ARTICLE IN PRESS

Journal of Phonetics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

# Stochastic time analysis of syllable-referential intervals and simplex onsets

Adamantios I. Gafos [a,b,]*, Simon Charlow [c], Jason A. Shaw [d,e], Philip Hoole [f]

[a] Universität Potsdam, Germany
[b] Haskins Laboratories, United States
[c] New York University, United States
[d] MARCS Institute, Australia
[e] School of Humanities and Communication Arts, University of Western Sydney, Australia
[f] Ludwig-Maximilians-Universität München, Germany

## ABSTRACT

We pursue an analysis of the relation between qualitative syllable parses and their quantitative phonetic consequences. To do this, we express the statistics of a symbolic organization corresponding to a syllable parse in terms of continuous phonetic parameters which quantify the timing of the consonants and vowels that make up syllables: consonantal plateau durations, vowel durations, and their variances. These parameters can be estimated from continuous phonetic data. This enables analysis of the link between symbolic phonological form and the continuous phonetics in which this form is manifest. Pursuing such an analysis, we illustrate the predictions of the syllabic organization corresponding to simplex onsets and derive a number of previously experimentally observed and simulation results. Specifically, we derive not only the canonical phonetic manifestations of simplex onsets but also the result that, under certain conditions we make precise, the phonetic indices of the simplex onset organization change to a range of values characteristic of the complex onset organization. Finally, we explore the behavior of phonetic indices for syllabic organization over progressively increasing sizes of lexical samples, thereby concomitantly diversifying the phonetic context over which these indices are taken.

## 1. Introduction

We offer a formal characterization of the temporal phonetic indices used in experimental work to assess syllabic organization. In this characterization, the temporal consequences of syllable parses are expressed as stochastic quantities whose statistics can be subject to analysis. These quantities are expressed in terms of continuous phonetic parameters referring to the timing of the consonants and vowels that make up syllables. These parameters can be estimated from continuous phonetic data. The resulting characterization serves as a link between qualitative syllable parses and their quantitative consequences. A formal characterization of this link enables a better understanding of the behavior of temporal stability indices under conditions which may affect timing. Specifically, we deduce closed-form expressions specifying, for a set of phonetic parameters, the expected values and variances of syllable-referential intervals used to infer phonological organization. The analysis is illustrated by deriving a number of previously observed experimental and simulation results. Finally, we explore how sampling over a lexicon influences the collective phonetics as expressed via stability metrics. Specifically, using articulatory movement data from Moroccan Arabic, we show that progressive increases in lexical size and concomitant diversification in phonetic contexts lead to a shift of the phonetic indices from values characteristic of the simplex onset organization to values characteristic of the complex onset organization.

## 2. Background

Linguistic and specifically syllabic structure shapes the continuous low-level temporal organization of articulatory movements during speech (e.g. Browman & Goldstein, 1988; Byrd, 1995; Byrd, Tobin, Bresch, & Narayanan, 2009; Honorof & Browman, 1995; Krakow, 1999; Sproat & Fujimura, 1993). Syllabic structure is language-specific. Whereas in English sequences such as [kru] 'crew' or [gli] 'glee' are parsed into a single syllable with a

---

* Corresponding author at: University of Potsdam, Linguistics and Center of Excellence in Cognitive Science, Karl-Liebknecht-Str. 24–25, 14476 Potsdam, Germany. Tel.: +49 30202373610; fax: +49 3319772095.

*E-mail address:* adamantios.gafos@uni-potsdam.de (A.I. Gafos).

0095-4470/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2013.11.007

Please cite this article as: Gafos, A. I., et al. Stochastic time analysis of syllable-referential intervals and simplex onsets. *Journal of Phonetics* (2014), http://dx.doi.org/10.1016/j.wocn.2013.11.007

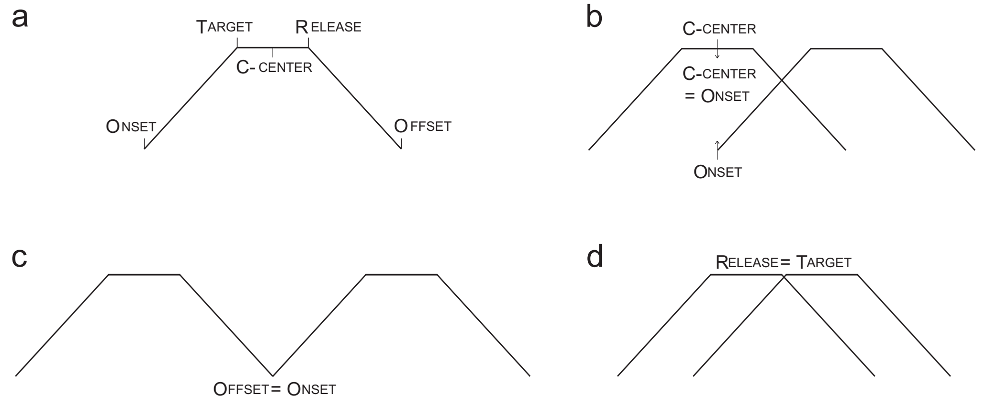

**Fig. 1.** Landmarks in gestural life and examples of different coordination relations. (a) Landmarks, (b) overlap, (c) no overlap and (d) more overlap.

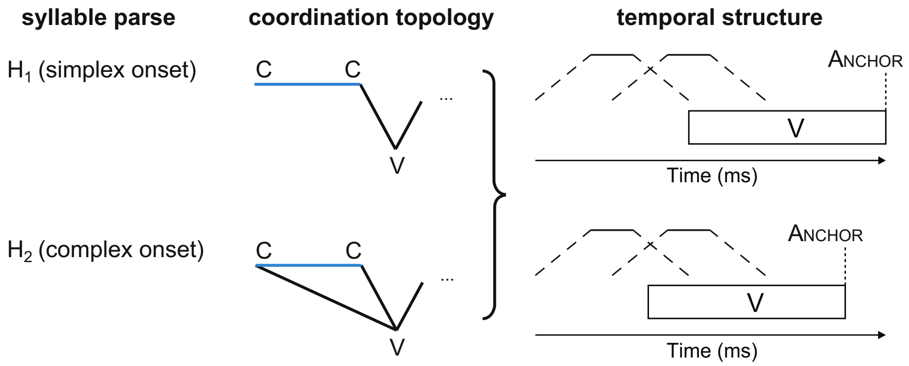

**Fig. 2.** Given any sequence of consonants and vowels, here CCVX, we ask: is the sequence parsed in terms of syllables of the simplex or the complex onset type? To evaluate the two hypotheses, $H_1$ vs. $H_2$, we project coordination topologies from hypothesized syllable parses. The topology on the top/bottom embodies temporal relations of the simplex/complex onset parse. Stochastic time predictions can be derived from these topologies, and their match to phonetic data can be evaluated.

complex two-consonant cluster as its onset (Kahn, 1976), in Moroccan Arabic similar sequences of segments are claimed to be parsed into two syllables, e.g. [k.ra] 'rent', [sk.ru] 'they got drunk', and [g.lih] 'he grilled', where '.' marks syllabic divisions (Dell & Elmedlaoui, 2002). In these forms, the syllables with the vowels [a], [u] and [i] can only include a single consonant as their onset, hence these are syllables with simplex onsets (and the syllables preceding these consist entirely of consonants [k], [sk] and [g]). Because syllabic structure is language-specific, the same or similar segmental sequences are expected to exhibit different temporal organization as a function of the syllabic structure imposed by the different linguistic systems involved. Experimental data illustrating differences in temporal patterning between English and Moroccan Arabic are discussed in the following section. In this background section, we introduce basic notions with the aim of assessing syllabic organization in the temporal dimension of speech.

A fundamental starting point is that the basic units of representation have internal temporal structure (Browman & Goldstein, 1986). For example, the action effecting the formation of the closure for a bilabial stop can be decomposed into a sequence of spatiotemporal events or landmarks, shown in Fig. 1a: Onset, the onset of movement toward the target of the gesture, Target, the point in time at which the gesture achieves its target (i.e. upper and lower lip contact for a bilabial stop), Release, the onset of movement away from the target of the gesture (i.e. upper and lower lip opening of contact for a bilabial stop), C-center, the midpoint of the interval between target and release (known as the gestural plateau), and Offset, the point in time at which active control of the gesture ends.

The landmarks in Fig. 1a delineate the internal temporal structure of gestures. Different gestures are related via gestural coordination relations. A gestural coordination relation is a relation between two gestures stating that a specified landmark (within the temporal structure) of one gesture is synchronous with a specified landmark of another gesture (Gafos, 2002). Coordination relations project corresponding coordination constraints into the grammar. Such constraints are instantiated using the notion of alignment, as developed in the Generalized Alignment theory of McCarthy and Prince (1993). Given two gestures, $G_1$ and $G_2$, the general form of a gestural coordination constraint is Align($G_1$, $\text{Landmark}_1$, $G_2$, $\text{Landmark}_2$). Schemas of phonetic consequences for different coordination constraints are shown in Fig. 1b–d. As can be seen in these examples, different constraints imply distinct amounts of temporal overlap between the coordinated gestures. The relation in Fig. 1b shows an overlapped pattern of coordination, that in Fig. 1c shows a relation of no overlap, and that in Fig. 1d shows a relation of more overlap than in Fig. 1b. Such relations allow us to express the fact that languages show different patterns of coordination in consonant clusters. For instance, when we state in our example of a coordination relation in Fig. 1d that the Release landmark of $C_1$ is aligned with the Target of $C_2$, this is one way of ensuring that there is no acoustic release (known as a close transition) within a cluster of two consonants, as in [bd] of English [rabd] 'robbed', not [bəd]. The latter acoustic output would correspond to the relation in Fig. 1b, a timing pattern which gives rise to an open transition, an interval between the Release of $C_1$ and the Target of $C_2$ when the vocal tract is not constricted, as in e.g. Piro (Anderson, 1974; Gafos, 2002; Matteson & Pike, 1958). For an example of the relation in Fig. 1c, see Gafos, Hoole, Roon, and Zeroual (2010) from homorganic sequences of consonants in Moroccan Arabic.

