

## Structure in mind, structure in vocal tract

Adamantios I. Gafos · Jens Roeser · Stavroula

Sotiropoulou · Philip Hoole · Chakir Zeroual

**Abstract** We update our understanding of the view that grammar regulates inter-segmental temporal coordination and present an extension of that view to a new do-

---

main: we argue that inter-segmental coordination is basic to prosody. It is the glue joining segments together differently in different languages (here, illustrated with examples from Arabic and Spanish) and orchestrates their unfolding in ways corresponding to constructs posited in theoretical analysis. The correspondence is one between organization in mind-brain and organization in vocal tract. Moreover, for *both* mind-brain and vocal tract, the organization is phonological and abstract. It is so because it holds over segments of various identities: in Arabic, the first segment in /bka/ is not prosodified as part of the same unit as /ka/ and this holds true also for /blat/, /klat/ and so on, regardless of sonority. In contrast, in English or Spanish, a different organization holds. Crucially, uniformity in organization (same organization presiding over sequences with varying segmental makeup) does not imply uniqueness of phonetic exponents: prosodic organization is pleiotropic, simultaneously expressed by more than one phonetic exponent. Finally, two properties of coordination relations are underscored: lawful flexibility and abstractness. The first is revealed in the degrees of freedom with which movements corresponding to any given effector begin; the second in invariances of task-relevant kinematic signatures regardless of the effectors implicated in any given segmental sequence. Once again, abstract phonological structure is mirrored in vocal tracts via coordination relations holding across physiology and the particular modes of its operation.

**Keywords** coordination · Moroccan Arabic · Spanish · Electromagnetic Articulatory · syllables · prosody · pleiotropy

## 1 Introduction

In this paper, we update our understanding of the view that grammar is involved in regulating inter-segmental temporal coordination and present an extension of that view to a new domain: we argue that inter-segmental coordination is basic to prosody. It is the glue that joins segments together, differently in different languages (here, illustrated with examples from Arabic and Spanish), and orchestrates their unfolding in ways that correspond to constructs posited in theoretical analysis. The correspondence is one between organization in mind-brain and organization in vocal tract: the latter mirrors the (language-specific) properties of the units produced. Moreover, for both mind-brain and vocal tract, the organization is phonological and abstract. In other words, abstract linguistic structure lives in both systems.

This paper is organized as follows. Sect. 2 (Background) reviews relevant aspects of the phonology and morphology of Moroccan Arabic. A previous theoretical treatment was based on data from transcriptions. Here, the data are Electromagnetic Articulometry recordings. Sect. 3 (Present aims) and 4 (Structure in the vocal tract) specify exactly which aspects of the theory we aim to link to data and why, and places our study in the context of other recent work. Specifically, in Sect. 3, after introducing the leading theoretical view on syllables in this language along with competing hypotheses, we present the theoretical postulate for which we seek evidence: the first segment in /bka/ ‘he cried’ is not prosodified as part of the same unit as /ka/ (and the same holds for /blat/ ‘be old’, /klat/ ‘eat’ and so on, regardless of sonority). We express this, in Sect. 4 (Structure in the vocal tract), in terms of a span of organization hypothesis: in a CCV, global organization ranges over the entire segmental complex whereas local organization ranges over the inner CV only. Metrification is local in

Moroccan Arabic but global in languages such as English or Spanish where some of the above sequences are claimed to be single syllables. Sect. 5 outlines the method of registering our data and our stimuli. Sect. 6 (Stability of local timing) demonstrates that in sequences with varying cluster size (VCV, VCCV, VCCV), it is a local timing interval, defined over the CV portion to the exclusion of any material outside of that CV, that exhibits stability. The span of organization indexed by this local timing interval coincides with the main theoretical position on syllables in this language. Sect. 7 (A closer look at sonority) examines timing as a function of the sonority of the intervocalic consonants. Its results point to the same conclusion as in the previous section: a local span of organization; we find stability of local timing regardless of the sonority profile of the prevocalic cluster. We conclude in Sect. 8 by emphasizing the importance for linguistic theory of inter-segmental coordination and the fundamentally spatio-temporal nature of the units which need to be invoked so that the gap between theory and data we take up in this paper can be closed.

## 2 Background

Chomsky and Halle (1968, p. 354) wrote that obstruent consonants (stops, fricatives and affricates) cannot form syllables by themselves or in combination with other consonants. However, subsequent work provided considerable evidence from phonotactics, morphology and versification that there exist languages wherein syllables are composed entirely of consonants and more specifically that nuclei in these syllables can be obstruents (Clements 1990; Dell and Elmedlaoui 1985, 1988, 2002; Prince

and Smolensky 2004). Moroccan Arabic (MA) has been taken to be an illustration of this case.

To those familiar with the work of Dell and Elmedlaoui on Imdlawn Tashlhyit Berber (Dell and Elmedlaoui 1985), a first approximation would be to say that MA is like Tashlhyit. However, the evidence suggests that MA goes further than Tashlhyit in the range of possible syllabic parses involving consonants. In Tashlhyit, the relative sonority of consonants plays a major role in determining syllabic form. Thus, in Tashlhyit, /zrb/ is syllabified with /r/ as a nucleus, /zrb/ ‘be in a hurry’, while the parse /z.rb/ with /z/ in a separate syllable (and /b/ as the nucleus of the second syllable) is not possible. In contrast, in MA, both syllabic parses are claimed to be possible. Dell and Elmedlaoui (2002), drawing from extensive analysis of the Lmnabha dialect of MA, offer /zrb/ (one syllable) and /z.rb/ (two syllables) as shown shown in Fig. 1, meaning ‘hurriedness’ and ‘be in a hurry’ respectively. Let us make it clear that we adopt notation from our sources. What appears in slashes, plus any additional diacritics, indicates only the phonologically relevant aspects of the claimed representations (here, its syllabic organization). We address issues of the relation between such notation and the phonetics below. In effect, given a consonantal string, the MA range of possible syllabic parses is greater than that of Tashlhyit.

The reason for MA’s greater range of syllabic parses, according to Dell and Elmedlaoui (2002), is that morphological templatic requirements take priority over sonority-driven syllabification in MA. Consider /qlb/ ‘to overturn’. There are two attested syllabifications of this sequence in MA, one of which again is excluded in Berber. When in its perfective form (/qlb/, ‘he overturned’), Dell and Elmedlaoui

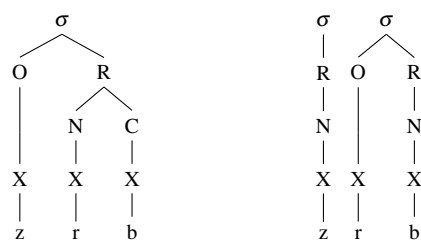


Fig. 1: /zrb/ ‘hurriedness’; /z.rb/ ‘he was in a hurry’

(2002) propose that a templatic constraint called FinL, which requires the final syllable to be light, gives rise to the parse (/qlb/, FinL)  $\rightarrow$  /q.lb/ (the same applies to all strong verbs). The parse /qlb/ violates this constraint, but /q.lb/ with onset /l/ followed by nucleus /b/ satisfies it (Dell and Elmedlaoui 2002, p. 295, 281). This is the parse shown in Fig. 1 (right). A different templatic requirement, FinH, which requires a final heavy syllable, is at work in the deverbal noun parse shown in Fig. 1 (left) as well as in a subset of underived nouns such as for /klb/ or /kəlb/ ‘dog’ whose lexical entry must be specified as (/klb/, FinH).

More relevant to the level of data description relevant in this study, let us consider some further sources of evidence that morphology plays a crucial role in regulating segmental temporal relations in MA. As in all languages known to have templatic word-formation, MA’s templatic morphology coexists with concatenative or affixal morphology. The coexistence of these two morphology types allows us to tease apart the respective contributions of word-formation type and phonology in timing patterns. When we fix the phonological make-up of consonant clusters to those comprising two homorganic stops (identical oral gestures involved) and vary morphology type, the following difference is observed (Gafos et al. 2010). In templatic morphology,

two homorganic stops are coordinated with a non-overlapped coordination relation, but in affixal morphology they are coordinated with an overlapped relation. In effect, timing depends on morphology type. Consider [!znat<sup>^</sup>t] (‘!’ denotes pharyngealization throughout; ‘<sup>^</sup>’ denotes an acoustic release, ultimately to be interpreted as the surface outcome of a coordination relation which we specify in the ensuing), the plural of [!zntit] ‘(dog’s) tail’ formed on the CCaC<sup>^</sup>C template. In this form, see Fig. 2 (left), there are two identical oral gestures, produced with two distinct closures, one for each [t], as shown by the two peaks in the articulatory movement signal (audio and movement signals were recorded using the Carstens AG500 3-D Electromagnetic Articulometry system; for more on the method, see Sect. 5). Turning to affixal morphology, two identical oral gestures across a morpheme boundary like those of [d] and [t] in [ʒbid-t(-u)] ‘I pulled (him)’ or [t] and [t] in [ʃmit-t(-u)] ‘I conned (him)’ show a single long closure spanning the two oral gestures of their respective consonants [d]-[t] or [t]-[t], the phonetic manifestation of what is known as a fake geminate. Fig. 2 (right) illustrates with a [d-t] sequence because doing so allows us to see that the resulting long closure exhibits heterogeneous voicing. This is not possible with lexical true geminate consonants which must be homogeneous in voicing, [tt] or [dd]. Tashlhyit exhibits a similar restriction which we take up below and which also demonstrates in a different way the phonological relevance of segmental temporal relations.

