Relating the sonority hierarchy to articulatory timing patterns
A cross-linguistic perspective

The sonority hierarchy is a central concept in phonology, one that is, arguably, not theory-dependent. It captures a robust typological generalization about preferred syllable structures cross-linguistically. This generalization is present, in slightly different formulations, in every phonological model of the major theoretical approaches – structuralist, functionalist, generativist¹. Equally robust, however, are the exceptions to this generalization, which are often hard to interpret and consequently hard to account for. Famous “troublemakers” in this category are, for example, Tashlhiyt Berber (see Dell and Elmedlaoui 2002 for theoretical implications) or Salish (Bagemihl 1991; Shahin and Blake 2004; Bird & Czaykowska-Higgins, this volume). Consonant sequences in these languages defy syllabification algorithms and principles of syllable organization based on sonority. Such data have challenged and catalyzed research, and depending on their accessibility, they have inspired new directions of study, improving our understanding of the syllable as a unit of information, processing, and production.

The goal of this chapter is to consider what sonority-based syllabic organization may mean when examined from an articulatory perspective. I propose that we have good reason to believe, based on results of experimental studies of consonant sequences in a variety of languages, that the sonority hierarchy can be best understood in its relation to articulatory timing. I will argue that the organizational role that has been attributed to the sonority hierarchy follows from language-specific properties of articulatory timing. The idea of the link between sonority and articulatory timing is not new. It is explicitly formulated by Mattingly (1981, 1998) in the concept of “parallel transmission” of information. Parallel transmission captures the essential coarticulatory properties of the speech signal that are crucial for maximum intelligibility and maximum speed in communication. We have now gained sufficient empirical knowledge to evaluate whether timing patterns of articulatory gestures can be related to syllabic organization via the concept of parallel transmission. The sonority hierarchy, as we know it, may have limited predictive power. It certainly captures one way of maximizing parallel transmission. But other options are attested in the world’s languages, and are predicted by aspects of articulatory timing. From a slightly different perspective, based on the articulatory study of liquids in American-English, Proctor and Walker (2012) have also proposed that understanding the relative sonority of segments can gain from considering their articulatory properties beyond degree of constriction.

The view expressed here assumes that the syllable exists and is an indispensable unit of linguistic organization (see Blevins 1995, Goldsmith 2011, for comprehensive reviews). Recent neurophysiological studies have inspired models that reinforce the role of syllable-sized units in speech processing (Doelling et al., 2014; Ghitza 2011, 2013; Giraud and Poeppel 2010; Howard and Poeppel 2012). The role of signal modulations is

¹ The strict CV model of phonology (Lowenstamm 1996), emerged from Government Phonology (Kaye et al., 1990; Kaye 1990), may be considered an exception, because sonority relations are rendered superfluous by the canonical CVCV syllable structure. However, Scheer (1996) proposed a theory of consonantal interactions, needed to account for restrictions on word-initial consonant clusters.
highlighted in a particularly relevant way in Ghitza’s (2011) Tempo model, where syllable-sized units with prominent energy peaks are tracked by the peripheral auditory system in the decoding process. Speech production models (Guenther 1995; Guenther et al., 2006; Bohland et al., 2010; MacNeilage and Davis 2000; Nam et al., 2009; Tilsen 2013) all include a model of the syllable. In terms of phonological representations, the proposal made here is consistent with the main tenets of Articulatory Phonology, henceforth AP (Browman and Goldstein 1992; Goldstein and Fowler 2003), and is related to the representation of syllabic organization in AP.

The chapter is organized as follows: in section 1 I introduce the sonority hierarchy as a phonological organizational principle of syllable structure. In section 2 I present three case studies of syllabification (Georgian, Slovak, Tashlhiyt) in terms of the traditional sonority hierarchy. In section 3 I discuss the same patterns with respect to properties of their articulatory timing. In section 4 I propose the connection between the sonority hierarchy and parallel transmission in articulatory terms. Section 5 contains a final discussion and conclusions.