The problem of assessing syllabic organization in timing data can then be schematized as in Fig. 2. Each syllabic parse can be mapped to a set of coordination relations, reflecting the organization underlying the segmental sequence over which the hypothesized syllabic parse holds.[1] The set of

---

[1] Throughout we refer to syllabification of sequences of segments, but the primitives related by our coordination relations are gestures, not segments. The background necessary to give content to a notion of inter-segmental coordination is in Gafos (2002). Here are the relevant definition and key assumption. Definition: inter-segmental coordination, "Two segments S1, S2 are coordinated with some coordination relation $\lambda$, S1 $\lambda$ S2, if the head gestures of these segments are coordinated as in $\lambda$" (Gafos, 2002, p. 284). It is assumed that the oral gestures of consonants are heads. The assumption that oral gestures drive coordination relations among consonants is also justified therein (Gafos, 2002, p. 296).

coordination relations comprises a coordination topology. Fig. 2 shows two contrasting coordination topologies corresponding to two hypotheses, a simplex onset parse ($H_1$) and a complex onset parse ($H_2$), on how the segmental sequence CCVX is parsed into syllables. Mnemonics are ‘C’ for any consonant, ‘V’ for any vowel, and ‘X’ for any string over the {C,V} alphabet. The coordination relations between consonants and vowels are indicated by lines between the segments so related. In the simplex onset topology, only the immediately prevocalic C enters into a coordination relation with the V; the other C is not parsed in the same syllable as the CV subsequence. In the complex onset topology, in contrast, both prevocalic Cs enter into coordination relations with the V, because both Cs are parsed in the same syllable as the V. These different coordination topologies act as mutually exclusive independent variables, e.g. in the example of Fig. 8, for any given CCV sequence, the parse in which both consonants are part of the onset (as per e.g. 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. the Arabic syllable structure). The task is to identify which coordination topology accounts best for experimental data from a language whose syllabic organization we are interested in assessing. This is done by mapping the abstract coordination topologies to concrete structure in time, as schematized in the right half of Fig. 2 and also exemplified below. See Shaw, Gafos, Hoole, and Zeroual (2009) for an application of this approach to data from Moroccan Arabic, and Shaw and Gafos (2010) for fitting opposing coordination topologies to data from English and Moroccan Arabic.

A set of coordination relations generates temporal structure which reflects the nature and complexity of these relations. To make these ideas concrete, consider the temporal structure corresponding to a CCV sequence, that is, a sequence of two consonants $C_1C_2$ followed by a vowel V. For concreteness, let us assume that the two consonants are coordinated as in our example coordination relation in Fig. 1d above, Align($G_1$, Release, $G_2$, Target). The temporal structure in question consists of a layout of articulatory landmarks defining each consonant in relation to its adjacent consonant and to the vowel. Our example coordination relation refers to landmarks in the consonantal plateaus, Target and Release. We thus proceed as follows: generate the plateau of $C_1$, a time interval delimited by the Target and Release landmarks of $C_1$—the duration of this interval can be set to the mean plateau duration of the experimental dataset being modeled; generate the plateau of the $C_2$, again an interval delimited by the consonant's Target and Release landmarks; and finally align the Release landmark of $C_1$ with the Target of $C_2$. 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 constraints of the sort described above. The other is a stochastic “motor” component. This latter component consists of a stochastic noise kick randomly shifting or dealigning the two events specified in the central plan away from each other.

Syllable structure enters crucially in the coordination relations between the vowel and the consonants. For a CCV sequence, the hypothesis that it is syllabified as C.CV, with a simplex onset, dictates that the vowel Onset is synchronous with the C-center of the prevocalic consonant only; in Fig. 2, in the temporal structure corresponding to the simplex onset parse (top half, under temporal structure), the vowel is shown to begin at the timestamp of that C-center. The hypothesis that it is syllabified as CCV, with a complex onset, dictates that the vowel Onset is synchronous with the C-center of the entire prevocalic cluster, which is defined as the midpoint of the c-centers of each consonant in the cluster; in Fig. 2, in the temporal structure corresponding to the complex onset parse (bottom half, under temporal structure), the vowel is shown to begin earlier than in the previous case, at the timestamp of C-center of the cluster.[2] Syllabic structure then determines the timestamp of the Onset landmark for the vowel with respect to the consonantal cluster. From this timestamp, we can unfold the structure of the vowel (the nucleus of the syllable) and derive the timestamp of what is known as the “anchor”, a time point 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 being modeled. In the stochastic model of syllable structure of Shaw et al. (2009) and Shaw and Gafos (2010), the end of the vowel (the anchor's timestamp) corresponds to a set of anchor distributions with the same mean for that timestamp but differing standard deviations. For example, in Shaw et al. (2009), a set of twenty anchors is used 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 variance is used as a stand-in for any source of variability in the temporal intervals spanning the hypothesized syllabic constituents. Such sources include speech rate, lexical statistics, and measurement error. These and other yet unknown factors introduce noise in 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 of Shaw et al. (2009) by systematically changing the variance of the anchor distribution.

Our description of coordination relations above is based on alignment of gestural landmarks. An alternative is coordination relations expressed in terms of phases. In the gestural model (Browman & Goldstein, 1990), gestures are defined using the dynamics of second-order mass–spring systems. In this model, a gesture is associated with an abstract 360° cycle. A phase corresponds to a point on the cycle of the oscillating body, and is expressed by number of degrees on the cycle (see Hawkins, 1992 for discussion). Coordination relations are expressed in terms of synchronizing phase angles. A mapping can be established between temporal organization as generated in the above discussion and the phasing-based description. Specifically, the spatio-temporal landmarks in the coordination relations above correspond to phase angles, as in Onset is at phase 0°, Target at phase 240°, C-center at phase 265°, Release at phase 290° and so on. For an illuminating discussion of the notion of phase, see Kelso and Tuller (1985). For experimental results concerning inter-gestural phasing, see Nittrouer, Munhall, Kelso, Tuller, and Harris (1988). For applications to syllable structure and modeling results using the phasing model see Nam and Saltzman (2003) and Goldstein, Nam, Saltzman, and Chitoran (2009). For an extension of the notion of phase to phrasal structure and experimental results, see Cummins and Port (1998).

To sum up, in research directed at uncovering experimental evidence for phonological organization, it is worth highlighting the abstractness of the latter concept. Phonological form (e.g. syllables, feet and so on), its grammatical constructs and organizational principles are invariant with respect to the effectors and the specifics of their physical instantiation involved in fleshing out this form. The task of evaluating syllable parses with experimental data is formulated here as the task of fitting abstract coordination topologies to the experimental data. This fitting can be expressed using at least two types of parameters, coordination topologies and some parameter or parameters that either individually or as a coalition perturb those topologies (e.g. for us, anchor variance). 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, Kelso, & Turvey, 1980, p. 13). Essential parameters specify the qualitative form of the system under study. For us, this corresponds to the syllabic parse of some sequence of phonological segments expressed in terms of a coordination topology. Perturbing this qualitative form is pursued by systematically varying some *non-essential* parameter, here, anchor variance. The effects of that perturbation on the stability indices of the qualitative organization make explicit the range of phonetic manifestations of that organization.

---

[2] A first expression of syllable structure in terms of such coordination patterns is Browman and Goldstein (2000), who introduce CV and CC “bonds” of different strengths and some notion of least-squares optimization when bonds are at odds. In a second analysis, Gafos (2002) shows how these temporal outcomes can be derived from the interaction of violable and competing CV and CC coordination constraints in language-specific grammars of gestural coordination. A third analysis has been given in terms of coupled oscillators, where timing relations are stated using the notion of phase, in Nam and Saltzman (2003).

Please cite this article as: Gafos, A. I., et al. Stochastic time analysis of syllable-referential intervals and simplex onsets. *Journal of Phonetics* (2014), http://dx.doi.org/10.1016/j.wocn.2013.11.007

## 3. Stability indices of syllabic organization

This section reviews results from previous experimental work and their proposed interpretations in terms of syllable structure. In furthering our understanding of the relation between experimental data and syllable structure, of particular interest are results in which phonetic indices of syllable structure do not pattern as expected. These cases along with the canonical data patterns both highlight the need and serve as a basis for the analytical take in the following section.

In those cases where speech production data are available, consonant clusters which are parsed as complex onsets seem to exhibit patterns of temporal stability that differ from clusters which are parsed in syllables with simplex onsets. These temporal differences were first quantified in terms of the relative stability of temporal intervals, which refer to hypothesized syllabic constituents of interest (Browman & Goldstein, 1988). To illustrate, Fig. 3 schematizes temporal differences between simplex and complex onsets for words beginning with one (CV), two (CCV), and three consonants (CCCV). For both simplex and complex onsets and for each word type (CV, CCV, CCCV), three temporal intervals — viz. horizontal lines connecting the small circles — are shown left delimited by landmarks in the word-initial consonant(s) and right delimited by a common anchor point (A). The three landmarks on the word-initial consonant(s) are the left edge (the point in time at which the first consonant achieves its target), the c-center (the mean of the c-centers of each consonant in the cluster), and the right edge (the point in time at which the last consonant releases its constriction). When clusters are parsed into simplex syllable onsets (Fig. 3: left), the duration of the right-edge-to-anchor interval is unperturbed by the addition of consonants to the word (an example with kinematic data follows). Consequently, this interval remains stable across CV, CCV and CCCV words. In contrast, when clusters are parsed into complex onsets (Fig. 3: right), the duration of the right-edge-to-anchor interval progressively shrinks as the segmental make up goes from CV to CCV to CCCV, to make room for the addition of an extra consonant to the syllable. Under this temporal alignment schema, the c-center-to-anchor interval remains stable across CV, CCV and CCCV words (or more precisely, the c-center-to-anchor interval remains more stable than both the right-edge-to-anchor and the left-edge-to-anchor interval across CV, CCV and CCCV words).

It is this dynamical insight of uncovering stabilities in syllable-referential intervals by perturbing the phonological string (here, increasing its size from CV to CCV to CCCV) which Browman and Goldstein (1988) first expressed with a stability analysis of English data. Subsequent experimental work has sought to expand the range of empirical investigations of the relation between syllable structure and temporal organization. For example, in some languages, the syllabic parse of consonant clusters depends on the identity of the clusters, e.g. Davis (1990) for Italian. Capitalizing on this case, Hermes, Grice, Mücke, and Niemann (2008) report data from Italian sCV clusters which they demonstrate differ from Italian unambiguous onsets (e.g. CRV clusters where R is a liquid consonant) in ways fully consistent with the schemas of Fig. 3. Pouplier (2012) offers a recent review of experimental studies on a number of languages.

Nevertheless, experimental work has also revealed cases where the evidence from the articulatory data seems hard to interpret and in some cases, to be discussed below, the data even points to conflicting conclusions. Such cases prompt the following question: How reliably do stability measures of temporal organization, extracted from the inherently variable and continuous phonetic signal, reflect syllable structure? It is one of our concerns in this paper to take up this question and experimental data that may speak to it, with the aim of showing how an analytical treatment of the behavior of stability indices may aid in the interpretation of the data. Specifically, we address this question for simplex onsets, by first illustrating in this section the predominance of the canonical temporal alignment schema for simplex onsets, turning to exceptions to this schema and possible interpretations of these exceptions next. In Section 4, we then take up an analysis which aims to make sense of the canonical and the exceptional patterns in the phonetic manifestations of simplex onsets.

We begin in Fig. 4 with an illustration of kinematic data from one speaker of Moroccan Arabic. This data exemplifies the canonical temporal pattern seen in experimental studies (Shaw et al., 2009, 2011) and is in good agreement with theoretical accounts supporting simplex onsets for this language (Dell & Elmedlaoui, 2002, Chapter 8). Fig. 4 shows for each word in the triad *bulha~sbulha~ksbulha* trajectories of the tongue tip (top panel), lower lip (middle panel), and tongue back (bottom panel). Ten tokens of each word (light blue lines) are displayed along with the average trajectory (black lines). The thick vertical lines indicate the right edge of the cluster, corresponding to the release of the immediately prevocalic consonant (vertical black line), and the c-center of the cluster, corresponding to the mean of the midpoints of all prevocalic consonants (vertical gray line). As the number of consonants increases from CV to CCV to CCCV, the position of the vertical grey line (c-center) shifts to the left while the position of the right edge of the consonants remains relatively stable.