In terms of temporal coordination relations, the contrast illustrated in Fig. 2 shows that in templatic morphology a sequence of two homorganic stops must be separated in time. In affixal morphology, instead, the single long closure indicates that the two

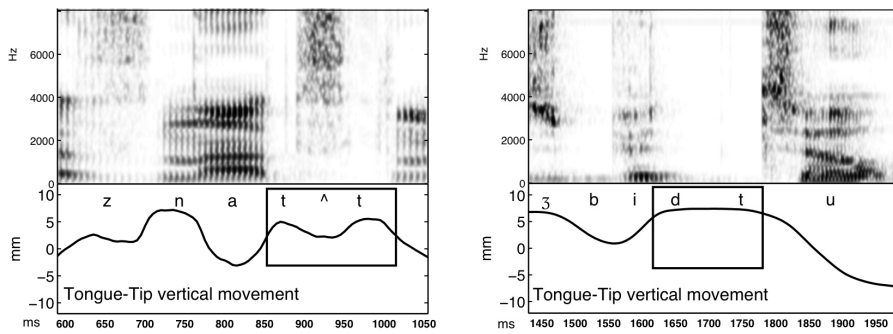


Fig. 2: Non-overlapped sequencing of two homorganic consonants within a morphological template (left) versus overlapped sequencing at the stem-affix juncture (right).

homorganic stops are coordinated in an overlapped scheme. Therefore, two distinct coordination relations are involved. This state of affairs cannot be expressed in an a-temporal phonology, where such temporal relations are not part of the representations.

For concreteness, let us introduce some definitions. A time lag between two sequential homorganic stops, as in the  $[t^t]$  part of Fig. 2, requires that the gesture of the second consonant begin late in the release phase of the gesture of the first. We refer to this coordination relation with the acoustic output  $[t^t]$  by the shorthand  $C \leftrightarrow C$ . Acoustic outputs resembling  $[C^C]$  in that the constrictions of the two consonants are separated by some time lag are also present in heterorganic clusters. Dell and Elmedlaoui (2002, p. 231) speak of a robust generalization in terms of a required intervening “audible release” between any two stops. We prefer the closely related but verifiable notion of “open transition” (Catford 1988, p. 118; see also Bloomfield 1933, section 7.9): in a sequence of two consonants  $C_1C_2$ ,  $C_1$ ’s release must precede  $C_2$ ’s constriction. In other words, there is an inter-constriction lag between  $C_1$ ,  $C_2$ . There is strong evidence for this temporal relation in MA stop-stop sequences. In our

ears, audible releases are not always present. Gafos et al. (2010) report no instances where C2's target occurs before C1 release. Speakers exhibit different mean durations for the inter-constriction gap. But a robust separation, whose duration could very well be that of a vowel in English or German, is always present.

Crucially, the temporal coordination relation involved in open transitions in heterorganic clusters is different from that in homorganic clusters even though both map to the acoustic output  $[C^{\wedge}C]$ . Fig. 3 illustrates this with articulatory traces from /bdat/ 'to start' (weak verb /bda/ + /ət/ 3pfs) and /dbal/ 'to tarnish' (perfective, 3ps). It can be seen that the two gestures overlap substantially (unlike in the homorganic case), but despite the substantial overlap the target of the second gesture is achieved after the release of the first (Gafos et al. 2010). This is shown by two vertical lines which demarcate the timepoints of the release of C1 and the target of C2 in each cluster. We use the shortcut  $C \circ C$  to refer to this coordination relation. In sum, the coordination relation for the acoustic output  $[t^{\wedge}t]$  is  $C \leftrightarrow C$  but for  $[d^{\wedge}b]$  and the fake geminates  $[dt]$ ,  $[tt]$  it is  $C \circ C$ . Fig. 4 gives schemas of the two relations. This review of the different coordination relations underlying the acoustic outputs  $[t^{\wedge}t]$  and  $[d^{\wedge}b]$  suffices to illustrate that mapping between temporal coordination relations and corresponding acoustics is not one to one and that morphology is involved in regulating segmental coordination in MA.

A theoretical formulation of coordination relations and their play out in grammars is in Gafos (2002). It was the term *temporal* coordination relations that was then used in definitions and grammar constraints referring to entities as in  $C \leftrightarrow C$  and  $C \circ C$ . In the backdrop of *a*-temporal phonological models in which that treatment of

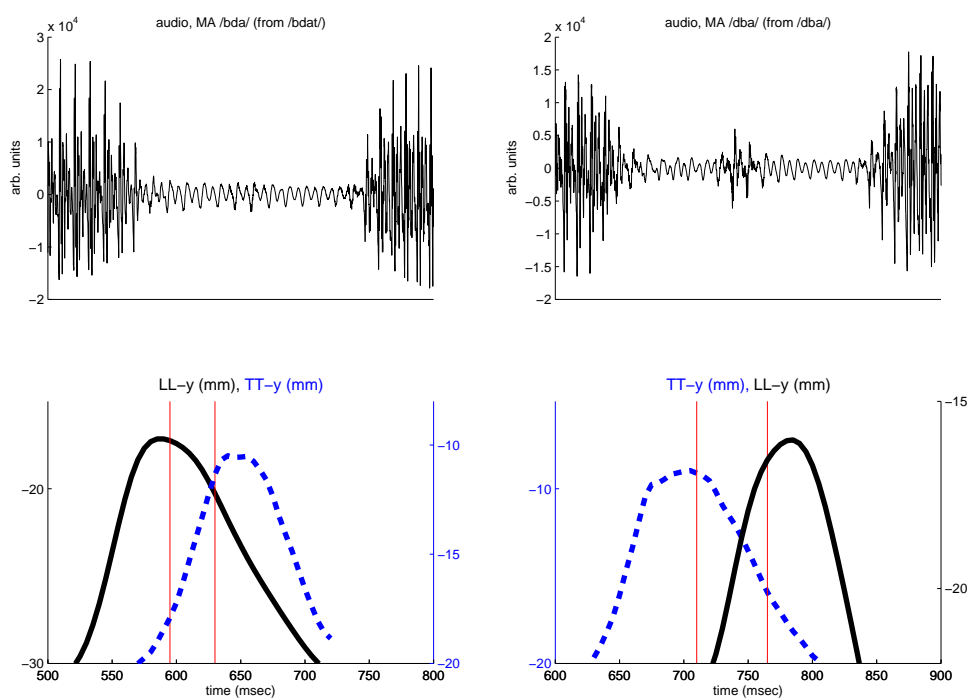


Fig. 3: Movement traces of lower lip (solid line) and tongue tip (dashed line) in /bda/, /dba/ clusters. The two vertical lines demarcate the timepoints of C1 release and C2 target, with the first always occurring before the second. This is known as an open transition. Whereas this open transition is seen in both clusters, an acoustic release is not evident in /bd/.

coordination relations was developed, this usage served the purpose of emphasizing the neglected dimension of representations for phonological theory. Here we use the term coordination relation or simply coordination. Furthermore, we use the term *timing* in the context of quantifying coordination relations by using notions of so-called ‘real time’ as in scales of milliseconds or absolute durations.

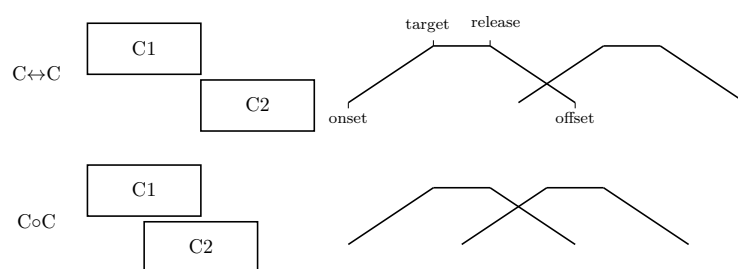


Fig. 4: In standard phonological representations, the only notion of temporal is the trivial left-to-right sequencing of segments. This notion is relevant but insufficient because it collapses the distinction between the two different coordination relations  $C1 \leftrightarrow C2$  and  $C1 \circ C2$  as shown here.  $C1$  precedes  $C2$  in both. But the relative arrangement of the characteristic landmarks (onset, target, release, offset) within the temporal structure of the two units coordinated is clearly different. This difference furthermore has phonological consequences (see text).

Shorthands such as  $C \leftrightarrow C$  or  $C \circ C$  are phonological entities. Specifically,  $C \circ C$  is a statement about language-particular organization, just as the shorthands for the symbols /b/ or /k/ denote (language-particular) phonological entities. In addition to being like a symbol,  $C \circ C$  is also a control structure of apparently the same abstractness and responsiveness to perturbation as that characterizing other areas of skill (Bernstein 1967). To appreciate the lawful flexibility and abstractness of these control structures, consider Fig. 5 showing data on how two speakers effect the open transition. The vertical line standing at the 0 of the abscissa is the  $C1$  release timepoint. Abscissa values indicate the  $C2$  onset timepoint relative to  $C1$  release. The majority of data, all points to the right of the vertical line, show  $C2$  onset starting before  $C1$  release as expected. Sometimes  $C2$  onset is after  $C1$  release, a 'late'  $C2$ . Nevertheless, all are

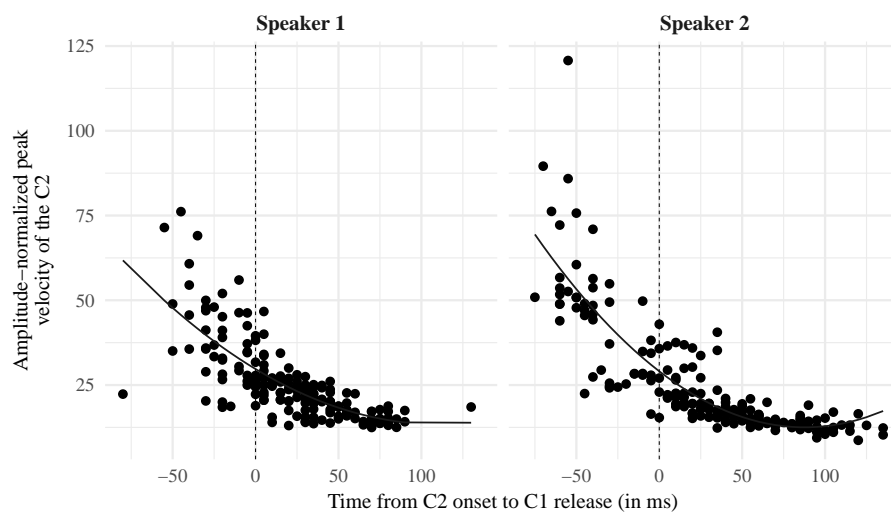


Fig. 5: In a C1C2 sequence, earlier C2 onset implies lower C2 closing phase amplitude-normalized peak velocity (regression lines from hyperbolic fits expressing this relation are drawn as curves; speaker 1:  $F(1,175) = 14.64$ ,  $p < .001$ ; speaker 2:  $F(1,176) = 75.82$ ,  $p < .001$ ). A standard dynamical system posited as a model of the control regime for the main oral constriction of a consonant is stated by the law  $\ddot{x} + b\dot{x} + k(x - x_0) = 0$  (Fowler et al. 1980; Saltzman and Munhall 1989) where  $x$  is constriction degree and  $x_0$  is the target or equilibrium (e.g., complete closure for a stop) of the system. Kinematic properties such as duration, amplitude of movement, velocity and their variability are *consequences* of the dynamical parameters  $b$  and  $k$  (damping and stiffness), whose tuning gives rise to an infinitude of trajectories. Dynamical laws do not represent explicitly every detail of an observed action but give rise to that detail. Amplitude-normalized peak velocity is related to the parameter  $k$ . Tuning  $k$  has a number of inter-related kinematic consequences one of which is the time to target (lower  $k$  means longer time to target). In effecting open transitions, the control regime seems to involve on-line adjustment of this parameter.