1. Sonority – phonology and phonetics

All definitions of the sonority hierarchy (Jespersen 1899, to Clements 1990, to subsequent work) consider that an alternation of peaks and troughs corresponds to the preferred cross-linguistic ordering of manner classes by degree of constriction (see Parker 2002 for a full chronological review of this concept). The preferred order is one of rising sonority in a syllable onset (obstruent < nasal < liquid < glide < vowel), and the opposite in a syllable rhyme. This generalization captures the fact that kla is preferred over kta, which is in turn preferred over lka.

The acoustic correlates of the peak/trough alternation are understood as maximal modulations in multiple parameters that vary simultaneously: amplitude, periodicity, spectral shape, F0 (Ohala & Kawasaki-Fukumori 1984; Ohala 1992; Parker 2002; Clements 2009). These acoustic correlates underlie speech intelligibility; however, they are not sufficient for characterizing the syllable as a unit of linguistic organization. They do predict the typologically common patterns of rising and falling sonority relative to the nucleus (e.g., kla). At the same time, there is an inconsistency between the generalization and the acoustic correlate of maximal modulation because the latter also predicts the less common sonority reversals such as lka. These are also alternations of troughs and peaks, whether the liquid is syllabic or not, therefore kla, lka should both be preferred over a sonority plateau kita. Sonority plateaus are not predicted, and yet, they are cross-linguistically more common than reversals.

Syllabic organization involves two related questions. One is quantitative, and concerns the number of elements that can be contained in a syllable onset, while the second one concerns the ordering of these elements. Cross-linguistically, an ordering that follows the sonority hierarchy results in a larger number of elements accommodated in a syllable onset. Many languages allow only one element in a syllable onset. Languages that allow more than one element usually follow the sonority hierarchy. Others, however, also allow sonority plateaus and reversals. In current phonological models the latter are treated as exceptions, and it is not clear that they need to be. I propose that a careful consideration of the articulatory timing patterns in some of the languages that have been
experimentally studied reveal new, more efficient generalizations regarding syllabic organization.

I argue that articulatory correlates relating to timing patterns between gestures can provide a deeper insight into organizational principles underlying the syllable because they can predict both common and uncommon patterns. The different acoustic consequences of these timing patterns help refine these predictions on a language-specific basis. For example, variation in timing lag between adjacent consonantal gestures can affect the way in which information about these consonants is maintained in the acoustic signal, and consequently perceived by the listener. Thus, a given, short lag duration may have different effects on stop-liquid and stop-stop sequences. For a given lag duration, a stop-liquid sequence such as /kl/ will consist of a constriction with an acoustic release and formant transitions into an open constriction. But for the same short lag duration, a stop-stop sequence such as /kt/ may lose information about C1 /k/. If the tongue tip gesture for /t/ begins before the release of /k/, then /k/ will be partly hidden by /t/. It will not have an acoustic release. Unless the /kt/ cluster is intervocalic, acoustic information about /k/ is only present in its release burst. The absence of an acoustic release may thus lead to misperception. At the same time, if the lag is too long in either a stop-liquid or a stop-stop sequence, in the presence of voicing, a vocalic transition can emerge between the two consonantal gestures. In this case the percept of a consonant sequence may be lost and replaced with a CV alternation.

As already mentioned, the generalizations that I propose are consistent with a phonological analysis of syllabic organization in the framework of AP. Therefore, before considering the three case studies in the next section, I first introduce the two main tenets of AP:

- articulatory gestures are at once discrete, combinatorial units of representation and units of continuous action in space and time;
- gestures have a temporal dimension. The discrete specifications of gestures are dynamic, and pairs of gestures are dynamically coupled. The coupling accounts for contextual variation. Common processes such as assimilation, insertion, deletion, are accounted for by variable relative timing between gestures.

Finally, in its current form (Goldstein and Fowler 2003) AP is compatible with, but does not crucially assume, the view that articulatory, rather than acoustic events, are the objects of speech perception.

My proposal is based on patterns of syllable structure in three languages – Georgian, Slovak, Tashlhiy and the articulatory timing patterns that characterize them. The relevant data are presented in section 2, and the experimental results in section 3.

