Analytical methods for assessing syllable structure in experimental data

Author(s) redacted
Institution(s) redacted

Abstract

We develop analytical methods in better understanding the relation between qualitative syllable parses and their quantitative consequences. The statistics of a symbolic organization corresponding to a syllable parse can be expressed in terms of continuous phonetic parameters. Analysis of the link between symbolic phonological form and the continuous phonetics thus becomes possible. Pursuing such analysis, we illustrate the predictions of specific syllabic organizations and derive a number of previously experimentally observed results.

1 Computational models

Gafos (2002) and Shaw et al. (2009) employ computational modeling methods for establishing an explicit link between qualitative phonological organization and continuous experimental data. Specifically, Shaw et al. (2009) develop a modeling paradigm which begins with a hypothesized syllable parse of a segmental string to generate temporal structure using a probabilistic version of temporal alignment constraints developed in Gafos (2002). The simulated data can then be compared to experimental data acquired through Electromagnetic Articulometry so that the performance of the hypothesized parse can be rigorously evaluated.

A schematic of the modeling paradigm is in Figure 1. Each syllable parse can be mapped to a coordination topology (Gafos 2002: 316), reflecting the temporal relations underlying the segmental sequence. Two contrasting coordination topologies corresponding to a simplex onset parse (H₁) and a complex onset parse (H₂) of a segmental substring CCVX are shown in Figure 1. Mnemonics are ‘C’ for any consonant, ‘V’ for any vowel, and ‘X’ for any string over the {C,V} alphabet. These topologies specify timing relations between consonants and vowels, indicated by lines between the segments so related. Formally, syllabic organization is diagnosed by estimating the proportion of variance in intervals (some spanning across and some spanning within the phonological constituents of interest) that can be explained by the hypothesized coordination topology. Different topologies act as mutually exclusive independent variables, e.g. in the example of Figure 1, for any given CCV sequence, the parse in which both consonants are part of the onset (as per the English syllable structure) is pitted against the parse in which only the prevocalic C is included in a syllable with the V (as per e.g. Arabic syllable structure). The task is to identify the topology accounting for the most variability in the data. For example, it is expected that for a CCV string in a language that does not admit complex onsets, the simplex onset topology explains more variability than the complex onset topology.

From a coordination topology, the model generates temporal structure that reflects this topology. Given a set of word types, e.g. CVX, CCVX, CC- CVX, the model generates articulatory landmarks defining the plateau of each consonant in relation...
to its adjacent consonants and to the vowel. The plateau of a consonant is defined as the interval demarcated by two landmarks, TARGET and RELEASE. The TARGET corresponds to the timestamp of the achievement of the consonant’s constriction, e.g. the timepoint where the tongue makes contact with the alveolar ridge during the formation of a [t]. RELEASE corresponds to the timestamp of the beginning of the movement away from that constriction.

To make these ideas concrete, we enumerate the steps for unfolding the temporal structure corresponding to a sequence of two consonants C1C2: generate the plateau of C1, a time interval delimited by the TARGET and RELEASE landmarks of C1 – the duration of this interval can be set to the mean plateau duration in the experimental data; generate the plateau of the C2, again an interval delimited by the consonant’s TARGET and RELEASE landmarks; and finally align the RELEASE landmark of C1 with the TARGET of C2. This ensures that there is no acoustic release within a cluster of two consonants, as in [bd] of English [rabd] ‘robbed’, not [bʰd].

The latter acoustic output would correspond to the acoustic release within a cluster of two consonants, such as in [bd] of English [rabd] ‘robbed’, not [bʰd].

Limiting attention to two segments timed in the way described above, the actual output timing relation between these segments is the product of two components. One is the deterministic “central plan” component expressed by alignment of landmark statements of the sort described above. The other is a stochastic “motor” component. At a first approximation, this latter component can be thought of as consisting of just noise randomly shifting or realigning the two events specified in the central plan away from each other. However, the motor component can also be tailored to specific clusters allowing us to more precisely encode timing relations that differ depending on the identity of the segments so related while keeping the central plan uniform across clusters.