exemplars of the C<sub>1</sub>C<sub>2</sub> coordination relation and, regardless, the following systematicity is evident. In a C<sub>1</sub>C<sub>2</sub> sequence, the earlier the C<sub>2</sub> onset is relative to C<sub>1</sub> release, the lower the amplitude-normalized peak velocity of the C<sub>2</sub>, also known as stiffness (see caption of Fig. 5). Intuitively, the earlier C<sub>2</sub> starts while C<sub>1</sub> is active, the more C<sub>2</sub> slows down, apparently to ensure C<sub>2</sub> attaining its target after the release of C<sub>1</sub>, hence effecting the open transition. The clusters examined were [bd, db, dg, gd, br, rb, kr, rk, kl, lk, lb, nk] each embedded in three word positions (initial, medial intervocalic, final). The presence of this relation is one indication of the abstractness of the principles that underwrite phonological entities such as C<sub>1</sub>C<sub>2</sub>. Such principles, whatever their specifics turn out to be, are invariant with respect to the effectors involved in fleshing out a CC sequence.

The above properties of coordination relations and their corresponding acoustics are not a peculiarity of MA. In the Sierra dialect of Popoluca (Zoquean, Mexico), the acoustic profiles of two-consonant sequences is variable in that sometimes an audible release appears between the two consonants but other times such a release is absent (Elson 1947). Presence of a release or open transition is found when “[t]he two members of the cluster are at different points of articulation” (Elson 1947, p. 16). This transition is manifested as aspiration between voiceless consonants, as in [kɛk<sup>h</sup>.paʔ] ‘it flies’, or as a “lenis shwa vowel” especially after nasals, as in [miŋ<sup>ə</sup>.paʔ] ‘he comes’. Absence of a release or close transition, which Elson characterizes as “the lack of development of any type of aspiration or shwa vowel”, is observed “[b]etween syllables in which the final consonant of the first and the initial consonant of the second are the same point of articulation” (ibid.), as for example in the first CC sequence

of [kɛk.gak<sup>h</sup>.paʔ] ‘it flies again’ (/gak/ is a repetitive morpheme). Stop transitions, thus, show surface variability in terms of open or close transition. This variability, however, can be seen to arise from an invariant statement at the level of temporal organization, that is, a relational invariant between gestures as dictated by C◦C.

If distinctions in coordination between segments are phonologically relevant, then we expect effects of such distinctions in phonological phenomena. We give two examples. Even though as we have seen the relation  $C \leftrightarrow C$  is attested in MA, it can be shown that it is actively avoided in templates. This avoidance is manifested through effects that show consistent deviance from a phonological norm just in case, when following otherwise normal procedures for how segments appear in templates,  $C \leftrightarrow C$  would be invoked. In its templatic word-formation, MA exhibits systematic geminate separability. We illustrate with words from the Professional noun CCaCC-i, the Plural CCaC<sup>^</sup>C (again, recall that ‘<sup>^</sup>’ denotes an acoustic release), and the Passive participle m-CCuC. The Professional noun of /s<sup>w</sup>kkaɾ/ ‘sugar’ is [skakɾ-i] ‘dealer in sugar’, the Plural of /fdan/ ‘field’ is [fdad<sup>^</sup>n] and the Passive participle of /kübb/ ‘pour’ is [m-kbub]. In each case, two consonant positions in the derived form are occupied by the two ‘halves’ of a base geminate, with an intervening vowel. An independent systematicity characteristic of several MA templates is that final consonant clusters are produced with an intervening release, as in /tqüb/ ‘puncture’ → [taq<sup>^</sup>b] (Active participle), /ngür/ ‘pester’ → [t-nag<sup>^</sup>r] (Reciprocal), /nimiru/ ‘number’ → [nwam<sup>^</sup>r] (Plural). The coordination relation in the final CC clusters in these templates is C◦C. The crucial point now concerns the behavior of geminate-final bases mapped to templates with a final CC cluster. In this case, base geminates never

separate into two halves with an intervening release. For example, /kǔbb/ ‘pour’ → [kabb] (Active participle), but not \*[kab<sup>^</sup>b], /šəmm/ ‘smell’ → [t-šamm] (Reciprocal), but not \*[t-šam<sup>^</sup>m], and /mxadd-a/ ‘pillow’ → [mxadd] (Plural), but not \*[mxad<sup>^</sup>d]. This is geminate in-separability. The generalization is that geminates separate when an intervening vowel is present, /kǔbb/ → [m-kbub], but not when the intervening element is a release, /kǔbb/ → [kabb], not \*[kab<sup>^</sup>b] (a full analysis is in Gafos 2002). The latter part of this generalization illustrates the avoidance of the temporal relation required for a release between two identical consonants. This is the C↔C relation, crucially a different relation from C◦C. Why must then [!znat<sup>^</sup>t] settle for the C↔C relation? This is so because the plural template has four C positions, CCaC<sup>^</sup>C, and the singular noun from which the plural is derived [!znat<sup>^</sup>t] ‘(dog’s) tail’ is composed of four separate consonants. The requirement that all consonants of the base noun appear in the derived form dominates default coordination preferences (Bat-El 1994; McCarthy 1984).

If distinctions in coordination between segments are phonologically relevant, then we also expect effects of such distinctions in phonological phenomena like assimilation. One example comes from Tashlhyit but the same can be illustrated with other Maghrebian languages. Tashlhyit obstruent sequences in certain environments show regressive voicing, devoicing, and rounding (Dell and Elmedlaoui 1996). Regressive assimilation is optional in some contexts and obligatory in others. These assimilations interact with the presence of an acoustic release in the following way. Heterorganic consonants in Tashlhyit are systematically produced with an acoustic release. The final cluster of /t-arag-t/ ‘gift’ (f-gift-fs) is realized as [k<sup>^</sup>t] where “k is pronounced

with a clearly audible release” (Dell and Elmedlaoui 1996, p. 386). We know that the final obstruent of /-arag-/ is voiced from the plural form [t<sup>^</sup>arag-in]. Thus, assimilation is compatible with the presence of release in heterorganic sequences. Consider now a sequence of homorganic stops. Such sequences in Tashlhyit have two variant realizations, with or without a release, as in [t<sup>^</sup>d], [td]. Regressive voicing is optional, hence both of these realizations are attested. But when regressive voicing assimilation does take place, the alternative realization with the intermediate release is unacceptable. Thus in /hra#t-dl/ ‘she just covered herself’ (just 3fs-cover), as Dell and Elmedlaoui write, “the sequence /t-d/ has three realizations in free variation: t<sup>2</sup>d, td, and dd; d<sup>2</sup>d is not acceptable” (p. 386), where ‘2’ stands for the audible release (which should be expressed in terms of coordination). Dell and Elmedlaoui (1996) specifically report: “When in a sequence of sibling stops, the first is released, it cannot assimilate to the second in phonation type or in secondary labiality” [sibling stops are homorganic stops with same value for Sonorant, e.g., /t, d/, /d, T/: authors] (p. 386), where ‘T’ stands for a pharyngealized /t/. Thus, assimilation is possible in /g, t/ sequences and results in the phonetic output [k<sup>^</sup>t], but not so in /t, d/ sequences which cannot surface as \*[d<sup>^</sup>d]. As we have seen, the coordination relation involved in [k<sup>^</sup>t] is not the same as that in [t<sup>^</sup>d]. The former is C<sup>o</sup>C, the latter is C<sup>↔</sup>C. Assimilation is sensitive to this distinction in coordination. It applies in the former but not in the latter case. Once again, this distinction is not expressible in a-temporal phonological representations. For such representations, the only relevant notion of temporal is the

trivial left-to-right sequencing of segments. This notion is relevant but insufficient because it collapses the distinction between  $C1 \circ C2$  and  $C1 \leftrightarrow C2$ .<sup>1</sup>

In sum, theoretical analysis and empirical results point to the thesis that grammar, as expressed in systematic patterns of a morphological and phonological nature, is involved in the coordination of consonant sequences in MA. The following observation both closes this review and opens the way to the next section. All phenomena so far involve consonant sequences. In large part, this is so because it is between consonants where audible releases could be most directly registered by the unaided ear of the linguist who is otherwise mindful of phonetic details (Bloomfield 1933; Dell and Elmedlaoui 1996; Heath 1987). What about vowels?

### 3 Present aims

We now turn our attention to inter-segmental coordination relations that involve consonants and vowels. Our ultimate aim is to link such relations to theoretical claims about prosody in MA. Why prosody? Catford (1977, p. 172) defines the term prosody as encompassing phenomena “characteristic not so much of the individual segments as of their relations to each other”. Looked at from this perspective, inter-segmental coordination appears to be *the* juncture where segments and prosody meet.