2. Sonority – three representative patterns

Syllable organization is compared in three languages: Georgian (Chitoran et al., 2002; Chitoran & Goldstein 2006), Slovak (Pouplier and Beňuš 2011), and Tashlhiyt (Ridouane 2008; Ridouane & Fougeron 2011). All three languages allow sequences of consonants, but their organization into syllables differs in terms of their behavior as syllable nuclei vs. syllable margins. Thus, Georgian allows only vowels as nuclei, Slovak allows vowels and liquids, and in Tashlhiy vowels and all consonants can be syllabic. The three syllabification patterns are illustrated below:
(1) Syllabification patterns according to the nature of the nucleus (nuclei are in bold):

a. Georgian – vowels only:
   - rbe.na ‘to run’
   - mtkna.re.ba ‘yawn’
   - t’k’bi.li ‘sweet’

b. Slovak – vowels, liquids:
   - mrak ‘darkness’
   - mřk ‘wink’
   - smrk ‘sniff’
   - blb ‘stupid guy’

c. Tashlhiyt – vowels, all consonants:
   - s.mun ‘accompany’ caus.
   - ts.mun ‘accompany’ 3fs. caus.

Considering the data in terms of the traditional definition of the sonority hierarchy, it follows that the relative sonority of consonantal segments is language-specific. Thus, in Georgian all consonantal segments are equivalent to one another in terms of sonority and all are equally distant from the vowels, because they can appear in any order in a syllable/word onset. This is not true of Slovak or Tashlhiyt, where sonority reversals are disallowed. In these languages, instead, we find a more fine-grained organization. The Slovak pattern suggests that liquids share the high sonority of vowels, so that when an actual vowel is missing, they can take over as nuclei. By the same criteria, in Tashlhiyt all consonants share the sonority of vowels.

The three languages consequently differ in their tolerance for complex onsets. Tashlhiyt is the most restricted, allowing only one segment in a syllable onset. Georgian is the most liberal, as it not only allows complex onsets, but it allows many combinations within an onset, whether they observe or violate the sonority hierarchy. The three patterns are summarized in (2). We see that an increase in the number of elements that can occupy a syllable nucleus implies a decrease in the complexity of the onset. Thus, Tashlhiyt strictly disallows complex onsets, but allows any segment in nucleus position. Georgian shows exactly the reverse pattern, with high onset complexity but only one type of segment (vowels) in nucleus position.

(2) A typology of nuclei and phonotactic (onset) complexity:

<table>
<thead>
<tr>
<th></th>
<th>Georgian</th>
<th>Slovak</th>
<th>Tashlhiyt</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NUCLEI</strong></td>
<td>vowels</td>
<td>vowels</td>
<td>vowels</td>
</tr>
<tr>
<td><strong>ONSETS</strong></td>
<td>complex onsets</td>
<td>syllabic liquids</td>
<td>syllabic liquids</td>
</tr>
<tr>
<td></td>
<td>sonority plateaus</td>
<td></td>
<td>syllabic nasals</td>
</tr>
<tr>
<td></td>
<td>sonority reversals</td>
<td></td>
<td>syllabic obstruents</td>
</tr>
<tr>
<td></td>
<td>k-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>kr-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rb-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mtkn-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>t’k’b-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>k-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>kr-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mř-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>t-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Such generalizations, however, cease to be useful beyond a purely descriptive level. They require further explanation as to why a segment can be more or less sonorous relative to another, on a language-specific basis. Why do we see in (2) an inverse relation between the nature of the elements that can fill a syllable nucleus (increasing from left to right in row 1) and the combinatorial complexity in a syllable onset (decreasing from left to right in row 2)?

I propose to examine the same data by comparing the syllabification patterns to the articulatory timing patterns characterizing sequences of consonants in the three languages. I turn to this comparison in the next section.

3. Articulatory timing – three representative patterns

I present in this section the relevant results from previous articulometer (EMA) studies on timing coordination in Georgian, Slovak, and Tashlhiyt. I begin by defining the four landmarks of an articulatory gesture, on which EMA measures of relative timing are based.