Syllabic structure enters crucially in the statement of these timing relations. For example, in a CCV sequence, the hypothesis that it is syllabified as C.CV, with a simplex onset, dictates that the vowel START is timed to the prevocalic consonant (only). The hypothesis that it is syllabified as C.CV, with a complex onset, dictates that the vowel START is timed to the center of the entire prevocalic consonantal cluster (Gafos 2002: 316-22 shows how these outcomes can be derived from interaction of competing coordination relations). Syllabic structure then determines the timestamp of the START landmark for the vowel. From this timestamp, we derive the timestamp of the anchor, a timepoint which is found all the way at the other end of the vowel, by adding a term corresponding to the vowel’s duration equal to the mean vowel duration in the experimental data. Based on this new timestamp, a set of anchor distributions is generated with the same mean but differing standard deviations. For example, Shaw et al. (2009) use a set of twenty anchors in which the standard deviation of the anchor increases from 0 ms in anchor 1 to 95 ms in anchor 20 in steps of 5 ms. Anchor variability is used as a stand-in for any source of variability in the intervals spanning within and across the hypothesized syllabic constituents. Such sources include rate, lexical statistics and measurement error. These and other yet unknown factors introduce noise in our experimental data. For instance, rate of speaking may vary from one stimulus production to another (Max & Caruso 1997) and lexical frequency and density (Munson 2001, Munson & Solomon 2004) may affect variability in articulation. Such variability is injected in the simulated data by systematically changing the standard deviation of the anchor distribution. In this way we ensure that variability is uniformly present across all intervals of interest: the intervals quantified in our data analysis are all right-delimited by a shared anchor, e.g. left edge to anchor, center to anchor and right edge to anchor.

To sum up the central idea, the task of evaluating syllable parses with experimental data has been formulated here as the task of fitting abstract coordination topologies to the experimental data (see Figure 1). This fitting can be expressed using two types of parameters, coordination topologies and anchor variability. In the study of biological coordination and complex systems more generally, these two parameters correspond respectively to the so-called essential and non-essential parameters describing the behavior of complex systems (Kugler et al. 1980: 13). Essential parameters specify the qualitative form of the system under study. For us, this corresponds to the symbolic parse of the phonological string. The fundamental hypothesis entailed in positing an abstract phonological organization isomorphic to syllable structure is that a syllable parse is a macroscopic organization uniform across a variegated set of segmental identities, lexical statistics and rate conditions, e.g. ‘plea’,
‘tree’, ‘glee’ are single syllables independent of speech rate, frequency or phonotactic probability (see Catford 1977: 13 on ‘phonological form’). All of these factors have left imprints on the articulatory patterns registered in experimental data. Crucially, we do not know and it may not be possible to predict for any given stimulus how each such factor or combination of factors has affected the intervals quantified. Taken together, then, these and other yet unknown factors have introduced noise in the intervals that will be measured. Therefore, in formulating the modeling problem of diagnosing syllable structure in experimental data, variability is one of the parameters manipulated in the fitting process.

2 Validity of phonetic indices

An important question for future research on uncovering the mental/syllabic organization of phonological form is: how reliably do stability measures of temporal organization, extracted from the inherently variable and continuous phonetic signal, reflect syllable structure? This problem has not been studied before. Through our approach, we can develop tools for studying the relation between qualitative phonological structure and experimental data. Using these tools we can expose the full range of the relation between a hypothesized symbolic syllabic organization and continuous stability metrics of that organization.

Past analyses of syllable structure in articulation examine the variance of structurally relevant intervals, some spanning across and some spanning within the phonological constituents of interest, extracted from strings with some hypothesized syllabic structure, e.g. the left edge to anchor, center to anchor and right edge to anchor intervals (Shaw, Gafos, Hoole, Zeroual, 2009, Byrd 1995, Honorof & Browman 1995, Browman & Goldstein 1988). In complex onset languages (e.g. English) it is expected that the center to anchor interval is more stable than the right edge to anchor interval (Figure 2, right). In simplex onset languages, the opposite pattern is expected (Figure 2, left).