Let us make explicit what we take to be a plausibly testable theoretical claim for MA syllables. Consider the range of converging evidence for syllable structure in

---

<sup>1</sup> See, among others, Angermeyer (2003), Benus et al. (2004), Bradley (2002, 2006, 2007), Davidson (2003, 2006), Hall (2003), Borroff (2007), Goldstein (2011), and Casserly (2012) for other analyses of phonological phenomena with grammar models based on gestural representations and or dynamical principles. See Pouplier (2011) for a review.

MA. One source of such evidence derives from patterns of seemingly puzzling variation in the phonetic forms of MA words. To wit, the word for ‘he sprinkled’ can be produced as [dr.dr] or [d<sup>ɔ</sup>r.d<sup>ɔ</sup>r], with a variably present voiced vocoid [ɔ], but not as [dr<sup>ɔ</sup>dr<sup>ɔ</sup>] (Benhallam 1980, p. 71). This variation can be explained by stating that the variably present voiced vocoid, [ɔ], can only occur after syllable onsets; [dr<sup>ɔ</sup>] is not possible because [dr] is not a legal syllable onset. The distribution of the vocoid is thereby explained parsimoniously by making reference to a ban on complex onsets. The distribution between high vowels and glides is also cleanly captured with reference to syllables which admit at most one consonant as part of their onsets. For example, the singular form of ‘son’ is [wld] and cannot be produced as [uld]. Its plural is formed by mapping these consonants onto the CCaC template. The resulting phonetic form of the plural is [u.lad] ‘sons’ and cannot be produced [wlad]. The ban on complex onsets plays out in the alternation between [w] and [u]. Because [wl] cannot be an onset, [w] is parsed into a separate syllable and surfaces as [u] in accordance with the broader cross-linguistic distribution of vowel-glide pairs (specifically, the generalization that vocalic features, shared in the vowel-glide pairs such as [u]~[w] and [i]~[j] surface variantly but systematically as a vowel in syllable nucleus position and as a glide elsewhere). Finally, work on MA versification also supports the conclusion that this language bans complex onsets (Dell and Elmedlaoui 2002; Elmedlaoui 2014). In Malhun songs, which conform to strict syllabic templates, word-initial consonant clusters cannot occupy a single beat. Such clusters are always split so that the first consonant, e.g., [g] of [glih], counts as an independent syllable (Dell and Elmedlaoui 2002, pp. 252–253).

We stress now that there exist subtle distinctions between individual theoretical analyses of MA syllables. Consider for example the word /b.ka/ ‘he cried’. Boudlal (2001) and Dell and Elmedlaoui (2002) prosodify this sequence by assigning a mora to /b/ which is in turn dominated by a syllable. Kiparsky (2003) assigns a mora which is not dominated by a syllable but links directly to the prosodic word (Boudlal 2001, pp. 62, 68; Dell and Elmedlaoui 2002, p. 252; Kiparsky 2003, pp. 159–160). Our aim here is not to evaluate whether the specific prosodic role assignments advocated in these analyses are appropriate. Even in this small set of analyses, the choices made at this high level of specificity do not ascribe to uniform criteria or aims. For example, Kiparsky (2003) makes explicit that his proposed choice is made for “the sake of concreteness” and that “[t]he question where unsyllabifiable moras are adjoined requires more study” (Kiparsky 2003, footnote 5). His approach is moreover motivated by keeping in perspective the cross-linguistic patterning in the prosodification of such “minor” syllables or non-syllable dominated moras such as the (in)ability to be stressed, carry tone, and the usually reduced range of segmental contrasts they host. Such considerations are not, at least not explicitly, in the scope of Dell and Elmedlaoui (2002) who are more involved in meticulously spelling out differences between the syllabic forms of MA and Tashlhyit. What we focus on here is that none of the analyses allows for a sequence of segments before a vowel to be in the same syllable as that vowel. That is, all three analyses share the claim that the cluster in a CCV is not a single syllable onset consisting of two consonants. In this regard, all three analyses contrast with previous treatments of MA which would consider strings such as /kra/ ‘rent’ and /skru/ ‘his plowshares’ as being monosyllables (Benhallam

1980, p. 78, 1990; Benkirane 1998, p. 346; Keegan 1986, p. 214). We take this to be a major distinction between the differing (in their details) theoretical proposals for which experimental evidence may have a chance to speak to theory.

We also do not aim to offer diagnostics for specific syllabic roles attached to individual segments in the phonetic record. Prosody and syllables more specifically refer to properties above that of the individual segment. It is thus at the level of *relations* between segments where generalizations are more likely to be found. If this conjecture is correct, it makes little sense to look at “physical correlates” of the contrast between an onset [t] versus a nucleus [t] within [t] itself.

#### **4 Structure in the vocal tract**

A theoretical argument for a postulate on mental organization draws evidence for that organization from several sources. Consider for instance the ban on complex onsets in MA wherein at least three distinct sources (patterns of variability in the presence/absence of a short vocoid, vowel-glide alternations, and evidence from syllable weight as revealed by patterns of metrification) are involved. A single theoretical postulate has a range of consequences in the spectrum of data considered usable evidence by the analyst. The more the consequences or the more seemingly disparate the phenomena it speaks to, the better the argument. We wish here to formulate our thesis for abstract organization in the vocal tract and its relation to prosody in a way that stands, in terms of the structure of the argument and the non-uniqueness of sources supporting it, at a comparable level to that of the theoretical analysis above.

Consider again the claim that MA bans complex onsets but English or Spanish admits these. We will keep with Spanish (henceforth, SP) because one conceivable source of explanation for such differences has been exemplified by a comparison between Arabic and Spanish (Steriade 1999). Consonant-liquid (CL) clusters are syllable onsets in Spanish but not in Cairene Arabic (henceforth, CairA). One source of cited evidence for this difference between SP and CairA is that speakers give syllabification judgments along the lines of /o.tros/ ‘others’, /a.bla/ ‘speaks’ for SP versus CairA /zak.ru/ ‘they studied’, /qab.lu/ ‘they met’ and so on. Stress assignment offers converging evidence, as in SP /i.dó.la.tra/, with the familiar stress on the antepenultimate when the penultimate syllable is light, versus CairA /ta.náb.la/ ‘extremely lazy (plural)’ with stress on the coda bearing penultimate. In syllable structure terms, then, SP disfavors /VC.LV/ (C is plosive, L is liquid) because, so the argument goes (Steriade 1999), speakers derive their judgments by consulting word edge phonotactics. Words ending in final stops are unattested in (the native lexicon of) SP and words starting with CL clusters are attested. CairA instead favors /VC.LV/, where the cluster is not part of a single syllable, because such clusters are impossible word-initially (although they are possible phrase-initially) and VC is possible word-finally. The latter leg of the argument is not applicable to MA and other Maghrebic languages (e.g., /!triq/, /klat/, /glih/, /kru/, /krina/, /trikku/ and so on are existing words in MA). At least in this case, another source of the difference between MA and SP should be sought.

We argue here that this source derives from differences in inter-segmental coordination which we take to be basic in expressing language-particular prosody, as

explained in the previous section. In its most succinct and general form, the hypothesis we wish to pursue can be stated by saying that adding a consonant to the left of a CV to obtain a CCV results in substantial reorganization of the spatio-temporal form of the internal CV in SP (or English) but not in MA. We express this using a notion of local versus global organization; a schema is in Fig. 6 (left). In a CCV sequence, global organization ranges over the entire segmental complex whereas local organization ranges over the inner CV subsequence only. Metrification is global in SP but local in MA. This “span of organization” difference between the two cases stands at a comparable level of generality to the statement that, using our example from the previous section, /b/ in /b.ka/ is not prosodified as part of the same unit as /ka/: as clarified in the preceding, this is our theoretical focus point here and this point stands over and above differences in specificity of formal execution of the individual theoretical analyses making the corresponding claim (Boudlal 2001, pp. 62, 68; Dell and Elmedlaoui 2002, p. 252; Kiparsky 2003, pp. 159–160).

We must now specify what we mean by reorganization in spatio-temporal form. To illustrate, we will compare the properties of the sequence /kla/ in MA vs SP. The MA vs SP comparison continues to be an apt one for our purposes for at least two further reasons. First, both languages show open transitions. Recall that inter-plateau interval duration is speaker-specific (but presence of the open transitions *per se* is not). In our sample from two MA and two SP (Castilian) speakers, for any given cluster C1C2, we normalize the inter-plateau interval (henceforth, IPI) by the total duration of the cluster as measured from target of C1 to release of C2.<sup>2</sup> The so

<sup>2</sup> The chosen sentential and phonetic context was maximally similar across MA, SP. In both, the clusters were at word initial position after /il, di \_\_por favor ‘say \_\_ please’ and /gibi \_\_ hnaja, ‘bring \_\_ here’. The

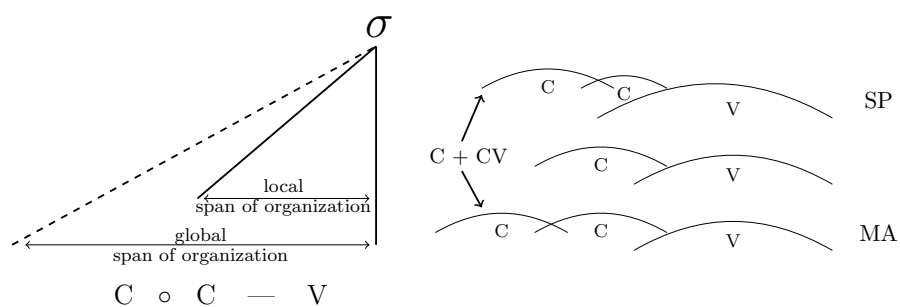


Fig. 6: Span of organization (local versus global) over a  $C \circ C - V$  sequence (left) and schematic of concomitant adjustments from /la/ to /kla/ in SP versus MA (right). Adding /k/ to the sequence /la/ (CV schema in right, middle panel) results in a number of readjustments which include substantial shortening of the /l/ in SP (refer to the shrunk second C of the CCV schema in the right, top panel) but not in MA and increase in overlap between the /l/ and the vowel in SP (refer to the mutual encroaching between the adjacent C, V of the CCV schema in the right, top panel) but not in MA.