(3) Landmarks of an articulatory gesture:

1 – gesture onset
2 – target onset = the point where the target constriction is reached
3 – target offset = the point where the constriction is released
4 – gesture offset

The particular context that interests us is the sequencing of consonantal gestures. The relative timing of these gestures has been shown to vary as a function of segmental, prosodic, speech style factors. Syllable position is a major factor that determines whether gestures are more or less overlapped in speech. For example, in a complex onset, as consonants are added to an onset (CV vs. CCV vs. CCCV), the rightmost consonant is expected to shift progressively closer to the nucleus vowel. The timing lag between the rightmost consonant and the vowel is expected to shorten. This rightward shift is interpreted as resulting from the multiple coordination patterns characterizing a complex onset: it is specifically hypothesized that the consonants in an onset are coupled to each other in anti-phase mode (moving away from each other), as well as being each coupled in-phase (moving closer) to the following vowel (Browman & Goldstein, 2000). A rightward shift is thus predicted in an onset if such a compromise exists between competing targets. Regardless of how many consonants are in an onset, the onset as a whole maintains a stable coordination relationship with the nucleus vowel. This hypothesis has been confirmed for a number of languages. The in-phase coupling is interpreted as characterizing a syllable onset.

In what follows I will review against this background the characteristic timing patterns for each of the three languages examined.

3.1 Georgian

Kinematic studies of Georgian consonant sequences revealed several patterns that are relevant for syllabic organization. The data were obtained with an electromagnetic
articulometer (EMA) system (Perkell et al., 1992). It was found that Georgian has a timing pattern sensitive to the order of constriction location (Chitoran et al., 2002; Chitoran and Goldstein 2006). Thus, front-to-back sequences (e.g., gb, gd, rk, pl), where the release of an anterior constriction is followed by a posterior constriction in the vocal tract, are systematically more overlapped, more co-produced, than back-to-front sequences (e.g., gb, gd, rb, kl). In the latter, a posterior constriction is released into a vocal tract that is still closed by an anterior constriction. Such back-to-front sequences were found to have a longer lag between the two C gestures. They are less overlapped, and in sequences where at least one consonant is voiced, a vocalic transition is visible in the acoustic signal. The vocalic transition is present predominantly in back-to-front stop-stop sequences (58%) such as gd-eba ‘to be thrown’, g-ber-av-s ‘is inflating you’, k’bil-i ‘tooth’, a-gd-eb-a ‘throw in the air’, da-gb-er-a ‘inflate’, those with minimal overlap. Only 23% of front-to-back stop-stop sequences show a similar vocalic transition, in the forms: bgera ‘sound’, dg-eb-a ‘stands up’, abga ‘saddle bag’, a-dg-eb-a ‘will stand up’. Figure 1 shows an example of a back-to-front stop-stop sequence with a long lag and a vocalic transition.

FIGURE 1. Example of long lag and vocalic transition in Georgian #gb in g-ber-av-s

A subsequent study (Goldstein et al., 2007) has shown that, when such a vocalic transition is present, there is no rightward shift, as mentioned above, indicative of a complex onset. This suggests that, in Georgian, a consonant that is part of a complex onset may be timed as a single onset relative to this vocalic transition. In ts’k’riala ‘shiny’ a vocalic transition can occur between [k’] and [r], and when it does, the three consonants in the sequence do not show the coordination pattern of an onset cluster. The relevant measure considered is the distance between the time when the tongue tip reaches the constriction for [r] and the time when the tongue body reaches the target for the vowel [i]. This distance is compared in three forms: riali ‘commotion’, k’riali ‘glitter’, ts’k’riala ‘shiny’. If [k’r] and [ts’k’r] were complex onsets, it would be expected that the target onset of [r] would move progressively closer to the target onset of [i], as more consonants are added to the onset.

For one Georgian speaker a rightward shift was not found between [k’riali] and [ts’k’riala]. This suggests that the mode of coordination in [ts’k’riala] is not that of a complex onset. Since these cases also show a vocalic transition, the absence of a rightward shift suggests that the vocalic transition is acting as a syllable nucleus itself. This is not a systematic result, and a definitive conclusion is premature. We also do not know yet whether the vocalic transitions are due to long timing lags or to characteristics of a voiced C1 release when C1 is a stop, or to both. Nevertheless, it is interesting to speculate about the possible phonologization of vocalic transitions. If vocalic transitions are generalized, are actively perceived by native speakers, and are associated with the specific timing pattern of Georgian CC sequences, it is conceivable that over time an active vowel gesture may be produced, that acts as a syllable nucleus. Thus, a particular coordination mode, one that favors long lags (a low degree of overlap), would have an acoustic consequence that corresponds to the signal modulation preferred in a sonority profile.
3.2 Slovak

