Alternatives to the sonority hierarchy for explaining segmental sequential constraints

John J. Ohala – Haruko Kawasaki-Fukumori

1. Introduction[1]

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* + *fricative* + *nasal* + *liquid* + *glide* + *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: 185 ff)[2] or Saussure (1916). There was work which predates these efforts, however. (Awedyk 1975 gives a good review of early concepts on the syllable.) Whitney (1874)[3] 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 we have encountered is in de Brosches (or de Brosses) (1765) who in his *Traité de la formation méchanique des langues, et de principes physiques de l'étymologie* presents a three-element hierarchy (without any name) consisting roughly of stops + liquids & glides + vowels (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 (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 we 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.

We will also offer some proposals which avoid these deficiencies and which constitute candidate explanations for universal tendencies of speech sound sequences.

2. Problems with the sonority hierarchy

2.1. 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 determine 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; Haugen 1956a, b).

2.2. 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. We 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 they 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 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.[4]

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] C 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 4).

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.[5]

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

(9) The pattern in (8) is found in Japanese, 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, Akóõsē, Yao, Zulu, Wukari Jukun, Mambila, Parintintin, Yuma, Toura, Amharic, Bella Coola and others.

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

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

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) Ainu, Capanahua, Guahibo, Cavineña, Huichol, Korean, Ignaciano Moxo, Dan, Totonaco, Tenango Otomi, Sinhalese, Japanese, Trique, Bulgarian, Mixtec, Yao.

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

(14) Northern Estonian, Karakatšan.

An example of (12) is the fact that Mid-Atlantic and 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 stop + /l/ or the typical restriction of distinctive pharyngealization to [acute] consonants and its absence from [grave] consonants (Ohala 1985a).

2.3. *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, -kʃ, -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* [bəlo] by simply lengthening the duration of the /l/. (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 do the extra syllable and the perceived [ə] come from?

2.4. *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* [swit], *quick* [kwɪk], *puce* [pjus], *beaut* [bjut], etc., as shown above. 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/, that 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 [α voice] [α voice] such as /ml- st-/).

Regarding common exceptions to the prediction of the SH, e. g., the abundance of /st- -ts/ clusters, one possibility 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 cropping up in apparent epidemics, was different from microbially transmitted infectious diseases like measles and that it was in fact due to a deficiency of the vitamin 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: 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 two 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. We are unaware of any empirical support for the features

[mobile] and [stable]. Moreover, common clusters like /st, sl/ as well as affricates such as /t͡s, k͡x/ 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/.

We 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.

3. The first proposal

3.1. The hypothesis

There are three essential elements to the alternative we offer to account for common cross-language sequential constraints:[6]

(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.[7] 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.

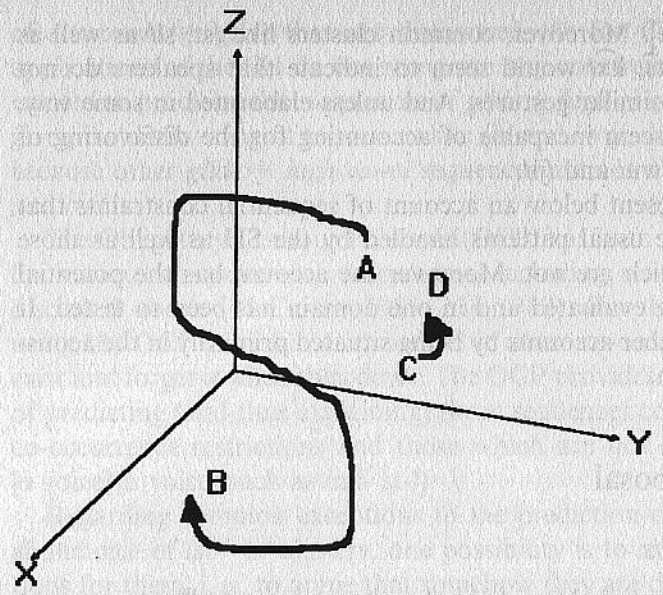

Figure 1. Schematic representation of the speech sound sequences' trajectories through the multi-dimensional acoustic space. X, Y, Z = dimensions of the acoustic space. Trajectory AB is better than trajectory CD because it is longer.