normalized IPI durations for the two languages are 0.36 ( $SD=0.17$ ) for MA and 0.36 ( $SD=0.14$ ) for SP. Second, voicing is comparable between the two languages, with voiceless stops in particular being of the short-lag VOT type. On the surface, then, across MA and SP, the profile of a /kla/ sequence is comparable. This would not be so if we had used English or German. These differ from MA in that they do not show

---

items quantified for MA were 26 repetitions of /klam/ and 36 repetitions of /lam/. For SP, 39 repetitions of /clapas, clema, clono/ and 50 repetitions of /lapa, lema, lomo/ were quantified.

robust open transitions in at least voiced stop-liquid clusters and in that their voicing systems involve a short-lag (voiced) versus long-lag (voiceless) VOT opposition.<sup>3</sup>

The crucial point now is that despite the surface similarities between MA and SP, when a /k/ joins /la/ to form /kla/, a set of readjustments take place in SP but not or less so in MA. A first readjustment is compression of the liquid. In MA, the liquid is only 3 ms shorter in CCV than in CV, whereas this shortening is 15 ms in SP. A second readjustment concerns the presence of a relation between IPI and liquid duration. In SP, as IPI increases, the liquid's duration decreases. A long lag between the /k/ and the liquid must be compensated in SP by shortening of the liquid. In MA, we find no such relation between the two (in fact, we find a weak positive relation in MA when using raw durations). Segments are appended linearly, one after another in MA, but in SP adding a segment has consequences for the spatio-temporal properties of its adjacent segments. Fig. 7, leftmost top versus bottom panel, shows the contrast in the presence versus absence of this relation in SP vs MA.

---

<sup>3</sup> See Katz (2012) for the potentially confounding role of VOT regarding the timing of English clusters with respect to the vowel and its import for syllables. Also, the two properties mentioned here are potentially related (see Bombien and Hoole 2013).

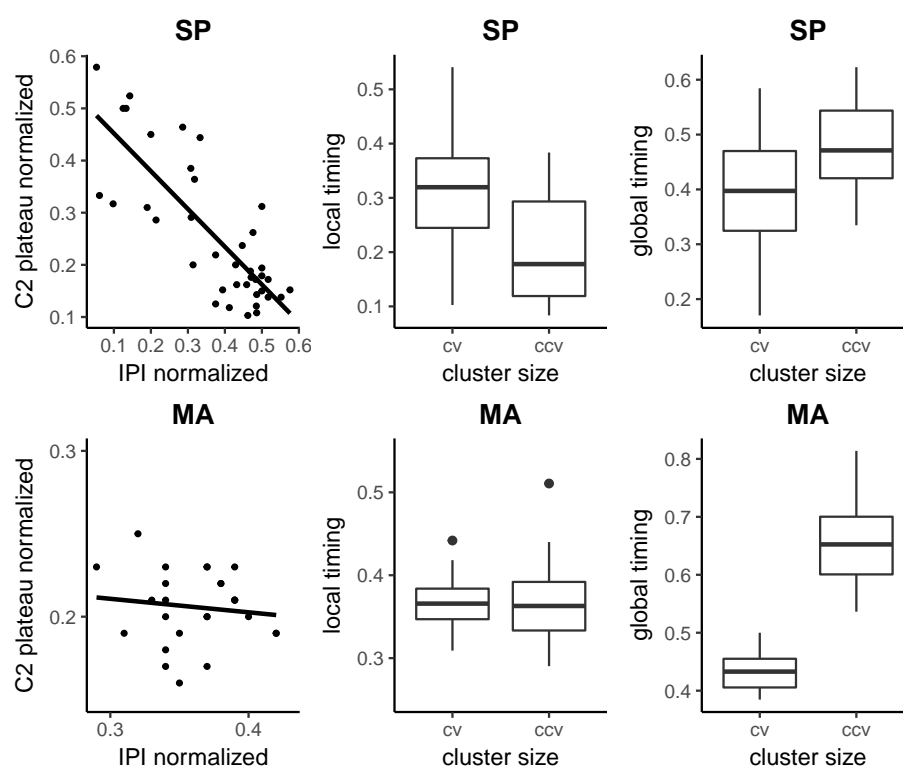


Fig. 7: Differential effects of adding /k/ to the sequence /la/ in SP vs MA: in SP (leftmost, top panel), there exists an inverse relation between the duration of the liquid and IPI (lag between the release of the /k/ and the target of /l/). This relation is absent in MA (leftmost, bottom panel) which appends segments next to one another as per its inter-segmental coordination patterns (SP:  $r = -.80$ ,  $n = 39$ ,  $p < .0001$ ; MA:  $r = -.12$ ,  $n = 26$ ,  $p = .55$ ). In SP, but not in MA, the overlap between /l/ and the vowel increases substantially as shown by the shortening of the local timing interval (middle, top panel) versus the lack of such shortening in MA (middle, bottom panel). Finally, the global timing interval changes substantially in MA (rightmost, bottom panel). This interval also changes in SP (rightmost, top panel), but less so than the local timing interval (middle, top panel). In terms of interval changes, SP and MA are mirror images of each other: in SP the local timing interval changes more than the global timing interval whereas the opposite is found in MA.

A third readjustment concerns the relation between the liquid and the tautosyllabic vowel. Let local timing refer to an interval that spans over the inner CV string (to the exclusion of any material before the CV such as the added consonant in the CCV) and global timing refer to an interval spanning over the entire string across both CV, CCV. Local timing, more specifically, is usually the interval delineated by some landmark on the immediately prevocalic consonant (e.g., the consonant's constriction release) up to some landmark on the vowel (e.g., the vowel's offset). For global timing, it is usually the interval delineated by the so-called c-center of the single consonant or consonant cluster up to the vowel offset that is used (the c-center of a single consonant or consonant cluster is the mean of the midpoints of the constriction plateaus of each consonant; Browman and Goldstein 1988). For the SP speakers, the local timing interval changes (from CV to CCV) more than the global timing interval (the change is 37 ms for local vs 3 ms for global timing). In MA, the opposite is found. For the MA speakers, the local timing interval changes (from CV to CCV) less than the global timing interval (the change is 38.3 ms vs 15.2 ms for global vs local timing). Fig. 7, middle and rightmost panels, present these results graphically. In terms of interval changes, SP and MA are mirror images of each other: in SP the local timing interval changes more than the global timing interval whereas the opposite is found in MA.

Overall, adding a segment to a sequence of segments in SP produces effects that ripple through the inner sequence in SP but not so or less so in MA. The set of properties above then serves as an indication of a global reorganization of the CCV sequence in SP but not in MA. Note that we do not seek a single or promote a privileged index

of reorganization. Pleiotropy is the important concept here: prosodic organization exerts multiple effects on the spatio-temporal properties of the segments that partake in that organization. That is, organization is simultaneously expressed by more than one phonetic exponent. Suffice it to say for current purposes that the correspondences we bring out between organization in mind-brain and vocal tract are in the formal algebraic sense indicative not of an isomorphism but a homomorphism between the two organization-encoding systems. The former relation must be bijective but the latter need not be so and there exists evidence (not crucial to our concerns in this article) that this distinction is important.

Crucially, for our current purposes, evidence on a potential correspondence between syllabic organization and temporal coordination from MA and in the majority of the other studies where this correspondence has been studied concerns word-initial clusters (but see Shaw and Gafos 2015 for word-medial evidence in English).<sup>4</sup> Consonants occupying the first position of a word tend to be longer or strengthened compared to instances of the same consonants at a non-initial position (Byrd and Saltzman 2003; Byrd et al. 2005; Fougeron and Keating 1997). Likewise, word-initial clusters have been shown to have less overlap than word-medial clusters in a variety of lan-

---

<sup>4</sup> In Tashlhyit, Ridouane et al. (2014) offer evidence from a metalinguistic task, a game wherein speakers responded to a given word by producing either its 'first part' or its 'second part'. The words used contained some word-medial clusters as in /usman/ wherein the possible answers for the 'first part' would be any of /u, us, usm/ and for the 'second part' /an, man, sman/. The summary of the reported results does not allow us to discern how many medial items were used or whether there were any differences between word-medial and word-initial clusters. In any case, participants overwhelmingly responded in a way that was interpreted as evidence for the ban on complex onsets, e.g., responses to [gli] 'guide' and [tɣwa] 'it is empty' were /g/+/li/ for the former and /tɣ+wa/ for the latter.

guages such as English, Tsou, and MA (Byrd 1996; Byrd and Choi 2010; Gafos et al. 2010; Hardcastle 1985; Wright 1996). Given such evidence on how word position affects the realization of consonants, it is reasonable to ask whether the timing patterns so far observed word-initially are also found word-medially. Such potential generalization is essential to the claim that the timing patterns so far observed reflect prosody in general and syllabic organization in particular, since in both word positions the syllabification is claimed to be the same. We turn to address this issue next.

## 5 Method and Stimuli

Speech movements and corresponding acoustics were recorded from four native speakers of the Oujda dialect of MA (three male, one female) while reading sentences at a comfortable speech rate containing various stimuli.<sup>5</sup> The stimuli, shown in Tab. 1, were embedded in the carrier sentence *zibi \_\_\_ hnaja*, ‘bring \_\_\_ here’. The carrier sentence and each stimulus within it were displayed in standard Arabic script on a computer screen. Every stimulus word was repeated about eight times, except for */halal/* which was repeated sixteen times. The stimuli in Tab. 1 consist of nine word

<sup>5</sup> Movements were recorded using the Carstens AG500 3-D Electromagnetic Articulatory (EMA) system (Hoole and Zierdt 2010; Hoole et al. 2003; Zierdt et al. 1999). Sensor coils attached to the articulators were recorded in an alternating electromagnetic field (Perkell et al. 1992). Three EMA sensors were attached to the tongue. These will be referred to as the ‘tongue tip’, ‘tongue mid’ and ‘tongue back’, and were located approximately 1.5 cm, 3 cm, 5 cm from the anterior tip (or apex) of the tongue, with the tongue at rest in the mouth. In addition, sensors were attached to the lower lip, the upper lip, the jaw, as well as to gums of the upper incisors, the bridge of the nose and the left and right side of the head behind the ears (the latter four sensors being used for head correction).