The main difference between Slovak and Georgian is that in Slovak the liquids /l/ and /ɾ/ can be syllabic. The relevant results for Slovak (Pouplier and Behuš 2011) show consistent differences in timing patterns between syllables with consonantal nuclei and those with vocalic nuclei. Articulatory (EMA) coordination measures were taken for the following comparisons:

Onset-nucleus coordination: CV (bib) vs. CL (blb, brm)
Complex onset coordination: CV (lak, raky, mog) vs. CCV (vlak, braky, smog)

CC coordination: CCV (klak, mrak) vs. CLC / CLC (klk, krk) vs. VCC (kalk, park)

Degree of overlap is measured by the duration of the plateau lag, given by the distance from the target offset (release) of C1 to the target onset (achievement of the target constriction) for C2. A longer lag indicates less overlap between the target plateaus of the two consonants. The relevant results are summarized below.

In terms of onset-nucleus coordination, the authors compared the peak velocity lag of the closing movement in bib vs. blb, brm. Peak velocity is the time point of maximum articulator velocity before the articulator reaches its target. No difference was found between bib with a vocalic nucleus and blb with a consonantal nucleus, but there was a significant difference between blb and brm: syllabic /ɾ/ has the longest peak velocity lag, differing significantly from syllabic /l/.

With respect to complex onset coordination, both vocalic (CV, CCV) and consonantal (CL, CCL) nuclei behave the same way. Neither shows the rightward shift expected as consonants are added to the onset. Although this result is surprising, two additional measures did show the expected difference in timing between simple and complex onsets. A significant difference was found between vocalic and consonantal nuclei in terms of the size of the effect. Syllables with vocalic nuclei showed a significantly greater change between simple and complex onset than syllables with syllabic consonants. Slovak stop-liquid sequences show generally little overlap. With respect to CC coordination, onset-nucleus stop-liquid sequences as in krb, klk were found to be less overlapped than the same stop-liquid sequences in onset-onset (krab, klak) or coda-coda cases (park, kalk). Onset-nucleus CL sequences (krb, klk) were less overlapped that nucleus-coda LC sequences (krk, klk).

The authors note that in both the onset-nucleus (krb) and the onset-onset (krab) cases, the reduced overlap often results in open transitions with vocoids emerging between the release of the stop and the tongue tip gesture of the liquid. The vocoid is attributed to the tongue body retraction gesture that is characteristic of liquids. In Slovak, /l/ is dark, having both an apical raising gesture and a tongue body retraction gesture. The apical trill /ɾ/ also shows an apical gesture and a retraction of the tongue dorsum. The vocalic transition is the acoustic consequence of the long timing lag between the onset consonant and the tongue tip gesture of the liquid, whether the liquid is in nucleus or onset position. It is therefore not the transitional vocoid itself that provides a sonority peak, but rather the substantial timing lag favoring it. The authors conclude that in a syllabic liquid the vocalic retraction gesture does not simply act as a vowel. Instead, both the consonantal tongue tip gesture and the vocalic retraction gesture together belong to
the nucleus, and timing between onset and nucleus is such that it results in an acoustic vocalic element intervening between two consonants.

Once again, similarly to the case of Georgian, a language-specific pattern of articulatory timing results in open transitions. In Slovak, this timing pattern is found for obstruents and liquids, whether they coordinate as onset-onset or onset-nucleus, in a way that is consistent with a rising sonority profile in syllable onsets.

### 3.3. Tashlhiyt

Tashlhiyt is radically different. In Tashlhiyt all consonants, not just liquids, can be syllabic. Complex onsets do not exist. A sequence of two different consonants is produced as an onset-nucleus sequence, even when a complex onset would actually be consistent with the sonority hierarchy. Thus, a form such as *gli* ‘guide’ is obligatorily disyllabic (*g.li*). A timing pattern consistent with this phonotactic generalization is thus expected. Tashlhiyt has been experimentally studied by Fougeron and Ridouane (2008), Ridouane and Fougeron (2011), Hermes et al. (2011), Ridouane et al. (2014).