(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. This is represented schematically in Fig. 1.

All communication systems, whether writing, Morse code, semaphore, sign language, AM, FM, or PCM encodings of audio signals in radio and telephony involve modulations of some carrier signal and 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, 1984, 1989, 1990) 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 (listeners' 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 *stripe*. 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.

3.2. *A phonetic simulation*

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 we 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 ɛ ɑ 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 differs 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 × F2 × 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 had 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 /we/ 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/ (however, see below, section 3.4).

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

3.3. Extrapolations

Based on Kawasaki's study and extending it to include other acoustic parameters, one can imagine generating sound sequences by stringing together various sounds randomly as done in (16a–d).

(16) (a) sʃ pt ji wu ɥy θf
(b) pwe dje dle
(c) ske ble gre sts
(d) ba sa t͡ʃa

Given the discussion above, we 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 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. (See also Lindblom 1984.)

3.4. 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). (See second proposal below.)

In addition to the salience created by sound sequences, their degree of "predictability" should also be factored in. In some cases, this may be independent of acoustic salience. For example, the salience of the transitions into or out of nasal consonants is not greatly reduced by the nasalization on vowels, and yet sequences of *nasal vowel + nasal consonant* (or vice-versa) are rarely distinctive (Kawasaki 1986). Here 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 a theory of dissimilation (Ohala 1981, 1986, 1993).

Another factor contributing to absence of certain otherwise salient sequences is the existence of other acoustically similar sequences. Acoustically and auditorily similar sequences would be subject to confusion and thus merger. Combine this with asymmetries in the direction of confusion

(Ohala, 1983a, 1985b, in press) 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. Kawasaki tested this by estimating the similarity of sound sequences by their formant trajectories. By this metric, /dl/ and /gl/ were quite similar, and this may help to explain in part the disfavoring of /dl/.

Although we 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 which is required for voicing (Catford 1977: 26 ff; Ohala 1983b, 1994). 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", we acknowledge that there are some difficulties with this concept. In spite of the computational usefulness of distance and spatial metaphors for the comparison of speech sounds there is 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 Y is as close to and thus as confusable with X, i.e., the confusion should be mutual and symmetrical. A "cost" measure ("cost" by which one sound becomes like another) would be more appropriate to characterize the transformation of one sound into another (by the listener) (Ohala 1983a, 1985b, in press).

4. The second proposal

Our second proposed alternative to the SH is quite speculative and does not yet have much empirical support, but like the first proposal, above, is capable of being tested.

Stevens (1980) and Stevens–Keyser (1989) suggested that the features differentiating speech sounds may vary in their robustness or salience,

but instead of the magnitude of modulation of acoustic parameters he characterized robust features as those which could be detected in a short time window, c. 40 to 50 msec., whereas non-robust features would take much longer, perhaps as much as 100 msec. So "robust" = *rapid* and "non-robust" = *slow*. (Of course, rapid detection of distinguishing features is probably at least partially dependent on magnitude of formant trajectories.) Canonical examples of rapid features would be [± continuant] and [± voice], and of slow ones (we suppose), [± sharp] (palatalized / non-palatalized) and [± flat] (labialized / non-labialized or pharyngealized / non-pharyngealized). Stevens (1980) noted that languages with small segment inventories, e. g., many of the Malayo-Polynesian languages, tend to use the rapid features exclusively to differentiate their phonemes. Slow features – which would no doubt include all the secondary articulations as well as aspiration, glottalization, retroflexion, and the like – are found only in languages like Hindi, Navaho, and Georgian that have large segment inventories that also include sounds differentiated by the more rapid features. This is an important idea linking phonological universals and phonetics.