This latter pattern is the predominant one in e.g. Arabic data and goes along with theoretical evidence supporting the simplex onset hypothesis for this language (Dell & Elmedlaoui 2002). But exceptions to this pattern can indeed be found. One such exception is shown in Table 1 from one Arabic speaker. When the data are quantified using the $C_{\text{max}}$ anchor, the right edge to anchor interval showed lower relative standard deviation (RSD) than the center to anchor interval. This is the canonical result for Arabic. However, when the data are quantified using a different anchor, the $V^\text{end}$ anchor, the inverse stability pattern is found: the center to anchor shows lower RSD than the right edge to anchor. This latter pattern is the same as that seen in English above and would seem to support the complex onset hypothesis. In sum, in one subset of measurements it is the right edge to anchor interval that is most stable, but in a different subset it is the center to anchor interval that is most stable. One response to such inconsistencies would be to conclude that temporal stability indices are unreliable in diagnosing syllabic organization (“everything goes”) or even that syllable structure does not and need not, as Kahn (1976: 16-17) asserted, have consistent phonetic indices. A more constructive and ambitious approach is to appreciate that the relation between abstract phonological organization and these indices may not be straightforward, and undertake the task of developing tools enabling the systematic study of this relation. More generally, the problem met here is a specific instance of a
larger problem in cognitive science, namely evaluating qualitative theoretical constructs with variable experimental data.

The modeling approach developed in Shaw et al. (2009) can be employed to develop analytical tools in making sense of such seemingly inconsistent results (Table 1) by illuminating how stability-based indices are modulated under variation in different parameters. Returning to our example in Table 2, we can use the model embodying the Simplex Onset Hypothesis to study the effect of variability on indices of temporal stability. Via the model we generate simulated data. Our focus now is on how the patterning of temporal stability indices changes as we change variability in the data which, it will be recalled, can be done in our simulations by systematically changing anchor variability. Results are in Figure 3, showing the relative standard deviation or RSD, y-axis, of three intervals (left edge to anchor, center to anchor, right edge to anchor) by anchors of increasing variability, x-axis.

It can be seen that at low levels of anchor variability, anchors 1-6, the right edge to anchor interval has the lowest RSD. This is the expected phonetic reflex of the Simplex Onset Hypothesis (see, again, Figure 2, left). But as anchor variability increases, the right edge to anchor RSD increases at a slower rate than the center to anchor RSD. A tipping point can thus be seen after which the center to anchor interval emerges as having better stability (lower RSD) than the right edge to anchor interval. The stability pattern has changed. Specifically, it has changed to an English-like pattern expected for languages instantiating the Complex Onset Hypothesis, even though the model generating the data here embodies the Simplex Onset Hypothesis. The mapping between intended syllable structure and stability patterns is not one-to-one, but we can also go further because modeling enables analysis of this mapping: though both stability patterns (right edge to anchor more/less stable than center to anchor) are consistent with the Simplex Onset Hypothesis, the model identifies stringent conditions for their occurrence. Given a corpus and two sets of intervals delimited by different anchors extracted from this corpus, the model embodying the Simplex Onset Hypothesis predicts the following implicational relationship. If one set of intervals shows center to anchor stability and the other shows right edge to anchor stability, then the former set of intervals must have higher overall variability than the later. Further, the opposite relationship is precluded; it is not the case that “everything goes”. Such predictions allow us to better diagnose syllable structure in the phonetic record. Returning to Table 1, it can be seen that across all intervals, the SD and RSD values for intervals right-delimited by the V^{\text{end}} anchor are higher than for the corresponding intervals right-delimited by the C^{\text{max}} anchor. It is only under such conditions of higher variability where the center to anchor interval may be found to show a stability advantage over the right edge to anchor interval. In short, this data pattern is predicted by the Simplex Onset Hypothesis. A similar fit between the simulations and the experimental data can be established for patterns involving the left edge to anchor interval.

3 Analysis of the simplex onset model

3.1 Linking symbolic form to continuous data

We aim at developing a formal substrate for making explicit the relation between qualitative form and its continuous experimental manifestations. Stochastic models of phonological (syllabic) organization can be employed as analytical tools for evaluating the relation between qualitative syllable parses and the entire range of their quantitative consequences. Since the models are fully explicit, a rigorous mathematical analysis of the models’ predictions can be carried out.