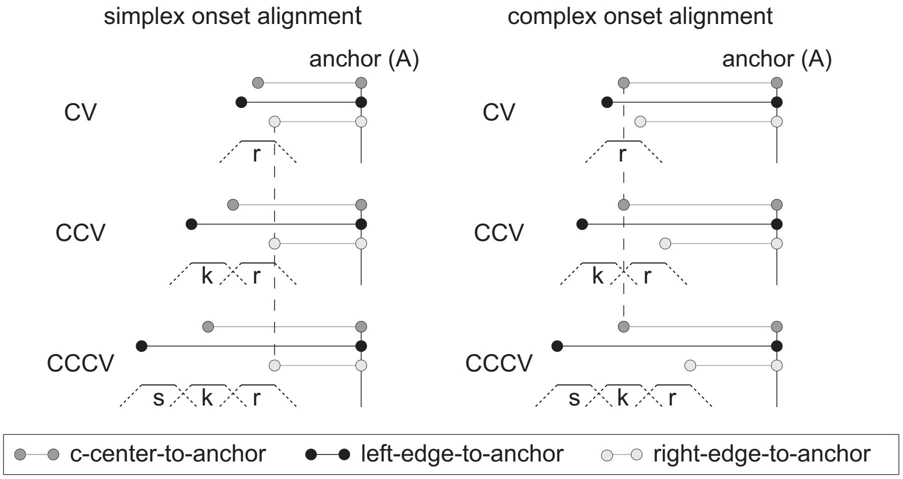

**Fig. 3.** Schematic representation of three intervals, left-edge-to-anchor, c-center-to-anchor and right-edge-to-anchor, delineated by points in an initial single consonant or consonant cluster and a common anchor (A). The alignment schema on the left/right represents experimentally observed temporal manifestations of the simplex/complex onset parse. These schemas have been used as phonetic heuristics for diagnosing syllable structure in experimental data.

Please cite this article as: Gafos, A. I., et al. Stochastic time analysis of syllable-referential intervals and simplex onsets. *Journal of Phonetics* (2014), http://dx.doi.org/10.1016/j.wocn.2013.11.007

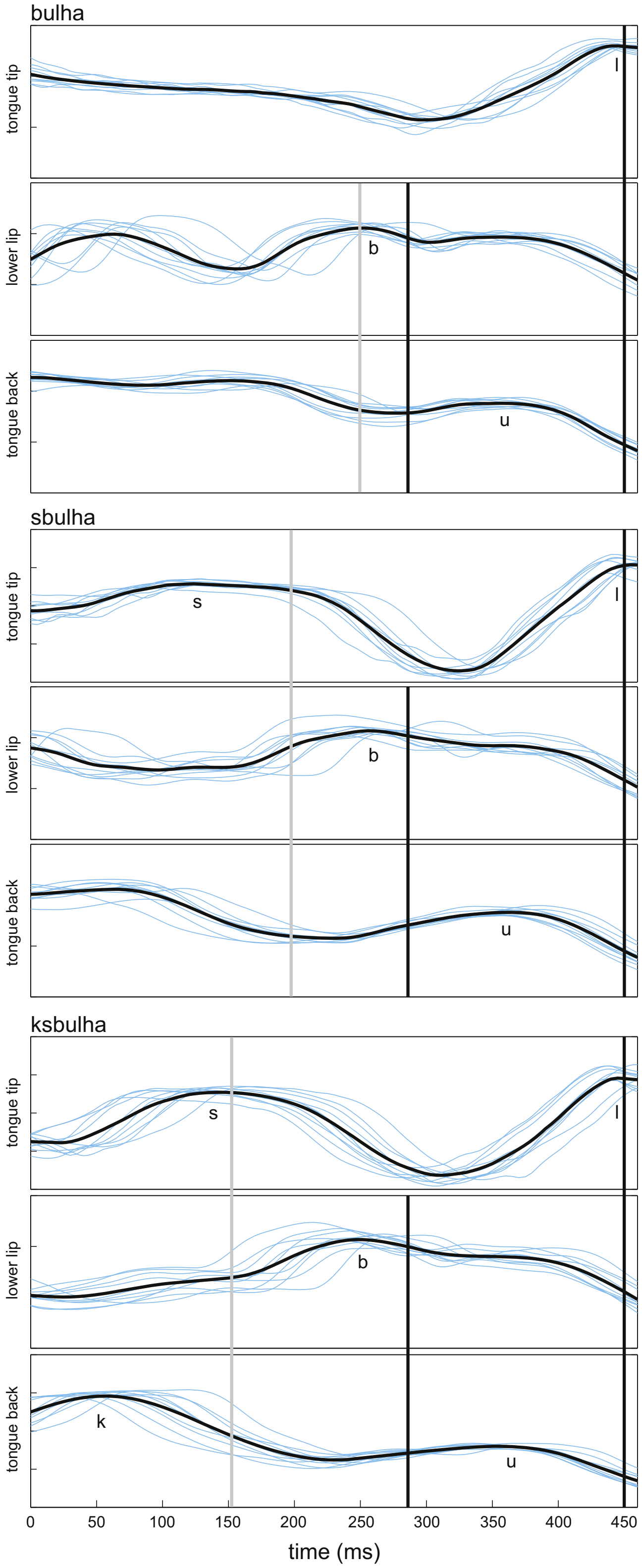

**Fig. 4.** Positional signals of sensors placed on the tongue tip, lower lip, and tongue back for 10 repetitions each of /bulha/, /sbulha/, and /ksbulha/ (light blue lines), along with average trajectories (black lines). The data are from one speaker of Moroccan Arabic (Ouzda dialect). For each word, the thick vertical lines represent, from left to right, the midpoint of the initial consonant cluster (c-center), the release of the prevocalic consonant [b] (right edge), and the point of maximum constriction in the postvocalic consonant [l] (anchor). Across the three words, the right-edge-to-anchor interval remains steady, while the c-center-to-anchor interval changes as a function of the number of prevocalic consonants. For American English, Browman and Goldstein (1988) report the opposite pattern whereby, as the number of prevocalic consonants is increased from one to three, the right-edge-to-anchor shrinks, while the c-center-to-anchor interval remains relatively stable. Their data incorporates one speaker's productions of *pot*, *sot*, *lot*, *spot*, *plot*, and *splot*. Figure adapted from Shaw et al. (2009). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

Please cite this article as: Gafos, A. I., et al. Stochastic time analysis of syllable-referential intervals and simplex onsets. *Journal of Phonetics* (2014), http://dx.doi.org/10.1016/j.wocn.2013.11.007

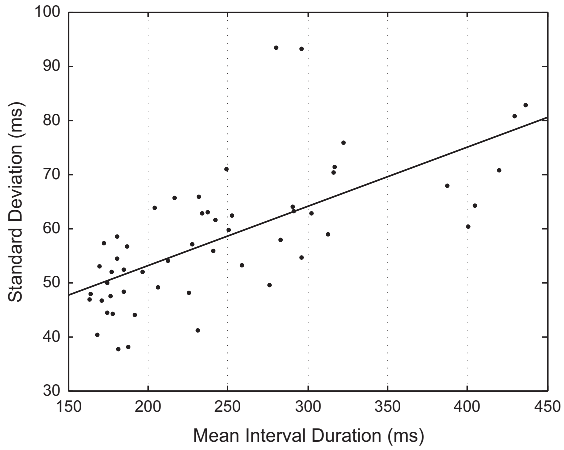

**Fig. 5.** The standard deviation of intervals (left-edge-to-anchor, c-center-to-anchor, right-edge-to-anchor) plotted against the mean duration of those intervals (plus a least-squares regression). Means and standard deviations were calculated across 61–62 repetitions by 4 Moroccan Arabic speakers of one of the nine words *lan*, *flan*, *kflan*, *bulha*, *sbulha*, *ksbulha*, *kulha*, *skulha*, *mskulha*. The six data points with mean duration greater than 375 ms are the left-edge-to-anchor intervals of the words beginning with three consonants (3 words ×2 anchor points).

**Table 1**
Relative standard deviation (RSD) of the left-edge-to-anchor (LEA), c-center-to-anchor (CCA), and right-edge-to-anchor (REA) intervals for three word sets (the triads *lan~flan~kflan*, *bulha~sbulha~ksbulha*, *kulha~skulha~mskulha*) spoken by four speakers of Moroccan Arabic (A, B, C, D). For all 12 cases, the REA has the lowest RSD, indicating relative right-edge-to-anchor stability, in line with a simplex onset parse (cf. Fig. 3). Data from Shaw et al. (2011).

| Speaker | RSD | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *lan~flan~kflan* | | | *bulha~sbulha~ksbulha* | | | *kulha~skulha~mskulha* | | |
| | LEA (%) | CCA (%) | REA (%) | LEA (%) | CCA (%) | REA (%) | LEA (%) | CCA (%) | REA (%) |
| A | 27.6 | 16.7 | **9.7** | 29.3 | 20.7 | **14.4** | 18.9 | 17.5 | **12.1** |
| B | 29.9 | 16.6 | **7.4** | 34.4 | 27.0 | **12.5** | 25.1 | 15.5 | **13.2** |
| C | 23.0 | 13.8 | **8.9** | 27.4 | 14.8 | **11.8** | 20.0 | 13.2 | **9.3** |
| D | 29.4 | 19.9 | **12.9** | 22.5 | 13.0 | **12.9** | 15.9 | 14.2 | **11.0** |

In the temporal organization corresponding to simplex onsets, the expected stability pattern is lower variability for the right-edge-to-anchor interval than for the c-center-to-anchor interval (see Fig. 3, right). A quantitative evaluation of this hypothesis adopts a conservative index of stability, the coefficient of variation, or relative standard deviation, henceforth RSD. Compared to other measures of stability, such as variance or standard deviation, RSD is conservative with respect to our hypothesis because it corrects for a well-known bias for longer intervals to be more variable. In the study of human motor behavior it is generally observed that the variance of an interval between two timed events is positively correlated with the mean duration of the interval (Wing & Kristofferson, 1973). This is true of kinematic speech data as well. We illustrate this by drawing from the dataset of Moroccan Arabic recordings in Shaw, Gafos, Hoole, and Zeroual (2011). Fig. 5 plots the mean duration of the three structurally relevant intervals ($x$-axis) against the standard deviation ($y$-axis). The intervals are those shown in Fig. 3: left-edge-to-anchor interval, c-center-to-anchor interval, and right-edge-to-anchor interval. These three intervals were calculated across 61–62 repetitions of 9 words, *lan* 'to become soft', *flan* 'someone', *kflan* '(nonce)', *bulha* 'her urine', *sbulha* 'her ear (of grain)', *ksbulha* 'to win for her', *kulha* 'to eat for her', *skulha* '(nonce)', and *mskulha* 'to hold for her', produced by 4 native speakers of Moroccan Arabic. To assess the hypothesis that, regardless of the composition of the initial clusters, the syllable containing the first vowel in these words includes only the prevocalic consonant as its onset, the consonant sequences beginning these words were chosen to exhibit a variety of sonority profiles, e.g. whereas for *flan* the sonority is rising, for *mskulha* it is falling (see Shaw et al., 2011 for further details of this corpus). The intervals for these word triads were quantified using two measurement techniques. In the first, the three intervals were right-delimited by the $C^{max}$ anchor, a time point located on the consonant after the first vowel in each of the words above. Specifically, this time point is identified by the velocity minimum of the articulator relevant to the production of the corresponding consonant (e.g. the [n] in *lan* versus the [l] in *bulha*) during the constriction phase of that consonant. In the second technique, the three intervals were right-delimited by the time point of the $V^{end}$ anchor. This anchor occurs at the end of the vowel gesture, and its time point is determined on the basis of a velocity threshold. Each of the 54 data points (9 words ×3 intervals ×2 anchor points) in Fig. 5 sums over 10–18 repetitions by 4 speakers. The key result is that, consistent with more general findings about human motor behavior, the mean duration of the intervals is positively correlated with the standard deviation of those intervals, here $r(53)=0.67$, $p<0.001$. By adopting RSD as our index of stability we correct for the inherent bias for shorter (timed) intervals to have greater stability.