A strict continuum of rapid and slow features has not been posited yet, let alone tested, but it strikes us that any such continuum that would emerge from further study might characterize a continuum of segment types that would resemble to some extent the SH. Stops, which occupy one end of the SH, are certainly rapidly differentiated from other sounds by virtue of the rapid rise in amplitude they create through their bursts and/or the rapid onset of following continuants. Glides, which occupy the other end of the SH, are differentiated from other sounds by relatively slow amplitude and formant transitions (Liberman et al. 1956). Whether the other ordered segment types in the SH (fricatives, nasals, liquids) could also be shown to be differentiated by more rapid to less rapid features remains to be established, but it is not on the face of it implausible. We speculate further that there could be an inherent rationale for the sequencing of segments according to whether they are differentiated by fast to slower features: given the possibility of co-articulation of segments, it would be highly efficient to simultaneously initiate both rapid and slow sounds because they will be manifested in the speech signal at different times: first the rapid sound and then later the slow one. (Admittedly, this reasoning fails to account for the apparent inverse sequencing of segment types post-vocalically.)

Our first and second proposals are not necessarily mutually exclusive; they could both be operative in shaping sound sequences. That the sound

sequence exceed some threshold magnitude in the modulation of acoustic parameters should be the first requirement; that the successive modulations be ordered so that slow ones follow rapid ones would be the second requirement.

5. Is the syllable necessary?

5.1. Introduction

No mention was made of syllables in the above two proposals. We believe this is 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, 1992b). 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 reason for it. But we think it is important to realize that syllables are logically subsequent, not antecedent, in 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 "syllables first, then segments". (See also MacNeilage–Davis 1990.) The reverse position, "segments first, then syllables", which we tentatively adopt, is for the most part the established view in phonology today. Nevertheless, we believe it is important to try to reinforce its logical basis and not simply accept it uncritically.

5.2. 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 (assuming that they are). We do not 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. We are

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.

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

5.3. *Syllables as epiphenomena from concatenating acoustically distinct speech sounds*

The fact that speech consists of alternate openings and closings of the vocal tract (or of alternate increases and decreases of the loudness) does not require the concept of the syllable or a hierarchy specifying the preferred order of segment types for its explanation. Starting from a given configuration of the vocal tract, the criteria proposed above for a "good" contrast would stipulate that the next configuration should be one that creates some acoustic-auditory difference with respect to the current one. That could be achieved by opening the tract more (if it were not already maximally open) or by reducing the opening (if it were not already maximally closed). Having then moved to this second configuration, the criteria for moving to the next, the third, one, are the same as before, and so on for any sequence of speech sounds. If the maximum opening is reached at any point, then there is no other option but to continue by closing the vocal tract; likewise, if the minimum opening is reached at any point, there is no other option but to continue by opening the tract. Just by virtue of seeking detectable changes in the acoustic signal one would create as an epiphenomenon, i. e., automatically, a sequence showing local maxima and minima in vocal tract opening or loudness. In a similar way one could find "peaks" (local maxima) in a string of random numbers as long as each succeeding number in the sequence was different from the preceding one.

5.4. *Syllables are the domain for synchronization of the segmental and suprasegmental stream*

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 F0 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 a pedal is parallel to the ground and in the frontmost position. The neuromotor system that produces speech may have its own constraints, as does the bicycle and its rider, which motivates chunking of what one might theoretically think should be just a continuous stream. That skill somehow plays a role in syllable production is suggested by the fact that speakers of English can easily 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).