As explained in the previous section, in our modeling paradigm the temporal consequences of syl-
lable parses are stochastic quantities whose statistics can be computed and plotted for further analysis. Figure 3 did this for the RSD values of the three intervals plotted there. Analytical expressions (equations) can also be computed for these. Letting \( x \) represent the center to anchor interval, its RSD is \( \text{RSD}(x) = \sqrt{\text{Var}(x)/E(x)} \), where \( \text{Var}(x) \) and \( E(x) \) are respectively the variance and mean (expected value) of that interval. The right-hand side of this equation can be expanded, given any corpus of data. As an example, given a corpus of CVC, CCVC and CCCVC words, we may derive expressions for the mean and variance of the center to anchor interval or any other intervals of interest. The mean is given by the expression \( E(x) = E(x|\text{cve}) \ast P(\text{cve}) + E(x|ccve) \ast P(\text{ccve}) + E(x|\text{cccv}e) \ast P(\text{cccv}e) \), where the terms \( E(x|\text{cve}) \) and \( P(\text{cve}) \) are respectively the conditional mean and probability of CVC words in the corpus (and so on for the rest of the terms). The conditional means in the equation above can be easily expressed in terms of the model parameters: consonantal plateau durations, inter-plateau durations, vowel durations, and their variances. These parameters exist at the level of the continuous speech signal. Hence, the statistics of a symbolic organization corresponding to a syllable parse can be expressed in terms of these lower-level parameters.

Once RSD expressions have been derived for the center to anchor and the right edge to anchor intervals, the next step is to identify their patterning. Equivalently, we can identify the conditions under which the center to anchor interval is more or less stable than the right edge to anchor interval. This is done by solving for inequalities such as \( \text{RSD}(x) > \text{RSD}(y) \), where \( x \) is the variable representing the center to anchor interval and \( y \) the right edge to anchor interval. The conditions under which these inequalities hold are expressed in terms of the basic parameters of the model. In turn, because the values for these parameters can be estimated from the corpus, given a corpus and its hypothesized syllabic organization, we can predict the patterns of stability that should be met in the experimental data from that corpus. This allows a rigorous evaluation and analysis of the relation between qualitative organization and experimental data specific to a speaker or a population of speakers. In the next section, we offer a concrete instantiation of the analytical approach sketched above.

3.2 Analysis

The random variables (and their variances) relevant to our analysis are: (a) consonant plateau duration, \( \pi \) (variance \( \tilde{\pi} \)), (b) the temporal distance between the offset of one consonantal plateau and the onset of another, \( \delta \) (variance \( \tilde{\delta} \)), and (c) the right-edge-to-anchor interval, \( \rho \) (variance \( \tilde{\rho} \)).

We define gestural landmarks as random variables expressed in terms of these values. We begin by fixing some arbitrary temporal value for the right edge, \( \tau \), and identify the temporal alignment of other landmarks with respect to it.1 Our mnemonics: ‘T’ denotes target or achievement of closure, ‘R’ denotes release of closure, and ‘A’ denotes anchor point.

\[
\begin{align*}
E(R_3) & := \tau \\
E(T_3) & := \tau - \pi \\
E(R_2) & := \tau - \pi - \delta \\
E(T_2) & := \tau - 2\pi - \delta \\
E(R_1) & := \tau - 2\pi - 2\delta \\
E(T_1) & := \tau - 3\pi - 2\delta \\
E(A) & := \tau + \rho
\end{align*}
\]

Using two familiar theorems from probability and stochastic processes—viz. (1) the variance \( \tilde{\zeta} \) of a random variable \( \zeta = \alpha + \beta \) (with \( \alpha, \beta \) independent random variables) is given by \( \tilde{\alpha} + \tilde{\beta} \), (2) for some coefficient \( n \) and random variable \( \zeta \) with variance \( \tilde{\zeta} = n^2 \tilde{\zeta} \)—the variance of each landmark is:

\[
\begin{align*}
\tilde{R}_3 & = \tilde{\tau} = 0 \\
\tilde{T}_3 & = \tilde{\pi} \\
\tilde{R}_2 & = \tilde{\pi} + \tilde{\delta} \\
\tilde{T}_2 & = 4\tilde{\pi} + \tilde{\delta} \\
\tilde{R}_1 & = 4\tilde{\pi} + 4\tilde{\delta} \\
\tilde{T}_1 & = 9\tilde{\pi} + 4\tilde{\delta} \\
\tilde{A} & = \tilde{\rho}
\end{align*}
\]

