F. 1916. *Cours de linguistique générale*. Paris: Payot.
- Sharf, D. J. & Beiter, R. C. 1974. Identification of consonants from formant transitions presented forward and backward. *Language & Speech* 17.110-118.
- Sievers, E. 1983. *Grundzüge der Phonetik*. [3rd ed.] Leipzig. Breitkopf & Hartel.
- Steriade, D. 1982. Greek prosodies and the nature of syllabification. M. I. T. doctoral dissertation.
- Stetson, R. H. 1928. *Motor phonetics*. La Haye: Martinus Nijhoff. [2nd ed. 1951. Amsterdam: North Holland Publishing Company; retrospective edition 1988, ed. by J. A. S. Kelso & K. G. Munhall. Boston: College-Hill.]
- Tuller, B., Kelso, J. A. S., & Harris, K. S. 1982. Interarticular phasing as an index of temporal regularity in speech. *Journal of Experimental Psychology. Human Perception & Performance*. 8.460-472.
- Whitney, W. D. 1874. *Oriental and linguistic studies*. Second Serie New York: Scribner, Armstrong, & Company.
- Yip, M. 1989. Feature geometry and cooccurrence restrictions. *Phonology* 6.349-374.
-