5.5. *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, we 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 /bə/, and of course the /lo/ transition is unchanged. Listeners don't require an actual /əl/ transition in order to conclude that there is a /ə/ 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. Menéndez Pidal (1926: 217–218) notes the existence in Spanish, sometimes sporadically, of a "*vocal relajada*" breaking up clusters of the sort Cl and Cr, e. g., *Ingalaterra*, *corónica*, *eg^{e}lesia*, *p^{e}redicto*. Similar examples can be found in many other languages. Something of the same

sort probably underlies the variable phonetic and phonemic analysis of syllabic [l̩ ɾ̩ n̩ m̩] in English as /əl/, /ən/, etc. Instrumental examination of words with these syllabic consonants as pronounced by most American English speakers usually gives little evidence of a separate vowel.

5.6. *Why is the CV structure so common?*

There are several indications that CV structures have a special status vis-à-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 commonsense 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 at 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 a 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. This means that coarticulation of a segment sequence X Y 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

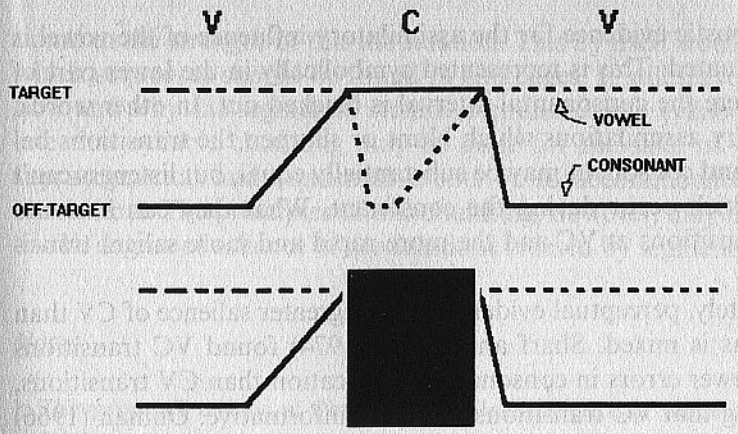

Figure 2. Schematic representation of consequences of asymmetry in anticipatory and perseveratory assimilation during a VCV utterance. TOP: dashed line represents vowel gestures; solid line, consonant gestures; horizontal axis is time; vertical axis represents degree of achievement of target configuration (top = target, bottom = off-target). Longer anticipatory assimilation symbolized as less steep slope of lines leading up to a target configuration; shorter perseveratory assimilation, as the steeper slopes of lines leading away from target. BOTTOM: same as upper figure but with the transitions during the consonant blacked out symbolizing their acoustic attenuation. The result is that the steeper, more salient, acoustic modulations will be heard at the CV not the VC juncture.

longer and slower approach to Y during X would make for a less strong modulation in terms of its rapidity. Translating this to a VCV sequence we would get the situation depicted schematically in Figure 2, 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 the 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 no asymmetry in them. However, the assimilations during the consonantal interval occur typically while substantial portions of the 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 blacked out. 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 those which occur during the consonant. What they can hear are 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. On the other hand, Kawasaki found the CV formant trajectories to be better differentiated acoustically among themselves than were the VC trajectories. Ohala (1990) presented original data and reviewed previous evidence that in intervocalic heterogeneous clusters of the sort /abga/ (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 the 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 it reflects special attention to articulation for the sake of maintaining the acoustic salience of the CV transition.

Finally, and perhaps most importantly, there is indication that the auditory system reacts more strongly to onsets than offsets of acoustic energy (Greenberg 1995).

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.

6. Conclusions

"Sonority" and its cousin "strength" do not exist and should be abandoned for the sake of explaining universal sequential constraints. They should be replaced by a measure of the degree of modulation in several acoustic parameters (amplitude, periodicity, spectral shape, F0) and the notion that the 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. It still leaves unanswered how and why the segmental stream is chunked into syllables. We 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

1. This paper is a revision of Ohala (1992a). We thank Natasha Warner for helpful comments.
2. 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 (521 ff).
3. Actually a few years earlier.
4. 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.
5. 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. See also Maddieson and Precoda (1992).
6. See previous discussion of this in Ohala 1980, Kawasaki 1982, Kawasaki-Fukumori 1992, and Ohala and Kawasaki 1984.
7. Bell and Hooper (1978: 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

Asher, R. E.–J. M. Y. Simpson (eds.)
1994 *The encyclopedia of language and linguistics.* Oxford: Pergamon.