We now define conditional C-center metrics (CCMs) and their variances. For any consonant, its C-center is defined as the midpoint between the target and the release landmarks of that consonant. The C-center (CC) of a cluster of consonants is defined as the mean of the individual C-centers of all the consonants in the cluster, and the CCM as the temporal difference between CC and the anchor point. We may then move from these expressions

\[ \text{Figure 4: Defining prevocalic consonantal gestural landmarks in terms of mean peak duration (}\tau\text{), mean inter-plateau-interval (}\delta\text{), and a reference point corresponding to the the release of } C_3 (R_3 = \tau) \text{. To define e.g. the expected value of } R_1, E(R_1), \text{we subtract two plateaus and two inter-plateau-intervals from } R_3 = \tau, \text{ giving } E(R_1) = \tau - 2\pi - 2\delta \text{.} \]

1 Here \( \tilde{\tau} = 0 \), but nothing hinges on the values of \( \tau \) or \( \tilde{\tau} \).
to their variances:
\[
E(\text{CCM}[cv]) := A - \frac{T_3 + R_3}{2}
\]
\[
E(\text{CCM}[ccv]) := A - \frac{T_2 + R_2 + T_3 + R_3}{4}
\]
\[
E(\text{CCM}[cccv]) := A - \frac{T_1 + R_1 + T_2 + R_2 + T_3 + R_3}{6}
\]
\[
\tilde{\text{CCM}}_{cv} = \tilde{A} + \frac{T_3 + R_3}{4}
\]
\[
\text{CCM}_{ccv} = \tilde{A} + \frac{T_2 + R_2 + T_3 + R_3}{16}
\]
\[
\text{CCM}_{cccv} = \tilde{A} + \frac{T_1 + R_1 + T_2 + R_2 + T_3 + R_3}{36}
\]

Simplifying using the above variance values:
\[
\tilde{\text{CCM}}_{cv} = \tilde{\varrho} + \frac{\pi}{4}
\]
\[
\text{CCM}_{ccv} = \tilde{\varrho} + \frac{6\pi + 2\delta}{16}
\]
\[
\text{CCM}_{cccv} = \tilde{\varrho} + \frac{19\pi + 10\delta}{36}
\]

Since we assume that the syllabic parse is that of the simplex onset, the expected right-edge metric (REM) value remains constant across word types. In each case, the REM is given by \( A - R_3 = \tilde{\varrho} \). Conditional REM variances, \( \text{REM}_{\{cv, ccv, cccv\}} \), are in each case \( \tilde{\varrho} \).

We wish to derive expressions for the unconditional expected values and variances of \{REM, CCM\}. The REM case is straightforward since neither the expected value of the REM nor its variance change across word types. It is immediate, then, that \( E(\text{REM}) = \varrho \) and \( \tilde{\text{REM}} = \tilde{\varrho} \). Therefore:
\[
\text{RSD}_{\text{REM}} = \frac{\text{SD}_{\text{REM}}}{E(\text{REM})} = \frac{\sqrt{\tilde{\varrho}}}{\varrho}
\]

For the CCM case, the unconditional expected CCM value is simply the weighted average of each conditional expected CCM value. Assuming that each word type occurs equally frequently, we have (using the following convention for notational perspicuity: \( \text{CCM}1 = \text{CCM}cv, \text{CCM}2 = \text{CCM}ccv, \text{etc.} \)):
\[
E(\text{CCM}) = \frac{1}{3} \sum_{n=1}^{3} E(\text{CCM}[n]) = \pi + \frac{\delta}{2} + \varrho
\]

To calculate the unconditional (i.e. combined) variance of the CCM, we use the following expression for combined variance, again assuming equal frequency of word types:
\[
\tilde{\text{CCM}} = \frac{1}{3} \sum_{n=1}^{3} [\text{CCM}[n] + (E(\text{CCM}[n]) - E(\text{CCM}))^2
\]

\[
= \frac{83\pi + 29\delta}{216} + \varrho + \frac{\pi^2}{6} + \frac{\pi\delta}{3} + \frac{\delta^2}{6}
\]

We now have all we need to express \( \text{RSD}_{\text{CCM}} \):
\[
\frac{\text{SD}_{\text{CCM}}}{E(\text{CCM})} = \sqrt{\frac{83\pi + 29\delta}{216}} + \tilde{\varrho} + \frac{\pi^2}{6} + \frac{\pi\delta}{3} + \frac{\delta^2}{6}
\]

To calculate the anchor variance \( \varrho \) at which the simplex onset parse starts to exhibit greater CCM stability than REM stability—the crossover point—we set \( \text{RSD}_{\text{CCM}} < \text{RSD}_{\text{REM}} \) and solve for \( \tilde{\varrho} \):
\[
\sqrt{\frac{83\pi + 29\delta}{216}} + \tilde{\varrho} + \frac{\pi^2}{6} + \frac{\pi\delta}{3} + \frac{\delta^2}{6} < \sqrt{\tilde{\varrho}} + \varrho + \pi + \frac{\delta}{2} + \varrho
\]

The right-hand side of the inequality gives the point of anchor variance beyond which the phonetic manifestation of the invariant symbolic parse (simplex onset) shifts from REM to CCM stability.