Using the conservative index of stability expressed by the RSD measure, the overwhelming majority of the data conforms to a hypothesis stated in terms of stability inequalities. For simplex onsets, the right-edge-to-anchor interval has a lower RSD than the c-center-to-anchor interval. Table 1 shows this to be the case for all combinations of triad and speaker for intervals right-delimited by the $C^{max}$ anchor (as reported in Shaw et al., 2011).

Despite the predominance of right-edge-to-anchor stability, exceptions to this pattern have been reported. Table 2 shows some such exceptions. The first four examples, *bal~dbal*, *tab~ktab*, *bula~sbula*, *bulha~sbulha~ksbulha*, were reported in Shaw et al. (2009). The fifth example, *lan~flan~kflan*, comes from Speaker A as reported in Shaw et al. (2011). The intervals for these word dyads or triads were quantified using the two different anchors introduced above, the $C^{max}$ anchor (as reported for the data in Table 1) and the $V^{end}$ anchor. Looking at the RSD values in Table 2, it can be seen that when using the $C^{max}$ anchor, the right-edge-to-anchor interval showed lower RSD than the c-center-to-anchor interval. This is the canonical result for Arabic. However, when the data were quantified using the $V^{end}$ anchor, the inverse stability pattern is found: the c-center-to-anchor interval shows

Please cite this article as: Gafos, A. I., et al. Stochastic time analysis of syllable-referential intervals and simplex onsets. *Journal of Phonetics* (2014), http://dx.doi.org/10.1016/j.wocn.2013.11.007

**Table 2**
Relative standard deviation (RSD) of the three intervals, left-edge-to-anchor (LEA), c-center-to-anchor (CCA), and right-edge-to-anchor (REA), for two different anchors $C^{max}$ and $V^{end}$ as reported in Shaw et al. (2009, 2011). RSD minima are in bold. For each word set, when the intervals for that word set are quantified using the $C^{max}$ anchor, the RSD minima are found for the REA; when the intervals for that word set are quantified using the $V^{end}$ anchor, the RSD minima are found for the CCA. The rightmost column shows the standard deviation of REA as an index of variability for each word set.

| Word sets | Anchor | Interval stability (RSD) | | | Variability index |
|---|---|---|---|---|---|
| | | LEA (%) | CCA (%) | REA (%) | |
| *bal~dbal* | $C^{max}$ | 20.5 | 9.7 | **5.1** | 15 |
| | $V^{end}$ | 27.5 | **22.7** | 25.3 | 63 |
| *tab~ktab* | $C^{max}$ | 6.8 | 5.7 | **5.5** | 14 |
| | $V^{end}$ | 12.2 | **7.7** | 10.0 | 26 |
| *bula~sbula* | $C^{max}$ | 22.0 | 11.1 | **7.3** | 19 |
| | $V^{end}$ | 14.6 | **6.5** | 6.9 | 26 |
| *bulha~sbulha~ksbulha* | $C^{max}$ | 24.6 | 15.9 | **11.2** | 22 |
| | $V^{end}$ | 23.9 | **17.8** | 18.2 | 41 |
| *lan~flan~kflan* | $C^{max}$ | 27.6 | 16.7 | **9.7** | 15 |
| | $V^{end}$ | 33.6 | **27.0** | 32.3 | 53 |

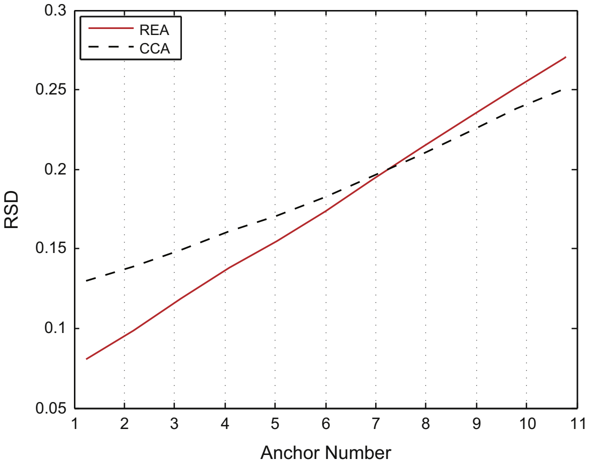

**Fig. 6.** Stability patterns generated by a stochastic model implementing the simplex onset parse. Relative standard deviation (RSD), *y*-axis, of the intervals right-edge-to-anchor (REA) and c-center-to-anchor (CCA) plotted as a function of anchor variance, *x*-axis. For anchors of low variance, anchors 1–7, the right-edge-to-anchor interval has the lowest RSD. Beyond anchor 7, however, REA becomes less stable (shows higher RSD) than CCA. The stability pattern has changed. This RSD reversal obscures the expected phonetic consequences of simplex onsets (where REA should have the lowest RSD). The main point illustrated is that the mapping between abstract syllabic organization and phonetic stability patterns is not one-to-one, because the same symbolic organization, that of simplex onsets, surfaces with the expected phonetics of simplex onsets for one range of anchor values (1–7) but also with the phonetics of complex onsets for another range of parameter values (8–11).

lower RSD than the right-edge-to-anchor interval. This latter pattern is the same as that seen in English (e.g. Browman & Goldstein, 1988; Honorof & Browman, 1995) 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 c-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, pp. 16–17) asserted, have consistent phonetic indices. In what follows, we pursue an explanation for these inconsistencies based on the approach of Shaw et al. (2009). In this approach, while the syllabic organization of a phonological sequence of consonants and vowels is kept constant, some parameters like consonant plateau duration, vowel duration or anchor variance are systematically changed. The effect of systematically changing these parameters on the phonetic indices of the assumed syllabic organization is studied. Specifically, the way in which the phonetic indices change as these parameters are scaled informs us about the dynamics of these indices.

To illustrate, Shaw et al. (2009) used a model embodying the simplex onset parse to study the effect of variability on indices of temporal stability. Via the model, they generated simulated data and studied how the patterning of temporal stability indices changes as anchor variance is systematically increased. Results are replicated in Fig. 6, showing the relative standard deviation or RSD, *y*-axis, of two intervals (right-edge-to-anchor, c-center-to-anchor) by anchors of increasing variance, *x*-axis.

It can be seen that at low levels of anchor variance, anchors 1–7, the right-edge-to-anchor interval has the lowest RSD. This is the expected phonetic reflex of simplex onsets (see, again, Fig. 3). But as anchor variance increases, the right-edge-to-anchor RSD increases at a faster rate than the c-center-to-anchor RSD. A crossover point can thus be seen after which the c-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 complex onsets, 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: both stability patterns (right edge-to-anchor more/less stable than c-center-to-anchor) are

Please cite this article as: Gafos, A. I., et al. Stochastic time analysis of syllable-referential intervals and simplex onsets. *Journal of Phonetics* (2014), http://dx.doi.org/10.1016/j.wocn.2013.11.007

consistent with the simplex onset parse. Furthermore, the simulation suggests that there are stringent conditions for the occurrence of each pattern. Given a corpus and two sets of intervals delimited by different anchors extracted from this corpus, the model embodying simplex onsets predicts the following implicational relationship: if one set of intervals shows c-center-to-anchor stability and the other shows right-edge-to-anchor stability, then the former set of intervals must correspond to an anchor with higher variance than the latter. 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 (Shaw et al., 2009). Specifically, it is only under such conditions of higher variability where the c-center-to-anchor interval may be found to show a stability advantage over the right-edge-to-anchor interval. Returning to Table 2, in the rightmost column we report an index of anchor variance, which we identify with the standard deviation of the right-edge-to-anchor interval because of the strong correlations between the standard deviation of this interval and the standard deviation of the other two intervals. For instance, in the dataset of Shaw et al. (2009), the correlation between the standard deviation of the right-edge-to-anchor interval and the standard deviation of the c-center-to-anchor interval was $r(12)=0.935$, $p<0.001$. In Table 2, the sets of measurements showing c-center-to-anchor stability (lower RSD for the c-center-to-anchor interval than for the right-edge-to-anchor interval) have a higher variability index than their corresponding sets of measurements (i.e. within the same word set) showing right-edge-to-anchor stability.[3]

The proposed interpretation of the apparently inconsistent behavior of stability indices above rests on a simulation. It is the main concern of the next section to show analytically that the behavior of the simulation is a consequence of how the temporal structure of simplex onsets behaves under different variability conditions.

## 4. Linking symbolic form to continuous data

In this section, we analytically treat the relation between a hypothesized symbolic syllabic organization and continuous stability metrics of that organization. Stochastic formalizations of phonological (syllabic) organization can be employed as analytical tools for evaluating the relation between qualitative syllable parses and their quantitative consequences. Since these formalizations are fully explicit, an analysis of their phonetic predictions can be carried out. In Section 4.1, we instantiate such an analysis for simplex onsets. In Section 4.2, we show that this analysis captures the behavior of the simulation reported in the previous section.

### 4.1. *Analytic expressions of stability patterns at the elementary level*

We proceed by first deriving analytic expressions for the stabilities of syllable-referential intervals. These expressions constitute the stability repertoire of different syllable parses. A simple consequence is that the symbolic, phonological organization of simplex onsets produces the stability pattern thought to be characteristic of that organization, but also the opposite stability pattern thought to be the hallmark of the complex onset organization. However, the two patterns are found under different parameter conditions which are made explicit.

Letting $x$ be any interval of interest, for example, an interval between two gestural landmarks, $x$'s RSD is defined as $\mathrm{RSD}_x=\sqrt{\tilde{x}}/E(x)$, where $\tilde{x}$ and $E(x)$ are, respectively, the variance and mean (expected value) of $x$. 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 c-center-to-anchor interval or any other intervals of interest.