Awedyk, Wiesław
1975 *The syllable theory and Old English phonology.* [Polska Akademia Nauk. Komitet Neofilologiczny] Wrocław: Wydawnictwo Polskiej Akademii Nauk.

Bell, Alan–Joan B. Hooper (eds.)
1978 *Syllables and segments*. Amsterdam: North Holland.
Borowsky, Toni J.
1986 *Topics in the lexical phonology of English*. [Unpublished doctoral dissertation, University of Massachusetts, Amherst.]
van den Broecke, Marcel P. R.–Antonie Cohen (eds.)
1984 *Proceedings of the 10th International Congress of Phonetic Sciences, Utrecht*. Dordrecht: Foris.
de Brosches, Charles
1765 *Traité de la formation méchanique des langues, et de principes physiques de l'étymologie*. 2 vols. Paris: Chez Saillant, Vincent, Desaint.
Bruck, Anthony–Robert A. Fox–Michael W. La Galy (eds.)
1974 *Papers from the Parasession on Natural Phonology, April 18, 1974*. Chicago: Chicago Linguistic Society.
Catford, John C.
1977 *Fundamental problems in phonetics*. Bloomington: Indiana University Press.
Crothers, John–James Lorentz–Don Sherman–Marilyn Vihman
1979 *Handbook of phonological data from a sample of the world's languages: a report of the Stanford Phonology Archive*. Language Universals Project: Stanford.
Cutting, James
1975 "Predicting initial cluster frequencies by phonemic difference", *Haskins Laboratories, Status Report on Speech Research*. SR-42/43. 233–239.
De Mori, Renato–Ching Y. Suen (eds.)
1985 *New systems and architectures for automatic speech recognition and synthesis*. [NATO ASI Series, Series F: Computer and System Sciences, Vol. 16] Berlin: Springer-Verlag.
Diver, William
1975 "Phonology as human behavior", *Columbia University Working Papers in Linguistics* 2: 27–57.
Docherty, Gerard J.–D. Robert Ladd (eds.)
1992 *Papers in Laboratory Phonology II: Gesture, segment, prosody*. Cambridge: Cambridge University Press.
Dudley, Homer
1940 "The carrier nature of speech", *The Bell System Technical Journal* 19: 495–515.
Elsendoorn, Ben A. G.–Herman Bouma (eds.)
1989 *Working models of human perception*. London: Academic Press.
Fromkin, Victoria A. (ed.)
1985 *Phonetic linguistics. Essays in honor of Peter Ladefoged*. Orlando, FL: Academic Press.
Greenberg, Steven
1995 "The ears have it: the auditory basis of speech perception", *Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, 13–19 August 1995*. Vol. 3: 34–41.
Halle, Morris–Horace Lunt–Hugh McLean (eds.)
1956 *For Roman Jakobson*. The Hague: Mouton.
Hankamer, Jorge–Judith Aissen
1974 "The sonority hierarchy", in: Anthony Bruck–Robert A. Fox–Michael W. La Galy (eds.), 131–145.