Tab. 1: Word triplets used as stimuli (twenty-seven words). Within each row, the number of the intervocalic word-medial consonants increases from one (VCV column), to two (VCCV column), to three (VCCCV column) consonants. For each triplet, the sonority profile of the three consonants is also shown as a sequence of three numbers (lower number means lower sonority). An asterisk (\*) denotes a nonce word.

Triplet	VCV	VCCV	VCCCV	Sonority
1	daba 'now'	nadba 'to moan with pain'	makdba 'lie'	111
2	kafa 'blanket or bed cover'	kamfa 'to grab (handle)'	makmfa 'handful'	132
3	naga 'camel'	sabga 'to be ahead of (precede)'	manbga 'fruit of jujube'	311
4	*baka nonce	maska 'to hold'	mamska 'chewing gum'	321
5	*hama nonce	hasma 'to be firm (decisive)'	mabsma 'smile'	123
6	qasa 'to suffer' (perf v 3ms)	qabsa 'box'	malbsa 'a dress'	412
7	taza (city name)	dabza 'to produce smth not useful'	madbza 'futility'	112
8	hala 'state'	hamla 'to be pregnant', 'to carry'	magmla 'louse'	134
9	ʒana 'to come (to us)'	ʒabna 'to give a ride'	maʒbna 'we did not give a ride'	213

triplets with a varying number of intervocalic consonants, from one to three. All VCV stimuli are parsed as V.CV, all VCCV as VC.CV, and all VCCCV as V.CC.CV. The vowel /a/ was chosen as it allows at once for maximal control of the phonetic environment in which the consonantal sequences in question are embedded and optimal delimitation of articulatory movements. The sonority profiles of the word-medial consonants in each triplet are given in the rightmost column of Tab. 1. Larger numbers indicate higher sonority. Thus, for example, in /mabsma/ the sonority profile is labeled as ‘123’ which means that the cluster starts with a segment of the lowest sonority (a stop), followed by a consonant of higher sonority (a fricative), which is in turn followed by a consonant of even higher sonority (a nasal), whereas in /mamska/ ‘321’ the order of sonority values is reversed. The variety of sonority profiles in the stimuli enables a most conservative test of the hypothesis that all word-medial consonant clusters are coordinated similarly regardless of the segmental composition or the sonority profile. This numerical indexing of sonority is used here for description and does not imply any theoretical commitment or dependency of any hypothesis on this numerical versus a natural class-based choice (such as obstruent vs nasal vs liquid) of mapping consonants on a sonority scale (Clements 1990).

Some comments on the stimuli are due. There are two nonce forms under VCV, [hama] and [baka]. Both are fully acceptable phonotactically. Consider, for example, [hama]. This sequence is found after the reflexive /t-/ as in [t+hama] (of the verb ‘to quarrel’, which is otherwise not attested in the dialect of our participants in its unprefixing simple form). Words under VCCCV, as in [makdba], consist of the negative /ma-/ prefixed to either a collective noun ([nbəg] ‘jujube fruit’, [gməl] ‘louse’) or a

deverbal noun ([kdub] 'lie' in its more frequent form or [kədb] in its less frequent form, [kmiʃ] 'handful', [lbiʃ] 'dress', [dbiʒ] 'futility'), followed by the instantiating suffix /-a/, except for the final stimulus where the form after /ma-/ is the perfective (past tense) of the verb [ʒbən] 'to bring' with /-na/ as the marker for first person plural subject. The entire construction is used in a double exclusion context as in 'it is neither a N nor a ... (different Noun)', where N stands for the noun given in the corresponding table cell (again, except the last word where the gloss given needs no further elaboration). All VCCCV stimuli are of the form ma-CəCC-a (except for the final stimulus where the third consonant of the CCC sequence is the /n/ of /-na/). What is the status of the short vowel ə in this template as far as the resulting form in our stimuli is concerned? According to Heath (1987), the most thorough analysis of both the morphology and phonology of a modern Arabic dialect (Hoberman 1995), these forms should be produced without the ə because this element is deleted by a rule of syncope triggered by the full vowel /-a/ (the rule is general in that it is triggered by any full vowel that happens to follow). In MA, just as many other vernaculars, short vowels are deleted when next to full vowels. Here are examples of the application of this rule referred to as Backward Syncope by Heath: [fχəd<sup>ʕ</sup>] 'thigh' → [fxd<sup>ʕ</sup>-i] 'my thigh', [gləs] 'sit' → [gls-na] 'we sat', [kəlb] 'dog' → [klb-ək] 'your dog', [χübz] 'bread' → [χ<sup>w</sup>bz-a] 'a loaf of bread', [kübb] 'pour' → [k<sup>w</sup>bb-i-t] 'I poured'. Heath notes that, in deliberate styles, syncope may be suspended when three or more consonants intervene between the triggering vowel and the short ə (Heath 1987, pp. 247, 248, 253). All our stimuli meet the conditions for the application of syncope. However, in what follows, we will take a conservative stance with respect to the import

of our results specifically for these forms. Even if some of the word-medial CCC sequences included a vocalic element at any level (phonetic, phonological, planning), our measures compare interval durations across VCV and VCCV as well as across VCCV and VCCCV *separately*. Thus, if it were possible to demonstrate, in one way or another, that (some of our) CCC sequences should be separated by a vowel (hence, they are not clusters), then our results would still hold true in this more restricted sense: in any VCCV sequence, where there is no evidence whatsoever that there is a vowel between the two consonants, the first consonant is not part of the onset of the second syllable.

## 6 Stability of local timing

Before evaluating stabilities of local versus global timing in our data, we review landmark identification for individual gestures and interval delimitation on the basis of these. Gestural landmark identification was based on the main articulator implicated in the formation of the constriction for each consonant. Fig. 8 illustrates delimiting the main oral gesture for a /g/, by using the positional and velocity signal of the sensor attached on tongue back. The landmarks shown are the onset (onset of movement towards the constriction), target (time point at which constriction begins), release (time point at which movement away from constriction starts), and offset (time point at which movement corresponding to control of this gesture is judged to end) (see Gafos 2002; Gafos et al. 2010). The two peaks in the lower panel correspond to velocity maxima, one towards the target or the closing phase of the movement (left peak

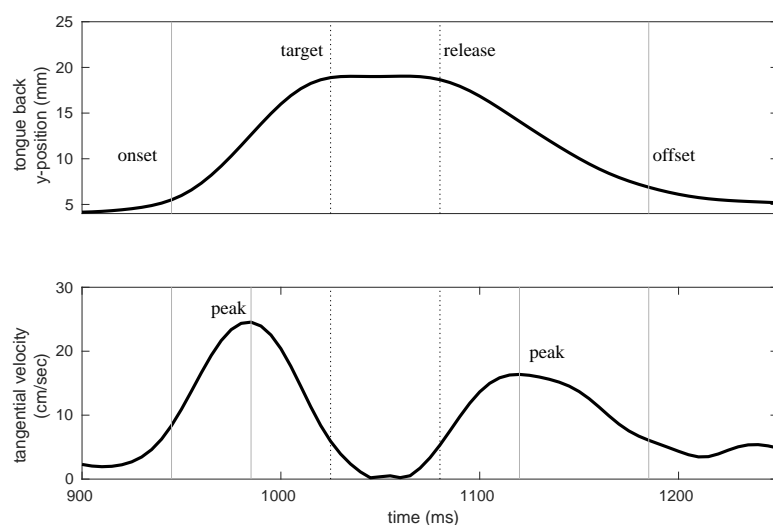


Fig. 8: Positional signal for a velar stop /g/ (upper panel) and corresponding tangential velocity (lower panel). The positional signal of the tongue back sensor in the vertical dimension (y-axis, in mm) is plotted as a function of time (x-axis, in ms). Higher values on the y-axis mean increases in tongue back height. The location of the articulatory landmarks, ‘onset’, ‘target’, ‘release’ and ‘offset’, are shown on the upper panel. The magnitude of the tangential velocity of the same sensor (in cm/sec) is also shown in the lower panel. In the lower panel, the two ‘peak’ labels correspond to the maxima in velocity for the closing phase (left peak) and the moving away or opening phase (right peak) from the target.

in Fig. 8) and the other away from the target or the opening phase of the movement (right peak in Fig. 8).

For any given VCV, VCCV, VCCCV sequence, we define two intervals, local timing and global timing. Local timing is the interval between the release of the

prevocalic consonant and a landmark at the end of or after the vowel which we refer to as the anchor. Global timing is the interval between the c-center (defined in Sect. 4) of the prevocalic consonant cluster and the anchor. To assess the robustness of our results, we used two different anchors (and thus two pairs of intervals). The first anchor was the acoustic offset in the subsequent /a/ vowel (as in Byrd 1995; Shaw et al. 2009). This anchor was manually labelled at the point of marked decline in energy in the higher formants (as seen in the synchronized spectrogram) during the end of the vowel /a/. It is a timepoint during the end of the vowel within the same syllable as the pre-vocalic consonant(s) whose relation to that vowel is at issue. In addition to this anchor, which does not necessarily correspond to a unique articulatory landmark, we used the timepoint of the peak velocity away from the target of the tongue mid sensor for the vowel. This landmark, which occurs at a kinematically well-defined location and thus can be identified algorithmically, offers a different (in this case articulatory) anchor with which to delimit the intervals of interest.

We now turn to evaluate two contrasting hypotheses, given in 1. and 2. below, about the organization of medial clusters with respect to their following vowel. These hypotheses are subsumed under a more general span of organization hypothesis as discussed in Sect. 4 (Structure in the vocal tract); see Fig. 6 therein. Here, we pursue these hypotheses within the same language as opposed to across Arabic and Spanish in Sect. 4, and within the word-medial context as opposed to the word-initial context in Sect. 4.