Fougeron and Ridouane (2008) use electropalatography (EPG) data (Hardcastle and Roach 1979) from one native speaker to study the kinematic properties of syllabic consonants and their timing relationships with surrounding segments as a function of syllable position. Their findings indicate that consonants in nucleus position are less variable than in onset or coda. They are less overlapped by a following non-coda consonant, and relatively more overlapped by a preceding onset consonant. There is more EPG contact in nucleus than in onset or coda position. In a detailed acoustic study of data from five more speakers, Ridouane and Fougeron (2011) focus on the vocalic transitions that characterize the Tashlhiyt timing pattern. The data examined consist of word-initial CC sequences. A vocalic transition is present if at least one of the consonants is voiced (especially the second one), and if the vocal tract is sufficiently open in the transition between the two consonants. They are found most frequently at C1 release, in sequences with minimal overlap. The presence or absence of a vocalic transition does not affect the acoustic duration of the onset cluster, proving that they are intrusive (excrecent) schwas rather than full phonologically inserted vowels.

Finally, the analysis of EMA data (Hermes et al., 2011) confirms the timing of simple onsets in a CC sequence, based on the absence of rightward shift in data such as *fik* ‘give yourself’ vs. *kfik* ‘give yourself’, vs. *tkfik* ‘she gave you’. The results are therefore consistent with those of Goldstein et al. (2007). The rightmost consonant remains stable in its timing relation to the nucleus /i/, as more consonants are added to its left. The exact same pattern is found for consonantal nuclei, as in *fnk* ‘they gave you’ vs. *kfnk* ‘they gave you’ vs. *tkfnk* ‘she buried you’.

The three language-specific timing patterns examined here all share the presence of minimal gestural overlap, which may be accompanied by vocalic transitions. These occur in back-to-front sequences in Georgian (*g’beravs*), before liquids with a retraction gesture in Slovak (*k’rb*, *k’rab*) and possibly in Georgian as well (*ts’k’rial*). In Tashlhiyt the vocalic transitions are the most frequent among the three languages, and are primarily related to voicing. I argue that this common characteristic that emerges from the cross-linguistic comparison is worth considering and subjecting to further testing because it can be highly informative with respect to phonotactic typology.
The sonority hierarchy falls short of explaining the combinatorial possibilities encountered in all three languages. Descriptively, the three languages differ with respect to the sonority hierarchy: many of the Georgian complex onsets violate the sonority profile, while the ones in Slovak do not. Tashlhiyt does not violate sonority because it does not have complex onsets. But Tashlhiyt challenges the sonority hierarchy with the following question: why is an obstruent-liquid sequence such as gli, with a perfect rising sonority profile, not parsed as a complex onset? The sonority hierarchy alone cannot account for this.

In the following section I argue that a valid organizational principle underlying the syllable is one that simultaneously enhances signal modulation and coarticulation as parallel transmission, in ways that include, but are not limited to, the sonority hierarchy.

4. Sonority and parallel transmission

“The overlapping of multiple gestures in speech makes possible parallel, hence rapid, transmission of information” (Mattingly 1998: 276). Mattingly explicitly proposes that the syllable organization based on sonority ranking can be interpreted as corresponding to the requirement for efficient speech communication. Parallel transmission is maximized “if less open constrictions are being released or applied in the presence of more open constrictions” (Mattingly 1981: 418). The ordering of constriction degrees corresponds to the preferred ordering of manner classes in a syllable onset: obstruent > nasal > liquid > glide > vowel. But it also corresponds more generally to other ways in which information can be encoded during constrictions and constriction releases, if variation in timing patterns is considered. Parallel transmission is ensured by coarticulation. Coarticulation is maximized when gestures are most co-produced, provided they do not obscure one other. Thus, releasing less open constrictions into more open ones favors maximum coarticulation with minimal loss of information as to the identity of the gestures.