Hardcastle, William J.–Alain Marchal (eds.)
1990 *Speech production and speech modelling*. Dordrecht, The Netherlands: Kluwer.
Hattori, Shirô–Kazuko Inoue (eds.)
1983 *Proceedings of the XIIIth International Congress of Linguists, Tokyo, 29 Aug. – 4 Sept. 1982*. Tokyo. [Distributed by Sanseido Shoten.]
Haugen, Einar
1956a "The syllable in linguistic description", in: Morris Halle–Horace Lunt–Hugh McLean (eds.), 213–221.
1956b "Syllabification in Kutenai", *International Journal of American Linguistics* 22: 196–201.
Janson, Tore
1986 "Cross-linguistic trends in the frequency of CV sequences", *Phonology Yearbook* 3: 179–195.
Javkin, Hector R.
1979 "Phonetic universals and phonological change", *Report of the Phonology Laboratory* (Berkeley) No. 4.
Jespersen, Otto
1897–1899 *Fonetik: en systematisk fremstilling af læren om sproglyd*. København: Det Schubotheske Forlag.
1904 *Phonetische Grundfragen*. Leipzig–Berlin: Teubner.
Jones, Charles (ed.)
1993 *Historical linguistics: Problems and perspectives*. London: Longman.
Kawasaki, Haruko
1982 *An acoustical basis for universal constraints on sound sequences*. [Unpublished doctoral dissertation, Univ. Calif., Berkeley.]
1986 "Phonetic explanation for phonological universals: The case of distinctive vowel nasalization", in: John J. Ohala–Jeri J. Jaeger (eds.), 81–103.
Kawasaki-Fukumori, Haruko
1992 "An acoustical basis for universal phonotactic constraints", *Language & Speech* 35: 73–86.
Kingston, John–Mary Beckman (eds.)
1990 *Papers in Laboratory Phonology I: Between the grammar and the physics of speech*. Cambridge: Cambridge University Press.
Kiritani, Shigeru–Hajime Hirose–Hiroya Fujisaki (eds.)
in press *Festschrift for Osamu Fujimura*. Berlin: Mouton de Gruyter.
Lass, Norman J. (ed.)
1980 *Speech and language. Advances in basic research and practice*. Vol. 3. New York: Academic Press.
Liberman, Alvin M.–Pierre C. Delattre–Louis J. Gerstman–Franklin S. Cooper
1956 "Tempo of frequency change as a cue for distinguishing classes of speech sounds", *Journal of Experimental Psychology* 52: 127–137.
Lindblom, Björn
1983 "Economy of speech gestures", in: Peter F. MacNeilage (ed.), 217–245.
1984 "Can the models of evolutionary biology be applied to phonetic problems?", in: Marcel P. R. van den Broecke–Antonie Cohen (eds.), 67–81.
1989 "Phonetic invariance and the adaptive nature of speech", in: Ben A. G. Elsendoorn–Herman Bouma (eds.), 139–173.

1990 "On the notion of 'possible speech sound'", *Journal of Phonetics* 18: 135–152.
MacNeilage, Peter F. (ed.)
1983 *The production of speech.* New York: Springer-Verlag.
MacNeilage, Peter F.–Barbara L. Davis
1990 "Acquisition of speech production: the achievement of segmental independence", in: William J. Hardcastle–Alain Marchal (eds.), 55–68.
Maddieson, Ian–Kristin Precoda
1992 "Syllable structure and phonetic models", *Phonology* 9: 45–60.
Mandelbrot, Benoit
1954 "Structure formelle des textes et communication", *Word* 10: 1–27.
Masek, Carrie S.–Robert A. Hendrick–Mary Frances Miller (eds.)
1981 *Papers from the Parasession on Language and Behavior.* Chicago: Chicago Linguistic Society.
McCarthy, John J.
1989 "Feature geometry and dependency: A review", *Phonetica* 45: 84–108.
Menéndez Pidal, R.
1926 *Orígenes del español.* Madrid: Hernando.
Ohala, John J.
1980 "The application of phonological universals in speech pathology", in: Norman J. Lass (ed.), 75–97.
1981 "The listener as a source of sound change", in: Carrie S. Masek–Robert A. Hendrick–Mary Frances Miller (eds.), 178–203.
1983a "The phonological end justifies any means", in: Shirô Hattori–Kazuko Inoue (eds.), 232–243.
1983b "The origin of sound patterns in vocal tract constraints", in: Peter F. MacNeilage (ed.), 189–216.
1985a "Around *flat*", in: Victoria A. Fromkin (ed.), 223–241.
1985b "Linguistics and automatic speech processing", in: Renato De Mori–Ching Y. Suen (eds.), 447–475.
1986 "Phonological evidence for top-down processing in speech perception", in: Joseph S. Perkell–Dennis H. Klatt (eds.), 386–397.
1990 "The phonetics and phonology of aspects of assimilation", in: John Kingston–Mary Beckman (eds.), 258–275.
1992a "Alternatives to the sonority hierarchy for explaining segmental sequential constraints", *Papers from the Parasession on the Syllable.* Chicago: Chicago Linguistic Society, 319–338.
1992b "The segment: Primitive or derived?", in: Gerard J. Docherty–D. Robert Ladd (eds.), 166–183.
1993 "The phonetics of sound change", in: Charles Jones (ed.), 237–278.
1994 "Speech aerodynamics", in: R. E. Asher–J. M. Y. Simpson (eds.), 4144–4148.
in press "Comparison of speech sounds: Distance vs. cost metrics", in: Shigeru Kiritani–Hajime Hirose–Hiroya Fujisaki (eds.).
Ohala, John J.–Jeri J. Jaeger (eds.)
1986 *Experimental phonology.* Orlando, FL: Academic Press.
Ohala, John J.–Haruko Kawasaki
1984 "Phonetics and prosodic phonology", *Phonology Yearbook* 1: 113–127.