Once RSD expressions have been derived for intervals of interest, the next step is to identify their patterning. Concretely, we can identify the conditions under which the CCA interval is *more or less stable* than the REA interval. This is done by considering the inequality $\mathrm{RSD}_{\mathrm{CCA}}<\mathrm{RSD}_{\mathrm{REA}}$, where '$\mathrm{RSD}_{\mathrm{REA}}$' represents the RSD of the right-edge-to-anchor interval and '$\mathrm{RSD}_{CCA}$' represents the RSD of the c-center-to-anchor interval. Given the previous definition of RSD, this yields the inequality in the below equation:

$$\frac{\sqrt{\widetilde{\mathrm{CCA}}}}{E(\mathrm{CCA})}<\frac{\sqrt{\widetilde{\mathrm{REA}}}}{E(\mathrm{REA})} \tag{1}$$

The conditions under which this inequality holds can be expressed in terms of a few phonetic parameters: consonantal plateau durations, inter-plateau durations, vowel durations, and their variances. These are continuous parameters whose values can be estimated from phonetic data. In other words, the statistics of a symbolic organization corresponding to a syllable parse can be expressed in terms of basic phonetic parameters. Moreover, 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 permits evaluation and analysis of the relation between qualitative organization and experimental data specific to a speaker or a population of speakers. In this section, we offer a concrete instantiation of this analytical approach.

Fig. 7 offers a schema of the basic random variables relevant to our analysis. It shows, in particular, the temporal landmarks and intervals relevant to the production of a CCCVX sequence: (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 $\tilde{\rho}$). Intervals which have been used in past experimental studies (e.g. Browman & Goldstein, 1988; Byrd, 1995; Honorof & Browman, 1995; Shaw et al., 2009) to diagnose syllabic organization can be expressed as linear combinations of these basic random variables.

---

[3] Why the $V^{end}$ anchor-delimited intervals show higher variability than the $C^{max}$ anchor-delimited intervals seems related to measurement variability and possibly to systemic factors. $V^{end}$ occurs somewhere at the end of the vowel movement, whereas $C^{max}$ occurs somewhere during the plateau of the postvocalic consonant. Vowels are characterized by slow movements of the tongue body and are identified in the articulatory labeling software using velocity thresholds. $V^{end}$ corresponds to the time point at which the velocity of the tongue body falls below some criterion value, expressed as a percentage of some local or global velocity extremum (e.g. local: velocity extremum in the release phase of the vowel in question; global: velocity extremum in some signal portion subsuming the vowel or the whole utterance). During that phase of vowels, there is typically no palatal contact and kinematics tend to be relatively flat. Instead, the time point for the $C^{max}$ anchor corresponds to an easily identifiable velocity minimum during the constriction phase of the consonant. Moreover, systemic factors may also be involved, in that for Moroccan Arabic, a language with a sparse vowel inventory, variability in vowel articulation and thus variability in delimiting the $V^{end}$ anchor landmark is to be expected.

Please cite this article as: Gafos, A. I., et al. Stochastic time analysis of syllable-referential intervals and simplex onsets. *Journal of Phonetics* (2014), http://dx.doi.org/10.1016/j.wocn.2013.11.007

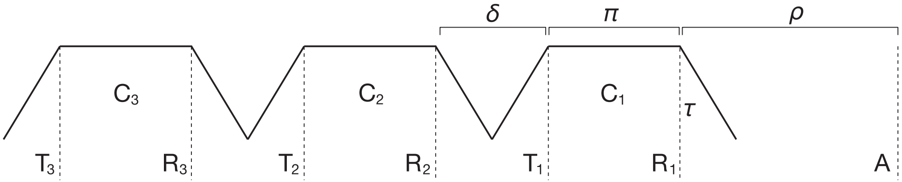

**Fig. 7.** Defining gestural landmarks in terms of mean plateau duration ($\pi$), mean inter-plateau-interval ($\delta$), and a reference point corresponding to the release of $\mathrm{C}_1$ ($E(\mathrm{R}_1)=\tau$). Relations can be expressed between landmarks, e.g. to define the expected value of $\mathrm{R}_3$, $E(\mathrm{R}_3)$, we subtract two plateaus and two inter-plateau-intervals from $E(\mathrm{R}_1)=\tau$, giving $E(\mathrm{R}_3)=\tau-2\pi-2\delta$.

We begin by fixing some arbitrary temporal value (a timestamp in milliseconds) for the right edge, $\tau$, and identify the temporal alignment of other landmarks with respect to it.[4] The resulting expressions are in (2) (recall that '$E(x)$' denotes the expected value of a random variable $x$). '$\mathrm{T}_n$' denotes the target or achievement of closure $n$, '$\mathrm{R}_m$' the release of closure $m$, and 'A' is the anchor point. (Note that we subtract from $\tau$ when the landmark of interest *precedes* $\tau$, and add to $\tau$ when the landmark of interest *follows* $\tau$.)

$$\begin{aligned}
&E(\mathrm{R}_1)=\tau, && E(\mathrm{T}_1)=\tau-\pi\\
&E(\mathrm{R}_2)=\tau-\pi-\delta, && E(\mathrm{T}_2)=\tau-2\pi-\delta\\
&E(\mathrm{R}_3)=\tau-2\pi-2\delta, && E(\mathrm{T}_3)=\tau-3\pi-2\delta\\
&E(\mathrm{A})=\tau+\rho
\end{aligned} \tag{2}$$

Along with the expected values of each interval $x$, $E(x)$, we are interested in $x$'s variance $\tilde{x}$ (recall that $\mathrm{RSD}_x$ is given by $\sqrt{\tilde{x}}/E(x)$). Variances of interest can be derived directly from the expressions in (2), using two familiar theorems from probability and stochastic processes (Papoulis, 1984)—viz. (i) the variance $\tilde{\zeta}$ of a random variable $\zeta=\alpha+\beta$ (with $\alpha$, $\beta$ independent random variables) is given by $\tilde{\alpha}+\tilde{\beta}$; (ii) for some coefficient $n$ and random variable $\zeta$ with variance $\tilde{\zeta}$, $\widetilde{n\zeta}=n^2\tilde{\zeta}$. Applying these to (2) gives the expressions in (3). (Note that squaring turns the negative coefficients in (2) positive.)

$$\begin{aligned}
&\widetilde{\mathrm{R}_1}=\tilde{\tau}=0, && \widetilde{\mathrm{T}_1}=\tilde{\pi}\\
&\widetilde{\mathrm{R}_2}=\tilde{\pi}+\tilde{\delta}, && \widetilde{\mathrm{T}_2}=4\tilde{\pi}+\tilde{\delta}\\
&\widetilde{\mathrm{R}_3}=4\tilde{\pi}+4\tilde{\delta}, && \widetilde{\mathrm{T}_3}=9\tilde{\pi}+4\tilde{\delta}\\
&\tilde{\mathrm{A}}=\tilde{\rho}
\end{aligned} \tag{3}$$

The approach of defining gestural landmarks as random variables can be applied to any phonological sequence. For concreteness, however, let us assume that, as is typical in experimental studies aiming to assess syllabic organization, we have a corpus of three types of words, CVX, CCVX and CCCVX. For such a corpus, the REA interval is coextensive with the part of the sequence between the immediately prevocalic consonant and the anchor, regardless of the number of consonants before the vowel. In other words, neither the expected value of the REA nor its variance changes across the three types of words, CVX, CCVX and CCCVX. In each case the conditional expected value of the REA is $\rho$, i.e. $E(\mathrm{REA}|\{\mathrm{cv,ccv,cccv}\})=\rho$.

We now turn to the CCA interval and its variance. For any consonant, its c-center is defined as the midpoint between the target and the release landmarks of that consonant. The c-center of a cluster of consonants is defined as the mean of the individual c-centers of all the consonants in the cluster. The CCA is the temporal difference between the c-center and the anchor point. This temporal difference depends on the specific segmental sequences in our word types. In (4), we give the expressions for the expected value of this difference for each word type, CVX, CCVX and CCCVX; E(CCA|cv) is the expected value of the CCA for the CVX type, E(CCA|ccv) for the CCVX type, and so on. We may then move from these expressions to their variances, given in (5), which in turn can be expanded in terms of the basic phonetic parameters, $\pi$, $\delta$, $\rho$, $\tilde{\pi}$, and $\tilde{\delta}$, by using the variance equations in (3)

$$\begin{aligned}
&E(\mathrm{CCA}|\,\mathrm{cv})=\mathrm{A}-\frac{\mathrm{T}_1+\mathrm{R}_1}{2}=\rho+\frac{\pi}{2}\\
&E(\mathrm{CCA}|\,\mathrm{ccv})=\mathrm{A}-\frac{\mathrm{T}_1+\mathrm{R}_1+\mathrm{T}_2+\mathrm{R}_2}{4}=\rho+\frac{2\pi+\delta}{2}\\
&E(\mathrm{CCA}|\,\mathrm{cccv})=\mathrm{A}-\frac{\mathrm{T}_1+\mathrm{R}_1+\mathrm{T}_2+\mathrm{R}_2+\mathrm{T}_3+\mathrm{R}_3}{6}=\rho+\frac{3\pi+2\delta}{2}
\end{aligned} \tag{4}$$

$$\begin{aligned}
&\widetilde{\mathrm{CCA}|\,\mathrm{cv}}=\tilde{\mathrm{A}}+\frac{\widetilde{\mathrm{T}_1}+\widetilde{\mathrm{R}_1}}{4}=\tilde{\rho}+\frac{\tilde{\pi}}{4}\\
&\widetilde{\mathrm{CCA}|\,\mathrm{ccv}}=\tilde{\mathrm{A}}+\frac{\widetilde{\mathrm{T}_1}+\widetilde{\mathrm{R}_1}+\widetilde{\mathrm{T}_2}+\widetilde{\mathrm{R}_2}}{16}=\tilde{\rho}+\frac{6\tilde{\pi}+2\tilde{\delta}}{16}\\
&\widetilde{\mathrm{CCA}|\,\mathrm{cccv}}=\tilde{\mathrm{A}}+\frac{\widetilde{\mathrm{T}_1}+\widetilde{\mathrm{R}_1}+\widetilde{\mathrm{T}_2}+\widetilde{\mathrm{R}_2}+\widetilde{\mathrm{T}_3}+\widetilde{\mathrm{R}_3}}{36}=\tilde{\rho}+\frac{19\tilde{\pi}+10\tilde{\delta}}{36}
\end{aligned} \tag{5}$$

The conditional expected values and variances above refer to individual word types, e.g. CVX, CCVX and CCCVX. Our aim is to assess the statistics of the whole corpus, as per the inequality in (1). To do so, we wish to derive expressions for the *unconditional* expected values and variances of the REA, CCA intervals. As before, the REA does not change across word types, and hence $E(\mathrm{REA})=\rho$. Likewise, since conditional REA variances, $\widetilde{\mathrm{REA}}|\{\mathrm{cv,ccv,cccv}\}$, are in each case $\tilde{\rho}$, the unconditional variance of REA is $\tilde{\rho}$

$$E(\mathrm{REA})=\rho \tag{6}$$

$$\widetilde{\mathrm{REA}}=\tilde{\rho} \tag{7}$$

---

[4] Here, for simplicity, $\tilde{\tau}=0$, but nothing hinges on the values of $\tau$ or $\tilde{\tau}$.