I propose that the attested cross-linguistic diversity of combinatorial restrictions, including the sonority hierarchy, follows from any possible coordination pattern that allows gestures to be maximally co-produced, and that also allows maximal modulation of the signal. Such a view covers the very common patterns that are in agreement with the sonority hierarchy, but it explains in addition why other linguistic patterns can also develop in some of the world’s languages, where sonority reversals and plateaus are well-formed. The preferred sequencing of applied constrictions is a sequencing that allows tighter intergestural coordination and faster transmission. This is the case of sequences like pla or kra, and it corresponds to the order of the sonority hierarchy, where more closed constrictions are released into gradually more open ones, allowing more gestural overlap. But if the order of constriction degree is reversed (as for mpa, rka), for example by the addition of consonantal prefixes, in this case some languages can parse the signal differently, preferring longer lags. When a C1 constriction is not released into a more open C2 constriction, a low degree of overlap allows the first constriction to have a release, which in and of itself contributes to the modulation of the signal. In this respect it is worth considering whether the presence of additional morpho-syntactic information may lead to changes in encoding. One hypothesis that emerges is that languages with uncommon phonotactic patterns are also languages whose morphologies are characterized by multiple consonantal prefixes attached to a root. If the ranking of faster
vs. accurate information transmission is encoded in patterns of articulatory organization, then in such languages longer timing lags would be predicted to be preferred, because they minimize loss of information.

The vocalic transitions, or excrescent vowels, that sometimes accompany these longer lags, provide peaks of energy, and are potentially phonologizable properties of the signal. The resulting signal still has a clear acoustic modulation, with sonority peaks and troughs. Longer lags can therefore be exploited in those languages that allow sonority reversals or plateaus in word onsets.

5. Discussion and conclusion

The question raised in this chapter concerns both the segmentation of the speech stream into syllables and the internal organization of a syllable. The brief comparison of timing coordination in three languages suggests the following interpretation of a general principle of syllable organization: what goes into a syllable onset is whatever can be maximally co-produced without minimizing signal modulation and without losing intelligibility. This can be achieved by an ordering of constrictions as the one captured by the sonority hierarchy, or by other types of ordering. In the latter case, longer lags maintain the alternations in the modulation of the signal.

Syllable-sized portions of the acoustic signal are known to play an important part in speech perception and comprehension. The studies of neural envelope tracking mentioned in the introduction have shown that the auditory cortical representations of the speech signal are sensitive to syllable-sized windows. An important piece of evidence comes from experimental studies of speech intelligibility in which the temporal speech envelope was manipulated. An experiment carried out by Ghitza & Greenberg (2009) involves time-compressed speech. The authors showed that the reduced intelligibility induced by a high compression factor can be overturned if the speech is “repackaged” by inserting silent gaps in-between successive intervals of compressed speech. Intelligibility increased when silent gaps of 20-120 ms were inserted, then decreased again for longer silent intervals (160 ms). Intelligibility was optimal when the distribution of the acoustic information in the time-compressed signal matched (was aligned with) that in the original, uncompressed signal.

In the authors’ interpretation, the resulting U-shaped performance implies that the auditory channel capacity is determined by the syllable-sized theta-frequency of neural rhythms, and that the appropriate unit to express speech information transfer rate is theta-syllables/s. MEG studies (Doelling et al., 2014) support this interpretation, showing that stimulus intelligibility is affected by the presence or absence of temporal fluctuations that occur at the syllabic (theta) rate. The role of temporal speech envelope information and entrainment to the input speech rhythm is supported by several experiments. It is a necessary component of speech comprehension. The importance of parsing the signal in syllable-sized chunks is reinforced, at least for recognizing syllables in speech without context.

If the temporal organization of the speech stream is indeed driven by the theta/syllable-rate, then it can be hypothesized that temporal properties of spoken language concur with this general property. It has been proposed by Pouplier and Beňuš (2011) that longer lags provide a favorable environment for syllabic consonants to emerge. I would add that they provide, more generally, an environment for energy peaks
to emerge. What counts as an energy peak further depends on details of the language-specific timing patterns, as well as on the lexical and morpho-syntactic makeup of a linguistic system. For example, the long lags in Georgian and Tashlhiyt may be related to the specific morphologies of these languages. Typologically they are quite different, but both morphological systems are rich in consonantal affixes stacked up, primarily preceding the root. Many of the ideas expressed here clearly await further verification. As a preliminary conclusion, successful encoding, transmission, and decoding of linguistic information crucially depend on intergestural timing patterns.

References


Bird, S. and Czajkowska-Higgins, E. (this volume)


Jespersen, O. (1897-1899) *Fonetik en systematisk fremstilling af laeren om sproglyd*. Copenhagen: Det Schuboteske Forlag