1. Local organization: the interval between the immediately prevocalic consonant and the vowel in a CV does not change as we add an extra consonant.

2. Global organization: the interval between the immediately prevocalic consonant and the vowel in a CV changes as we add an extra consonant.

An interaction of organization span (local vs global) and cluster size is expected under both 1. and 2. above, but the two hypotheses predict different interactions. The local organization hypothesis predicts that as cluster size increases, the local timing interval does not change but the global one does (e.g., Shaw et al. 2009). In contrast, the global organization hypothesis predicts that as cluster size increases, the local timing interval changes (decreases), but the interval corresponding to global timing does not (e.g., Browman and Goldstein 1988).

For each of our two intervals (local and global timing), interval duration mean, standard deviation (*SD*) and relative standard deviation (*RSD*) are reported in Tab. 2 across all four participants and repetitions. These descriptive statistics are shown for the interval delimited by the articulator anchor in the upper panel of the table and for the interval delimited by the acoustic anchor in the lower panel. The across participant interval means are the means of the medians (of each interval duration within each participant and each cluster), as medians have the well-known property of being more resistant to extreme values than the mean. The *RSD* is the “coefficient of variance” or  $c_v$  defined as the ratio of the standard deviation  $\sigma$  to the mean  $\mu$ ,  $c_v = 100 \frac{\sigma}{\mu}$  (see Frank and Althoen 1995 cited in Shaw et al. 2009, pp. 58–59). The ratio was multiplied by 100 to express values in percentage. A low (high) *RSD* value indicates low (high) variability.

Fig. 9 represents graphically the unaggregated data corresponding to these descriptive statistics. The interval durations for the articulatory anchor are shown in the

Tab. 2: Descriptive statistics in terms of interval duration mean (in ms), *SD* (i.e., standard deviation) and *RSD* (in %) for each triplet across all participants and repetitions. The upper panel shows the interval duration delimited by the articulatory landmark and the lower panel shows interval durations delimited by the acoustic landmark.

		Global timing			Local timing		
	Triplet	<i>mean</i>	<i>SD</i>	<i>RSD</i>	<i>mean</i>	<i>SD</i>	<i>RSD</i>
Articulatory anchor	ba~dba~kdba	202.27	43.02	21.44	144.58	16.89	12.00
	fa~mfa~kmfa	229.51	49.02	22.19	159.17	21.78	15.25
	ga~bga~nbga	212.22	43.09	20.96	150.00	20.81	15.44
	ka~ska~mska	223.89	48.13	21.97	158.96	22.60	15.16
	ma~sma~bsma	207.71	45.71	22.30	145.21	18.03	12.73
	sa~bsa~lbsa	216.11	40.99	19.49	151.88	23.52	16.26
	za~bza~dbza	215.16	45.55	21.30	150.00	21.98	14.94
	la~mla~gmla	194.88	37.83	19.64	134.79	14.19	10.73
	na~bna~ɣbna	206.44	44.32	21.58	142.50	12.09	8.54
Acoustic anchor	ba~dba~kdba	239.01	39.94	16.73	181.17	12.09	6.78
	fa~mfa~kmfa	249.42	43.55	17.44	179.10	15.37	8.47
	ga~bga~nbga	240.56	40.69	16.84	178.07	14.58	8.24
	ka~ska~mska	242.34	39.19	16.17	179.05	11.90	6.59
	ma~sma~bsma	241.74	42.85	17.77	179.92	12.04	6.73
	sa~bsa~lbsa	246.07	36.03	14.78	180.95	13.72	7.42
	za~bza~dbza	241.86	44.45	18.54	175.66	16.58	9.28
	la~mla~gmla	221.94	41.61	18.87	162.87	13.06	8.17
	na~bna~ɣbna	227.45	43.15	18.89	163.69	10.74	6.33

---

two panels on the left and the interval durations for the acoustic anchor are shown in the two panels on the right. In Fig. 9, the slope of the line through the mean of the clouds of the data points illustrates how interval duration changes as a function of cluster size. It can be seen that global timing changes (increases) as function of cluster size, as seen by its positive slope. In contrast, local timing does not vary as a function of cluster size. The same pattern is found for all triplet sets and for both anchors. This pattern will be assessed statistically in what follows.

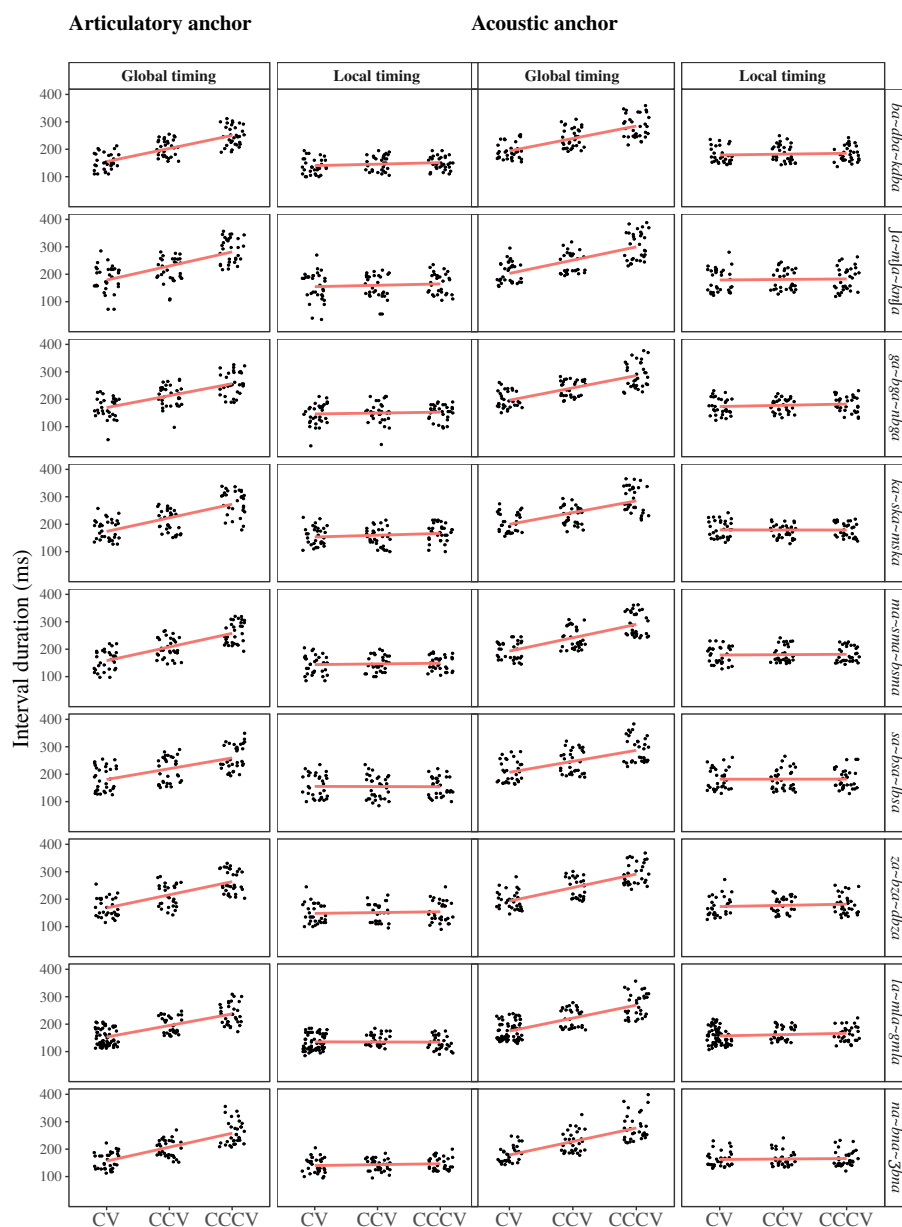


Fig. 9: Interval duration (in ms) presented in Tab. 2 plotted over cluster size (x-axis) and interval type (columns). The data for the articulatory anchors are shown in the two panels on the left, the data for the acoustic anchor are shown in the two panels on the right. Each row represents a different triplet of words from the corpus. The two columns correspond to the two intervals (global timing, local timing). The linear regression line illustrates the change (or the absence of change) in interval duration as cluster size is increased from one (CV) to two (CCV) to three (CCCV) consonants. Interval values are rendered within each cluster size, CV, CCV, CCCV, in such a way that any two datapoints with (nearly) identical interval durations within each of these categories are dispersed, for visibility reasons, on different positions in the graph (i.e., the x-axis is not continuous).

To assess how interval durations depend on cluster size, interval durations were submitted to linear mixed effects models.<sup>6</sup> Interval durations were logarithmically transformed for the purpose of model fitting to correct for the right skew observed in the raw interval durations (but all depictions of interval durations keep to the raw measures). This transformation was supported by the Box-Cox test (Box and Cox 1964). The statistical modelling results for the main effects and the interaction of cluster size and interval type are presented in the upper panel of Tab. 3, for both anchors. The intercept was specified as cluster size CCV, the reference level for the other two cluster size types. The predictor variable interval type was centred. Random slopes that did not contribute to the model fit were removed. Cluster size remained

---

<sup>6</sup> Model predictors were the main effects of cluster size and interval type and their interaction. Significant interactions were examined in nested contrasts comparing cluster sizes within interval type. Contrasts between CV vs. CCV and CCV vs. CCCV within every interval type were specified as sum contrasts by assigning  $-1$  to a baseline level and  $1$  to the corresponding contrast level,  $0$  otherwise. Sum coding allows us to determine the effect magnitude for changes in cluster size for each interval type. All models were fitted with full random effects structure, including random intercept adjustments for each subject, triplet set, and repetition nested within stimulus word (Barr et al. 2013). Parsimonious models were achieved by removing random slopes that did not contribute to the model fit (Bates et al. 2015b). The most parsimonious model included random by-subject and by-triplet set slope adjustments for cluster size and interval type. Varying random intercepts and slopes allows us to take individual differences (i.e., between subjects) into account as well as variance differences depending on the triplet set and variance associated with different repetitions of the same stimulus word. All data processing and analysis was performed in *R* (R Core Team 2013). The *lmerTest* extension of *lme4* (Bates et al. 2015a) was used to estimate the degrees of freedom for the *t*-statistics, by means of the Satterthwaite approximation, providing relatively conservative *p*-values (Kuznetsova et al. 2013). All *R*-scripts and data are available on <https://figshare.com/s/a5055436db9a4e463b58>.