### 3.3 Comparison with stochastic simulation

Shaw et al. (2009) discuss a stochastic simulation of a similarly defined simplex onset model of gestural coordination. The simulation takes as inputs values for \( \pi, \delta, \varrho, \tilde{\pi}, \) and \( \tilde{\delta} \), as well as a “starting value” for \( \tilde{\varrho} \), stochastically generates sets of \( n \) tokens based on these values, and calculates statistics over the simulated lexicon. The simulation gradually increases the value of the anchor variance \( \varrho \). When \( \varrho \) passes a certain critical value, \( \text{RSD}_{\text{CCM}} < \text{RSD}_{\text{REM}} \)—i.e. CCM becomes the more stable of the two.

Following the mathematical instantiation of the simplex onset model in the previous section, we can now analytically predict the simulation’s behavior under arbitrary combinations of input values. For example, assume \( \pi = 30\text{ms}, \delta = 40\text{ms}, \varrho = 267\text{ms}, \) and \( \tilde{\delta} = 400\text{ms} \). Given these values, our analytical solution determines the crossover point (the point at which there is a shift from REM to CCM stability) to be \( \varrho \approx 2500.2\text{ms}^2 \).

In terms of standard deviation (\( \sigma_{\tilde{\varrho}} \)), the crossover point is then \( \approx 50.002\text{ms} \).

This prediction is borne out by the simulation. When \( \pi = 30\text{ms}, \delta = 40\text{ms}, \varrho = 267\text{ms}, \) and \( \tilde{\pi} = \tilde{\delta} = 400\text{ms} \), with 3000 averaged simulations of about ten thousand tokens, the observed crossover point is \( \sigma_{\tilde{\varrho}} \approx 50.017\text{ms}, \) or \( \tilde{\varrho} \approx 2501.7\text{ms}^2 \). See Figure 5.

We offer an additional example. Set \( \pi = 20\text{ms}, \delta = 10\text{ms}, \varrho = 234.8\text{ms}, \) and \( \tilde{\pi} = \tilde{\delta} = 100\text{ms} \). The
analytical solution predicts that CCM will overtake REM as the more stable interval beyond $\sigma_\theta \approx 30\text{ms}$ ($\tilde{\sigma} \approx 899.984\text{ms}^2$). When the simulation is fed these parameters, 3000 averaged simulations of about ten thousand tokens yields a crossover point of $\sigma_e \approx 30.008\text{ms}$ ($\tilde{\sigma} \approx 900.48\text{ms}^2$). See Figure 6.

### 3.4 Other sources of variability

The simulations reported in Figures 5, 6 modulate right-edge-to-anchor variance in order to progressively introduce variability into a simulated lexicon. Another way to introduce variability is to construct lexica composed of tokens from a diverse range of contexts, i.e. without controlling for segmental identity, speech rate, or speaker.

We analyzed speech movement data collected with Electromagnetic Articulometry (Perkell et al. 1992, Hoole et al. 2003) from three native speakers of Moroccan Arabic (MA). Each speaker produced between 13 and 15 tokens of (k)sbulha, (k)skulha, and (m)skulha, for a total of 386 tokens. Using the MATLAB–based software MVIEW, we extracted landmark timestamps corresponding to constriction targets and releases of prevocalic consonants. Additionally, we extracted timestamps corresponding to the achievement of maximum degree of constriction for the postvocalic consonant. This landmark was treated as the anchor point relative to which our REMs and CCMs were calculated. As in the analytical treatment above, REM was defined as the temporal distance from the release of the immediately prevocalic consonant to the anchor point, whereas the CCM was defined as the temporal distance between the anchor and a derived landmark corresponding to the mean of the midpoints of each prevocalic consonant (the midpoints for each consonant in turn corresponding to the mean of that consonant’s target and release timestamps).