Please cite this article as: Gafos, A. I., et al. Stochastic time analysis of syllable-referential intervals and simplex onsets. *Journal of Phonetics* (2014), http://dx.doi.org/10.1016/j.wocn.2013.11.007

Deriving the unconditional expected value and variance of the CCA interval—respectively, $E(\text{CCA})$ and $\widetilde{\text{CCA}}$—is slightly more involved. First, $E(\text{CCA})$ can be expressed as the weighted average of each conditional expected CCA value. Assuming that each word type occurs equally frequently, we have (8) (for any set of numbers $S$, $\overline{S}$ is the arithmetic mean of $S$). To calculate the unconditional variance of the CCA, we use the expression for combined variance in (9), again assuming equal frequency of word types

$$E(\text{CCA}) = \overline{\{E(\text{CCA}|\,x) : x \in \{\text{cv,ccv,cccv}\}\}} = \frac{1}{3}[E(\text{CCA}|\,\text{cv}) + E(\text{CCA}|\,\text{ccv}) + E(\text{CCA}|\,\text{cccv})] = \pi + \frac{\delta}{2} + \rho \tag{8}$$

$$\begin{aligned}\widetilde{\text{CCA}} &= \overline{\{\widetilde{\text{CCA}|\,x} + (E(\text{CCA}|\,x) - E(\text{CCA}))^2 : x \in \{\text{cv,ccv,cccv}\}\}} \\ &= \frac{1}{3}\begin{bmatrix} \widetilde{\text{CCA}|\,\text{cv}} + (E(\text{CCA}|\,\text{cv}) - E(\text{CCA}))^2 & + \\ \widetilde{\text{CCA}|\,\text{ccv}} + (E(\text{CCA}|\,\text{ccv}) - E(\text{CCA}))^2 & + \\ \widetilde{\text{CCA}|\,\text{cccv}} + (E(\text{CCA}|\,\text{cccv}) - E(\text{CCA}))^2 & \end{bmatrix} = \frac{83\tilde{\pi} + 29\tilde{\delta}}{216} + \tilde{\rho} + \frac{\pi^2}{6} + \frac{\pi\delta}{3} + \frac{\delta^2}{6}\end{aligned} \tag{9}$$

We now have all we need to express $\text{RSD}_{\text{REA}}$ and $\text{RSD}_{\text{CCA}}$—and thus (1)—in terms of the basic random variables of our model. This is done by plugging the values obtained in (6)–(9) into the definition of RSD, as in the following equations:

$$\text{RSD}_{\text{REA}} = \frac{\sqrt{\widetilde{\text{REA}}}}{E(\text{REA})} = \frac{\sqrt{\tilde{\rho}}}{\rho} \tag{10}$$

$$\text{RSD}_{\text{CCA}} = \frac{\sqrt{\widetilde{\text{CCA}}}}{E(\text{CCA})} = \frac{\sqrt{\dfrac{83\tilde{\pi} + 29\tilde{\delta}}{216} + \tilde{\rho} + \dfrac{\pi^2}{6} + \dfrac{\pi\delta}{3} + \dfrac{\delta^2}{6}}}{\pi + \dfrac{\delta}{2} + \rho} \tag{11}$$

Concretely, as an illustration of our approach, we ask here whether there is a parameter space for which the simplex onset parse begins to display timing properties consistent with a complex onset parse. We ask, in particular, whether there exist conditions of variance $\tilde{\rho}$ beyond which the simplex onset parse exhibits greater CCA stability than REA stability—that is, the crossover point beyond which a simplex onset parse begins to have a phonetic manifestation consistent with a complex onset parse. We thus consider the inequality $\text{RSD}_{\text{CCA}} < \text{RSD}_{\text{REA}}$, as in (12), and solve for $\tilde{\rho}$

$$\begin{aligned}&\text{RSD}_{\text{CCA}} < \text{RSD}_{\text{REA}} = \frac{\sqrt{\dfrac{83\tilde{\pi} + 29\tilde{\delta}}{216} + \tilde{\rho} + \dfrac{\pi^2}{6} + \dfrac{\pi\delta}{3} + \dfrac{\delta^2}{6}}}{\pi + \dfrac{\delta}{2} + \rho} < \frac{\sqrt{\tilde{\rho}}}{\rho} \\ &\tilde{\rho} > \frac{\rho^2\left(\dfrac{83\tilde{\pi} + 29\tilde{\delta}}{216} + \dfrac{\pi^2}{6} + \dfrac{\pi\delta}{3} + \dfrac{\delta^2}{6}\right)}{\left(\pi + \dfrac{\delta}{2} + \rho\right)^2 - \rho^2}\end{aligned} \tag{12}$$

The variance, $\tilde{\rho}$, identified by the inequality in (12) is the crossover point after which the invariant symbolic form of the simplex onset organization surfaces with phonetic indices taken in previous experimental work to be instantiations of the complex onset organization.

To sum up the main point, though stability of the REA over the CCA interval is taken in experimental work as an indication of simplex onset organization (Section 3), analysis shows that the inverse stability relation can also be a consequence of the same organization. The conditions yielding emergence of the two stability patterns (REA more stable than CCA, and its inverse) can be analytically stated. More generally, the mapping between syllable structure and stability patterns is not one-to-one. This is not an unprecedented result. Ladefoged (1988), in a discussion of the mapping between discrete phonological entities and their continuous phonetic manifestations, respected the elusiveness of their relation and offered instructive examples. What is novel in our approach is the demonstration that this mapping can be studied analytically. In the specific problem studied here, the relation between syllabic organization and stability patterns, we have seen that analysis allows one to identify stringent conditions for the occurrence of each phonetic pattern.

### *4.2. 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 $\pi$, $\delta$, $\rho$, $\tilde{\pi}$, and $\tilde{\delta}$, 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 $\tilde{\rho}$. When $\tilde{\rho}$ passes a certain critical value, $\text{RSD}_{\text{CCA}} < \text{RSD}_{\text{REA}}$—i.e. CCA becomes more stable than REA.

Given the formal characterization of simplex onsets in the previous section, we can analytically predict the simulation's behavior under arbitrary combinations of input values. For example, assume $\pi = 30$ ms, $\delta = 40$ ms, $\rho = 267$ ms, and $\tilde{\pi} = \tilde{\delta} = 400\ \text{ms}^2$. Given these values, our analytical solution determines the crossover point (the point at which there is a shift from REA to CCA stability) to be $\tilde{\rho} \approx 2500.2\ \text{ms}^2$. In terms of standard deviation ($\sigma_\rho$), the crossover point is then ≈50 ms.

This is borne out by the simulation. When $\pi = 30$ ms, $\delta = 40$ ms, $\rho = 267$ ms, and $\tilde{\pi} = \tilde{\delta} = 400\ \text{ms}^2$, with 3000 averaged simulations of about ten thousand tokens, the crossover point is $\sigma_\rho \approx 50$ ms or $\tilde{\rho} \approx 2501.7\ \text{ms}^2$. See Fig. 8.

We offer an additional example. Set $\pi = 20$ ms, $\delta = 10$ ms, $\rho = 234.8$ ms, and $\tilde{\pi} = \tilde{\delta} = 100\ \text{ms}^2$. The analytical solution predicts a shift from CCA to REA stability beyond $\sigma_\rho \approx 30$ ms ($\tilde{\rho} \approx 900\ \text{ms}^2$). With these parameters, 3000 averaged simulations of about ten thousand tokens yield a crossover point of $\sigma_\rho \approx 30$ ms ($\tilde{\rho} \approx 899.7\ \text{ms}^2$). See Fig. 9.

Please cite this article as: Gafos, A. I., et al. Stochastic time analysis of syllable-referential intervals and simplex onsets. *Journal of Phonetics* (2014), http://dx.doi.org/10.1016/j.wocn.2013.11.007

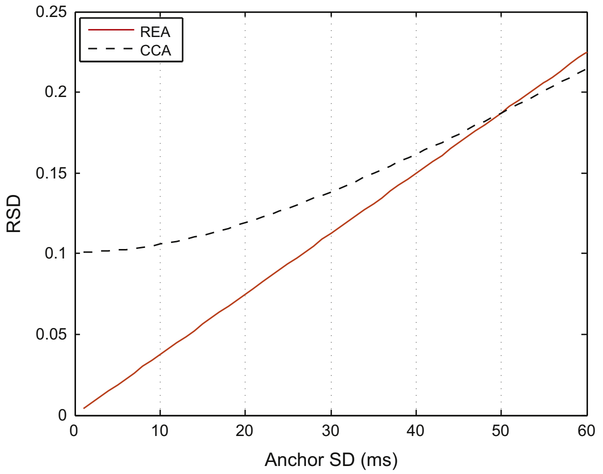

**Fig. 8.** Simulated RSDs as functions of anchor SD with $\pi = 30$ ms, $\delta = 40$ ms, $\rho = 267$ ms, and $\tilde{\pi} = \tilde{\delta} = 400$ ms$^2$. Observed shift to CCA stability beyond anchor SD ≈50 ms.

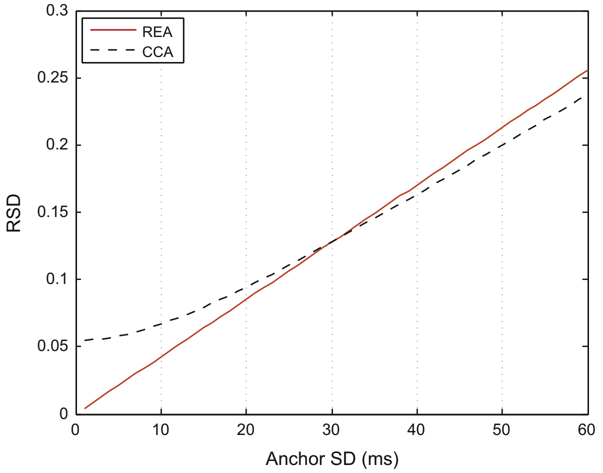

**Fig. 9.** Simulated RSDs as functions of anchor SD with $\pi = 20$ ms, $\delta = 10$ ms, $\rho = 234.8$ ms, and $\tilde{\pi} = \tilde{\delta} = 100$ ms$^2$. Observed shift to CCA stability beyond anchor SD ≈30 ms.

## 5. Stability patterns at the collective level

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. For example, ‘plea’, ‘tree’, and ‘glee’ are single syllables independent of speech rate, frequency or phonotactic probability. All of these factors will leave their 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 introduce noise in the intervals that are taken to be informative about syllable structure. Therefore, in formulating the problem of diagnosing syllable structure in experimental data, this variability must be one of the parameters whose effect in modulating stability patterns must be appreciated.

The simulations reported so far directly modulate anchor variance in order to progressively introduce variability in syllable structure-referential intervals. In other words, anchor variance serves as a parameter whose scaling perturbs the coordination topology of a hypothesized syllabic organization. The effects of that perturbation on the temporal indices of the syllabic organization are studied. However, it is in practice difficult to change this parameter systematically within a given speaker's dataset. In this section, we seek a way to introduce variability in syllable structure-referential intervals which opens up the possibility of using experimental data. Specifically, we consider sample lexica composed of tokens from a diverse range of contexts, i.e. without controlling for segmental identity, speech rate, or speaker. How do the phonetic indices, which are taken to be informative about the syllabic structuring of phonetic form, behave as we nudge and push at the relevant syllable structure-referential intervals by progressively introducing bigger and more variable lexica? Using existing experimental data, by progressively increasing the size of the lexical sample over which stability indices are quantified, we can study the effect of lexical size (and its concomitant variability) on the assumed phonetic indices of syllabic organization.