Ohala, John J.–James Lorentz
1977 "The story of [w]: an exercise in the phonetic explanation for sound patterns", *Berkeley Linguistic Society, Proceedings, Annual Meeting* 3: 577–599.
Öhman, Sven E. G.
1966 "Perception of segments of VCCV utterances", *Journal of the Acoustical Society of America* 40: 979–988.
Perkell, Joseph S.–Dennis H. Klatt (eds.)
1986 *Invariance and variability in speech processes.* Hillsdale, NJ: Lawrence Erlbaum.
Price, Patti J.
1980 "Sonority and syllabicity: Acoustic correlates of perception", *Phonetica* 37: 327–343.
Pulgram, Ernst
1970 *Syllable, word, nexus, cursus.* The Hague: Mouton.
Ruhlen, Merrit
1976 *A guide to the languages of the world.* Stanford.
Saporta, Sol
1955 "Frequency of consonant clusters", *Language* 31: 25–30.
de Saussure, Ferdinand
1916 *Cours de linguistique générale.* Paris: Payot.
Sharf, Donald J.–Robert C. Beiter
1974 "Identification of consonants from formant transitions presented forward and backward", *Language & Speech* 17: 110–118.
Sievers, Eduard
1893 *Grundzüge der Phonetik.* (3rd edition.) Leipzig: Breitkopf & Härtel.
Steriade, Donca
1982 *Greek prosodies and the nature of syllabification.* [Unpublished doctoral dissertation, M.I.T.]
Stetson, Raymond H.
1928 *Motor phonetics.* La Haye: Martinus Nijhoff.
[1951] [2nd edition. Amsterdam: North Holland Publishing Co.]
[1988] [Retrospective edition by: J. A. Scott Kelso–Kevin G. Munhall (eds.). Boston: College-Hill.].
Stevens, Kenneth N.
1980 Discussion. *Proceedings, Ninth International Congress of Phonetic Sciences.* Vol. 3. Copenhagen: Institute of Phonetics, University of Copenhagen. 185–186.
Stevens, Kenneth N.–Samuel J. Keyser
1989 "Primary features and their enhancement in consonants", *Language* 65: 81–106.
Tuller, Betty–J. A. Scott Kelso–Katherine S. Harris
1982 "Interarticular phasing as an index of temporal regularity in speech", *Journal of Experimental Psychology. Human Perception & Performance* 8: 460–472.
Whitney, William Dwight
1874 *Oriental and linguistic studies.* Second Series. New York: Scribner, Armstrong, & Co.
Yip, Moira
1989 "Feature geometry and cooccurrence restrictions", *Phonology* 6: 349–374.