REM and CCM RSD statistics were calculated for seven different lexica $S_1$–$S_7$, such that for any lexica $S_n$ and $S_{n+1}$, $S_n \subset S_{n+1}$. The precise makeups of each lexicon are given in Table 2. Each lexical update incorporates a new set of CVC, CCVC, and CCCVC speaker productions.

As the lexicon grows, new tokens of words with diverse segmental makeups and from different speakers are incorporated into the sets over which stability indices are calculated. Unsurprisingly, then, the variance of both REM and CCM increase as a function of lexical size. However, this increase in variability has a disproportionate effect on the RSD of the REM, causing a stability reversal between $S_3$ and $S_4$, charted in Figure 7, with RSD figures in Table 3.

It is revealing that the quantitative pattern of stability indices changes or tends to an English-like pattern expected for languages instantiating complex onsets. This behavior provides a stochastic temporal basis for the Onset Maximization Principle (Clements & Keyser 1983, Blevins 1995) – stating that languages prefer to syllabify multiple consonants with their following vowel rather than

<table>
<thead>
<tr>
<th>Lexicon</th>
<th>Tokens Produced</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1$</td>
<td>bulha + flanB + kflanB</td>
</tr>
<tr>
<td>$S_2$</td>
<td>k bulhaB + sbulhaB + ms kulhaA</td>
</tr>
<tr>
<td>$S_3$</td>
<td>k lanB + sbulhaA + kskulhaA</td>
</tr>
<tr>
<td>$S_4$</td>
<td>k lanA + flanC + kskulhaB</td>
</tr>
<tr>
<td>$S_5$</td>
<td>k bulhaA + sbulhaC + kflanA</td>
</tr>
<tr>
<td>$S_6$</td>
<td>k bulhaC + skulhaA + ms kulhaB</td>
</tr>
<tr>
<td>$S_7$</td>
<td>full paradigms for $A$, $B$, $C$</td>
</tr>
</tbody>
</table>

Table 2: Constructed lexica $S_n \subset S_{n+1}$. Notational conventions: ‘word’ denotes the set consisting of all of speaker $X$’s productions of word, and ‘+’ denotes set union.
splitting the consonants across different syllables (e.g. V.CCV preferred over VC.CV). Thus, when syllabic organization is assessed over a broad range of contexts such as across many words differing in the phonic identity of their constituent sounds, across many speakers and speech rates, the temporal stability patterns tend to the same qualitative state characteristic of complex onsets.

Through our analytical methods, we can thus achieve a system-level, quantitative understanding of how categorical organizational principles of linguistic systems may be supported by or emerge from statistical properties of the lower-levels in which linguistic form is conveyed. Specifically, we can imagine a phonological change scenario in which a learner exposed to outputs from a simplex onset language makes use of only a subset of the possible temporal measures, and thus hypothesizes that the ambient language has complex onsets. This way of approaching the ‘basis’ of higher-level organizational principles like Onset Maximization can be paralleled to Ohala’s paradigm for sources of sound change (Ohala 1981), but in the stochastic realm of stability samples over a lexicon. Crucial to this analytical understanding and a novelty of our formal approach is our stance on the role of variability in behavioral data. Because our approach explicitly takes into account variability, it harnesses this factor as a useful source of information about mental organization. Instead of ignoring it or treating it as a nuisance, variability in our approach is a tool for elucidating the relation between symbolic form and its complex behavioral instantiations.

4 Conclusions

Using methods from probability theory and stochastic processes, we derive analytic expressions for the relation between syllabic organization and phonetic quantitative indices of that organization. We proceed by building appropriately parameterized models or replicas of phonological (syllabic) parses. Studying parameter changes in these syllable replicas enables us to make explicit the phonetic predictions of the corresponding symbolic phonological organizations. These predictions are stochastic quantities whose statistics can be computed and plotted for further analysis. This analysis shows that the relation between syllabic organization and phonetic indices shows tipping points after which the phonetic indices change to values resembling the canonical manifestation of a distinct syllabic organization. We can thus derive classical effects observed in the experimental literature on the phonetic consequences of different hypothesized syllable structures. Further, we can expose the range of validity of these effects and analytically predict their breakdown.

References