We analyzed speech movement data collected with Electromagnetic Articulometry (Perkell et al., 1992; Hoole & Zierdt, 2010) from three native speakers (‘A’, ‘B’, and ‘C’) of Moroccan Arabic. Speakers A, B, and C each produced between 13 and 15 tokens of ((*k*)*f*)*lan*, ((*k*)*s*)*bulha*, and ((*m*)*s*) *kulha*, for a total of 386 production tokens distributed across 27 types (three speakers × three possible *CVX[5] word types × three prevocalic possibilities—i.e. CVX, CCVX, or CCCVX).

---

[5] Where ‘*’ is a possibly empty string of consonants.

Please cite this article as: Gafos, A. I., et al. Stochastic time analysis of syllable-referential intervals and simplex onsets. *Journal of Phonetics* (2014), http://dx.doi.org/10.1016/j.wocn.2013.11.007

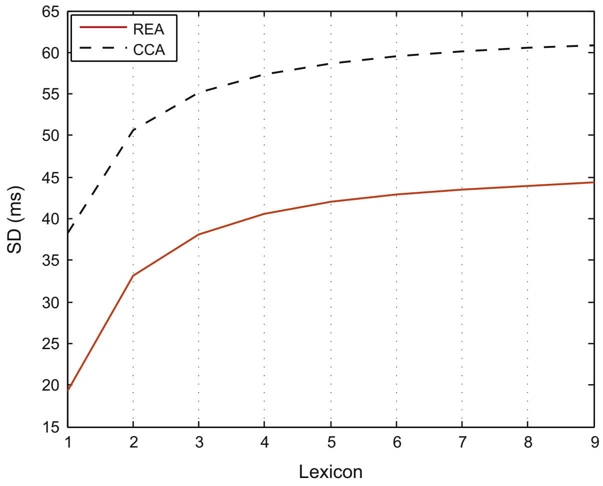

**Fig. 10.** $\mathrm{SD}_{\mathrm{REA}}$ and $\mathrm{SD}_{\mathrm{CCA}}$ as a function of lexical size. Initial segregation into speaker-word-type triads, with $9! = 362\,880$ averaged simulations. Both $\mathrm{SD}_{\mathrm{REA}}$ and $\mathrm{SD}_{\mathrm{CCA}}$ increase monotonically as a function of lexical size.

Using Mark Tiede's Matlab-based software Mview, we extracted landmark timestamps corresponding to constriction targets and releases of prevocalic consonants. Additionally, we extracted timestamps corresponding to the velocity minimum during the constriction for the postvocalic consonant (i.e. $C^{max}$). This landmark was treated as the anchor point for the calculation of the REA and CCA intervals. As in the analytical treatment above, REA was defined as the temporal distance from the release of the immediately prevocalic consonant to the anchor point, and CCA 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).

We performed an initial segregation of the production tokens into nine type-triads. Each triad contained all the productions *for a given speaker* of **lan*, **bulha*, or **kulha*. For instance, a single triad might consist of the REA and CCA data for all of Speaker A's productions of {*lan*, *flan*, *kflan*}, or the REA and CCA data for all of Speaker C's productions of {*kulha*, *skulha*, *mskulha*}.

Unselectively combining these triads—i.e. without controlling for word type, speaker, etc.—allows us to progressively assemble lexica with increasing degrees of variability. Given the nine initial triads just discussed, there are $9! = 362\,880$ ways to progressively assemble a lexicon containing all nine, since the possible ways to assemble lexica are isomorphic to the permutations of the nine initial triads. First, we randomly pick one of the nine triads, thereby obtaining an initial lexicon $S_1$. Next, we randomly pick one of the remaining eight triads, $t_2$, thereby obtaining a composite lexicon $S_2 = S_1 \cup \{t_2\}$ (thus, for any $S_n$ and $S_{n+1}$, $S_n \subset S_{n+1}$, and for any $m$, $|S_m| = m$). Proceeding in this way, we eventually arrive at $S_9$, which incorporates all 386 production tokens in the dataset. For any sequence of lexica $S = \langle S_1, \ldots, S_9 \rangle$ obtained in this way, we call $S$ a “(lexical) simulation”. Constructing lexica randomly in this way allows us to offer a *general* characterization of the statistical indices which emerge in progressively bigger lexica (given the initial segregation)—i.e. we ensure that whatever the pattern of quantitative indices emerging in this way, it is not due to an accidental choice of any given simulation.

It is possible to computationally assess the average statistical properties of the 9! possible lexical simulations: construct the simulations, iteratively calculate statistics over each, and average the resulting 9! sequences of statistical values of interest. More concretely, each simulation $S$ determines a sequence of nine $\langle \mathrm{SD}_{\mathrm{REA}}, \mathrm{SD}_{\mathrm{CCA}}, \mathrm{RSD}_{\mathrm{REA}}, \mathrm{RSD}_{\mathrm{CCA}} \rangle$ tuples. Given a simulation $S$, $S_n$ will contain some production tokens over which we unselectively calculate $\langle \mathrm{SD}_{\mathrm{REA}}, \mathrm{SD}_{\mathrm{CCA}}, \mathrm{RSD}_{\mathrm{REA}}, \mathrm{RSD}_{\mathrm{CCA}} \rangle$; $S_{n+1}$ will contain $S_n$'s tokens, along with a new batch of tokens, over which we unselectively calculate $\langle \mathrm{SD}_{\mathrm{REA}}, \mathrm{SD}_{\mathrm{CCA}}, \mathrm{RSD}_{\mathrm{REA}}, \mathrm{RSD}_{\mathrm{CCA}} \rangle$; and so on until $S_9$. Repeating this process 9! times yields 9! sequences of nine tuples of statistical values, which we can use to calculate an averaged sequence of nine tuples of statistical values.

We used a script to characterize the average statistical properties of the dataset's possible simulations along these lines. With the initial segregation into nine speaker-word-type triads detailed above, the average statistical properties of the simulations are as follows. 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 average standard deviation/variance of REA and CCA increases monotonically as a function of lexical size (Fig. 10). However, this increase in variability has a disproportionate effect on the RSD of the REA: though average $\mathrm{RSD}_{\mathrm{REA}}$ and $\mathrm{RSD}_{\mathrm{CCA}}$ both increase monotonically as a function of lexical size, a stability reversal is observed between $S_3$ and $S_4$, see Fig. 11, with average RSD values in Table 3.

Plotting $\mathrm{RSD}_{\mathrm{REA}}$ and $\mathrm{RSD}_{\mathrm{CCA}}$ as functions of the anchor SD $= \mathrm{SD}_{\mathrm{REA}}$ yields Fig. 12, which is very much in line with the simulated results plotted in Figs. 8 and 9. As there, the slope of $\mathrm{RSD}_{\mathrm{CCA}}$ is lower than the slope of $\mathrm{RSD}_{\mathrm{REA}}$, which yields a crossover point. In Fig. 12, $\mathrm{RSD}_{\mathrm{CCA}}$ overtakes $\mathrm{RSD}_{\mathrm{REA}}$ as the more stable interval beyond an $\mathrm{SD}_{\mathrm{REA}} \approx 38.4$ ms.

It is revealing that, as lexical size increases, the quantitative patterns of stability indices, instead of wandering around randomly, change progressively or tend to the pattern expected for languages instantiating complex onsets like English—i.e. relative CCA stability. This behavior provides a stochastic temporal basis for the Onset Maximization Principle (Blevins, 1995; Clements & Keyser, 1983), viz. that languages prefer to syllabify multiple consonants with their following vowel rather than splitting the consonants across different syllables, e.g. V.CCV preferred over VC.CV. In other words, when syllabic organization is assessed over a broad range of contexts such as across a variegated set of words differing in the phonic identity of their constituent sounds, across different speakers and idiosyncratic speech rates, the resulting temporal stability patterns tend to the qualitative state characteristic of complex onsets.

The main implication of the above results is that variability contributes to shaping categorical organizational principles at the global level. Through our 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. This way of approaching the ‘basis’ of

Please cite this article as: Gafos, A. I., et al. Stochastic time analysis of syllable-referential intervals and simplex onsets. *Journal of Phonetics* (2014), http://dx.doi.org/10.1016/j.wocn.2013.11.007

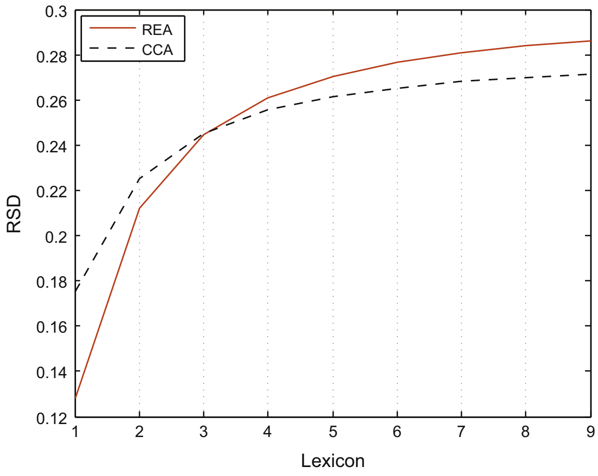

**Fig. 11.** $RSD_{REA}$ and $RSD_{CCA}$ as a function of lexical size. Initial segregation into speaker-word-type triads. 9!=362 880 averaged simulations. CCA overtakes REA as the more stable index beyond $S_3$.

**Table 3**
$RSD_{REA}$ and $RSD_{CCA}$ as functions of lexical size. Initial segregation into speaker-word-type triads. 9!=362 880 averaged simulations. RSDs for lexica showing stability reversals given in bold. CCA overtakes REA as the more stable index beyond $S_3$.

| Lexical size | $RSD_{REA}$ | $RSD_{CCA}$ |
|---|---|---|
| $S_1$ | 0.1280 | 0.1752 |
| $S_2$ | 0.2123 | 0.2254 |
| $S_3$ | 0.2445 | 0.2452 |
| $S_4$ | **0.2608** | **0.2554** |
| $S_5$ | **0.2704** | **0.2614** |
| $S_6$ | **0.2766** | **0.2653** |
| $S_7$ | **0.2808** | **0.2680** |
| $S_8$ | **0.2839** | **0.2699** |
| $S_9$ | **0.2861** | **0.2714** |

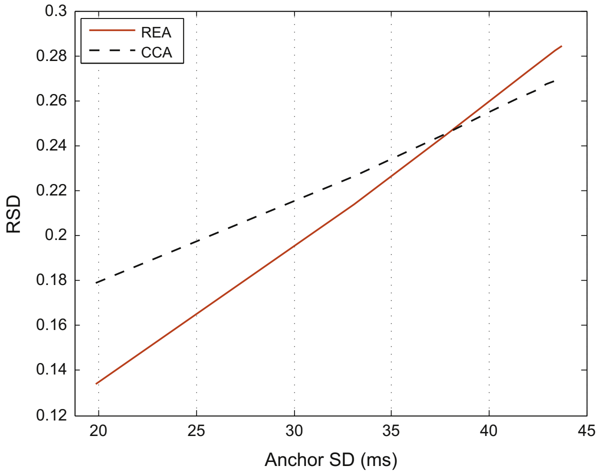

**Fig. 12.** $RSD_{REA}$ and $RSD_{CCA}$ as functions of anchor SD = $SD_{REA}$. Initial segregation into speaker-word-type triads. 9!=362 880 averaged simulations. CCA overtakes REA as the more stable index beyond $SD_{REA} \approx 38.4$ ms.

higher-level organizational principles like Onset Maximization can be paralleled to Ohala's paradigm for sources of sound change (Ohala, 1981), with the distinction being that the demonstration above is at the stochastic realm of stability samples over a lexicon.[6]

---

[6] Perhaps a more apt parallelism can be constructed by a combination of works of Goldstein (1983) and Harrington (2007). Goldstein (1983) points to the differential way in which (articulatory) variability affects acoustics; specifically, that uniformly distributed articulatory variability dictates specific (i.e. non-uniform) directional changes in acoustics, which in turn may advance sound change along the directions of those changes. Harrington (2007) argues that reorganization in a vowel inventory may be instigated by statistical properties of the lexicon; specifically, that the fronting of [u] in English RP is driven by the predominance of phonetic contexts (in the RP lexicon) where the vowel follows consonants with a high F2 locus.

Please cite this article as: Gafos, A. I., et al. Stochastic time analysis of syllable-referential intervals and simplex onsets. *Journal of Phonetics* (2014), http://dx.doi.org/10.1016/j.wocn.2013.11.007

The main point brought to forefront in the above is the importance of understanding the behavior of phonetic indices at the collective level. Variability shapes stability patterns at the collective level in coherent ways. Crucial both to the analysis of the preceding section and the simulation results in this section 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 phonological 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 phonetic instantiations.

## 6. Conclusions

We derived analytic expressions for the relation between qualitative syllabic organization and phonetic quantitative indices of that organization. These expressions make explicit the phonetic, temporal consequences of phonological (syllabic) parses. These consequences are stochastic quantities whose statistics can be subject to analysis. We pursued such an analysis for the phonological organization corresponding to simplex onsets. In this analysis, we derived effects observed in the experimental literature on the expected phonetic consequences of the simplex onset organization. Crucial to a better understanding of the relation between qualitative phonological form and its continuous phonetic manifestations are also cases where the presumed phonological organization fails to turn out with its expected consequences. The analysis pursued in this paper allows one to expose the range of validity of the phonetic manifestations of some syllabic organization by making explicit their dependence on parameter values, such as the parameter of anchor variance. Specifically, we illustrated that the relation between the syllabic organization of simplex onsets and phonetic indices shows crossover points after which the phonetic indices change to values resembling the canonical manifestation of a distinct syllabic organization, namely, the organization corresponding to complex onsets. Finally, we explored the effect of lexically induced variability on the phonetic indices of simplex onsets.

## Acknowledgments

Thanks are due to two anonymous reviewers and the editors for detailed comments and valuable feedback which have been extremely helpful in improving this paper. We also wish to thank audiences at the Workshop on Dynamic Modeling of Articulation and Prosodic Structure (University of Cologne), the Syllables Workshop (University of Münster), the University of Geneva, the West Coast Conference in Formal Linguistics 28 (Los Angeles), the University of Delaware, and Brown University, where parts of this work have been presented, for questions and comments. This material is based upon work supported by National Science Foundation Grant no. 0922437 (to Haskins Laboratories), German Research Council Grant no. HO3271/3 (to PH), and ERC Grant no. 249440 (to AIG). Fig. 4 is adapted with permission from Shaw et al. (2009). Syllabification in Moroccan Arabic: evidence from patterns of temporal stability in articulation. *Phonology* 26, 187–215, © Cambridge University Press 2009. Any shortcomings of interpretation, analysis and exposition in this paper rest solely with the authors.

## References

Anderson, S. (1974). *The organization of phonology*. New York: Academic Press.
Blevins, J. (1995). The syllable. In: J. Goldsmith (Ed.), *Handbook of phonological theory* (pp. 206–244). Oxford: Blackwell.
Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology*, *3*, 219–252.
Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, *45*, 140–155.
Browman, C. P., & Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory gestures. *Journal of Phonetics*, *18*, 299–320.
Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP, Bulletin de la Communication Parlée*, *5*, 25–34.
Byrd, D. (1995). C-centers revisited. *Phonetica*, *52*, 285–306.
Byrd, D., Tobin, S., Bresch, E., & Narayanan, S. (2009). Timing effects of syllable structure and stress on nasals: *A real-time MRI examination*. *Journal of Phonetics*, *37*, 97–110.
Clements, G. N., & Keyser, S. J. (1983). CV phonology: A generative theory of the syllable. In *Linguistic inquiry monographs* (Vol. 9). Cambridge, MA: MIT Press.
Cummins, F., & Port, R. F. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*, 145–171.
Davis, S. (1990). Italian onset structure and the distribution of *il* and *lo*. *Linguistics*, *28*, 43–55.
Dell, F., & Elmedlaoui, M. (2002). *Syllables in tashlhiyt berber and in Moroccan Arabic*. Dordrecht, Boston: Kluwer Academic Publishers.
Gafos, A. (2002). A grammar of gestural coordination. *Natural Language and Linguistic Theory*, *20*, 269–337.
Gafos, A., Hoole, P., Roon, K., & Zeroual, C. (2010). Variation in timing and phonological grammar in Moroccan Arabic clusters. In: C. Fougeron, B. Kühnert, M. d'Imperio, & N. Vallé (Eds.), *Laboratory phonology 10: Variation, detail and representation* (pp. 657–698). Berlin, New York: Mouton de Gruyter.
Goldstein, L. (1983). Vowel shifts and articulatory–acoustic relations. In: A. Cohen, & M. P.R. van den Broecke (Eds.), *Tenth international congress of phonetic sciences* (pp. 267–273). Dordrecht: Foris Publications.
Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. In: G. Fant, H. Fujisaki, & J. Shen (Eds.), *Frontiers in phonetics and speech science: Festschrift for wu zongji* (pp. 239–249). Beijing: Commercial Press.
Harrington, J. (2007). Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. In: J. Cole, & J. Hualde (Eds.), *Laboratory phonology*, Vol. 9 (pp. 125–143). Berlin: Mouton de Gruyter.
Hawkins, S. (1992). An introduction to task dynamics. In: G. J. Docherty, & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 9–25). Cambridge: Cambridge University Press.
Hermes, A., Grice, M., Mücke, D., & Niemann, H. (2008). Articulatory indicators of syllable affiliation in word initial consonant clusters in Italian. In *8th international seminar on speech production*, Strasbourg, France, 8–12 December 2008.
Honorof, D. N., & Browman, C. P. (1995). The center or edge: *How are consonant clusters organized with respect to the vowel?* In: K. Elenius, & P. Branderud (Eds.), *Proceedings of the XIII international congress of phonetic sciences*, Vol. 3 (pp. 552–555). Sweden: Stockholm.
Hoole, P., & Zierdt, A. (2010). Five-dimensional articulography. In: B. Maassen, & P. H.H. M. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 331–349). Oxford University Press.
Kahn, D. (1976). *Syllable-based generalizations in English phonology* (Ph.D. dissertation). MIT (New York: Garland, 1980).
Kelso, J. A. S., & Tuller, B. (1985). Intrinsic time in speech production: Theory, methodology, and preliminary observations. *Haskins Laboratories: Status Report on Speech Research* SR-81 (pp. 171–205).
Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures. I. Theoretical lines of convergence. In: G. E. Stelmach, & J. Requin (Eds.), *Tutorials in motor behavior* (pp. 3–47). North-Holland Publishing Company.
Krakow, R. (1999). Physiological organization of syllables: *A review*. *Journal of Phonetics*, *27*, 23–54.
Ladefoged, P. (1988). The many interfaces between phonetics and phonology. *UCLA Working Papers in Phonetics*, *70*, 13–23.
Matteson, E., & Pike, K. I. (1958). Nonphonemic transition vocoids in Piro (Arawak). *Miscellanea Phonetica*, *3*, 22–30.
Max, L., & Caruso, A. J. (1997). Acoustic measures of temporal intervals across speaking rates: *Variability of syllable- and phrase-level relative timing*. *Journal of Speech, Language and Hearing Research*, *40*(5), 1097–1110.

McCarthy, J., & Prince, A. (1993). Generalized alignment. In: G. Booij, & J. van Marle (Eds.), *Yearbook of morphology* (pp. 79–153). Dordrecht: Kluwer Academic Publishers.

Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language and Hearing Research*, *44*(4), 778–792.

Munson, B., & Solomon, N. P. (2004). The effects of phonological neighborhood density on vowel articulation. *Journal of Speech, Language and Hearing Research*, *47*(5), 1048–1058.

Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. In *Proceedings of the XV international congress of the phonetic sciences* (pp. 2253–2256). Barcelona, Spain.

Nittrouer, S., Munhall, K. G., Kelso, J. A. S., Tuller, B., & Harris, K. S. (1988). Patterns of interarticulator phasing and their relation to linguistic structure. *Journal of the Acoustical Society of America*, *84*, 1653–1661.

Ohala, J. J. (1981). The listener as a source of sound change. In: C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Papers from the parasession on language and behavior* (pp. 178–203). Chicago: Chicago Ling. Soc.

Papoulis, A. (1984). *Probability, random variables, and stochastic processes*. New York: McGraw Hill.

Perkell, J. S., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabieta, I., & Jackson, M. T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America*, *92*, 3078–3096.

Pouplier, M. (2012). The gestural approach to syllable structure: *Universal, language- and cluster-specific aspects*. In: S. Fuchs, M. Weirich, D. Pape, & P. Perrier (Eds.), *Speech planning and dynamics* (pp. 63–96). Berlin: Peter Lang.

Shaw, J., & Gafos, A. (2010). Quantitative evaluation of competing syllable parses. In *ACL SIGMORPHON*-11, 11*th meeting of ACL special interest group in computational morphology and phonology* (pp. 54–62). Uppsala, Sweden.

Shaw, J., Gafos, A., Hoole, P., & Zeroual, C. (2009). Syllabification in Moroccan Arabic: *Evidence from patterns of temporal stability in articulation*. *Phonology*, *26*, 187–215.

Shaw, J., Gafos, A., Hoole, P., & Zeroual, C. (2011). Dynamic invariance in the phonetic expression of syllable structure. *Phonology*, *28*, 455–490.

Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, *21*, 291–311.

Wing, A. M., & Kristofferson, A. B. (1973). Response delays and the timing of discrete motor responses. *Perception and Psychophysics*, *14*, 4–12.

Please cite this article as: Gafos, A. I., et al. Stochastic time analysis of syllable-referential intervals and simplex onsets. *Journal of Phonetics* (2014), http://dx.doi.org/10.1016/j.wocn.2013.11.007