Tab. 3: Linear mixed effects models for the log interval duration. The upper panel displays main effects and the interaction of cluster size and interval type. In the lower panel, nested contrasts between cluster size within interval type are summarised. The statistical results are shown for both anchors – the articulatory anchor on the left, the acoustic anchor on the right. Colons ‘:’ denote interactions.

Interaction model	Articulatory anchor			Acoustic anchor		
	<i>est.</i>	<i>t</i>	<i>p</i>	<i>est.</i>	<i>t</i>	<i>p</i>
(Intercept)	5.15	54.98	< 0.001	5.3	64.54	< 0.001
CCCVCV	0.13	5.22	< 0.001	0.1	12.06	< 0.001
CVCCV	-0.13	-5.43	< 0.001	-0.11	-6.46	< 0.001
Local-Global	0.17	22.35	< 0.001	0.15	30.07	< 0.001
{CCCVCV}:{Local, Global}	0.1	12.89	< 0.001	0.1	22.89	< 0.001
{CV, CCV}:{Local, Global}	-0.1	-12.59	< 0.001	-0.09	-21	< 0.001
<b>Nested contrasts</b>						
CCV-CV (Global)	0.23	14.08	< 0.001	0.19	19.47	< 0.001
CCCVCV (Global)	0.23	13.73	< 0.001	0.2	33.32	< 0.001
CCV-CV (Local)	0.03	1.64	0.17	0.02	1.62	0.14
CCCVCV (Local)	0.02	1.5	0.19	0.01	2.06	0.07

as a random slopes factor for varying subject intercepts and for varying triple set intercepts. Model comparisons revealed no difference between this model and the model with fully specified random effects (Bates et al. 2015b).

The significant interactions in Tab. 3 indicate that the effect of cluster size is dependent on interval type. The difference between cluster size within interval types was inspected in a linear mixed effects model with a nested contrast structure that

specifies the individual comparisons between CCV and both CV and CCCV within each interval type. The outcome of the model is shown in the lower panel of Tab. 3. The duration of the global timing interval increased as a function of increasing cluster size, but that of the local timing interval remained invariant. This pattern was observed for the interval duration of both anchors. In other words, regardless of anchor type, the interval duration increases for the global timing interval as consonants are added in front of a CV but the local timing interval remains constant. The fitted interval durations predicted by the linear mixed effects models are illustrated in Fig. 10.

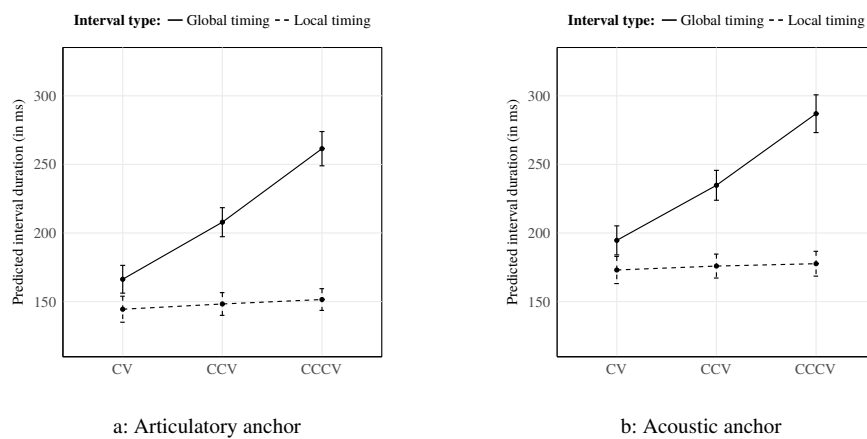


Fig. 10: Predicted interval duration (y-axis) with 95% confidence intervals. Values are predicted from the linear mixed effects model in Tab. 3. The two interval types, global and local timing, are shown in different line types and cluster size is plotted on the x-axis.

In sum, the within interval type comparisons in Tab. 3 showed significantly increasing interval durations from CV to CCV and from CCV to CCCV for the global timing interval. The difference between CV and CCV and that between CCV and CCCV for the local timing interval was not statistically significant. Given the interactions of interval type and cluster size, we can conclude that the local timing interval was not influenced by increasing cluster size, while the global timing interval increased as a function of cluster size. This is the stability pattern predicted by the local organization span hypothesis.

### 7 A closer look at sonority

There is a stark contrast between the variety of consonant clusters permitted in MA and the simplicity of its hypothesized ban on complex onsets: in any sequence of consonants followed by a vowel, only the immediately prevocalic consonant is in the same syllable as the vowel and this is so regardless of the sonority profile of the prevocalic consonants. We now turn to examine this last point more closely by considering whether the sonority profile of the consonant sequences plays a role in modulating the timing of the consonants with the subsequent vowel.<sup>7</sup>

In a first descriptive take, Fig. 11 shows the three interval durations as a function of cluster size and sonority profile.<sup>8</sup> A uniform pattern of how intervals change or do

---

<sup>7</sup> In the previous analyses, we aimed to isolate effects of cluster size and interval type and effectively treated sonority as a random factor (because triplet was a random factor and each triplet has its own sonority profile). Here, we do not treat the sonority profile as a random factor anymore.

<sup>8</sup> We use intervals delimited by the articulatory anchor. The results from intervals delimited by the acoustic anchor are qualitatively identical to those we present here.

---

not change seems to hold across the nine different sonority profiles whose numerical coding is shown on the rightmost legend of the graph. In other words, there is no indication from this descriptive look that sonority profile is a player in modulating interval durations and hence the corresponding interval stabilities.

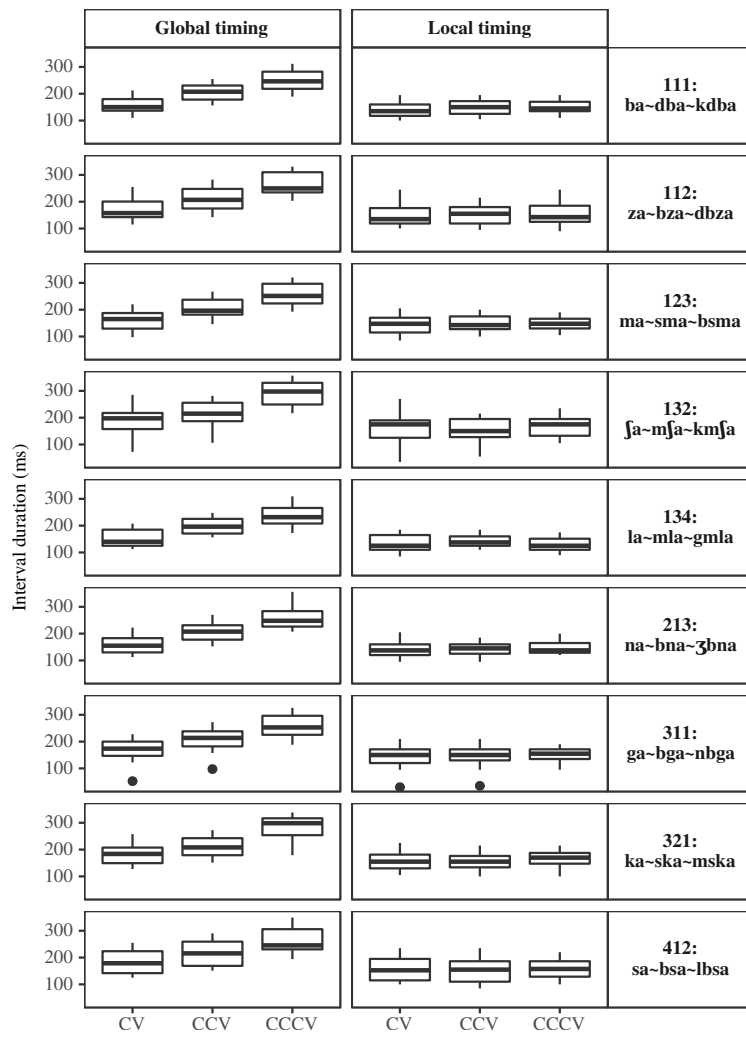


Fig. 11: Interval stability across sonority profiles as shown in the rightmost legends of the table: the duration of the local timing interval remains relatively stable across CV, CCV, CCCV, whereas that of the global timing interval increases as each consonant is added to the CV.

To assess this picture statistically, for each word triplet, we fitted a linear mixed effects model for each sonority profile, comparing interval durations across CV and CCV and across CCV and CCCV, for each of the two interval types. If the stability of interval durations depends on the sonority profile of the consonantal sequences, one would expect to see that the differences (in duration) by interval type depend on the specific triplet (as each triplet has a unique sonority profile). Fig. 12 shows the  $t$ -values for each cluster size comparison. Across triplet and hence across different sonority profiles, the local timing interval was found to be least affected by cluster size changes. This is shown by the fact that consistently across all triplets  $t$ -values for this interval hover near or below 2, corresponding to the lower bound of a significance level of  $\alpha = 0.05$  (Baayen 2008).<sup>9</sup>

For the global timing interval, consistently across triplets,  $t$ -values are well above 2. This corroborates our main analyses on the stability advantage for the local timing pattern. These data do not provide evidence that the sonority profile of the consonantal sequences affects interval stability.

---

<sup>9</sup> A potential minor exception may be the uniformly level case, shown in the top-left panel, where the  $t$ -value is slightly above 2 indicating a minor influence of either the profile or the specific segments in it (or both). Assessing these options properly would require data from several different level sonority profiles.

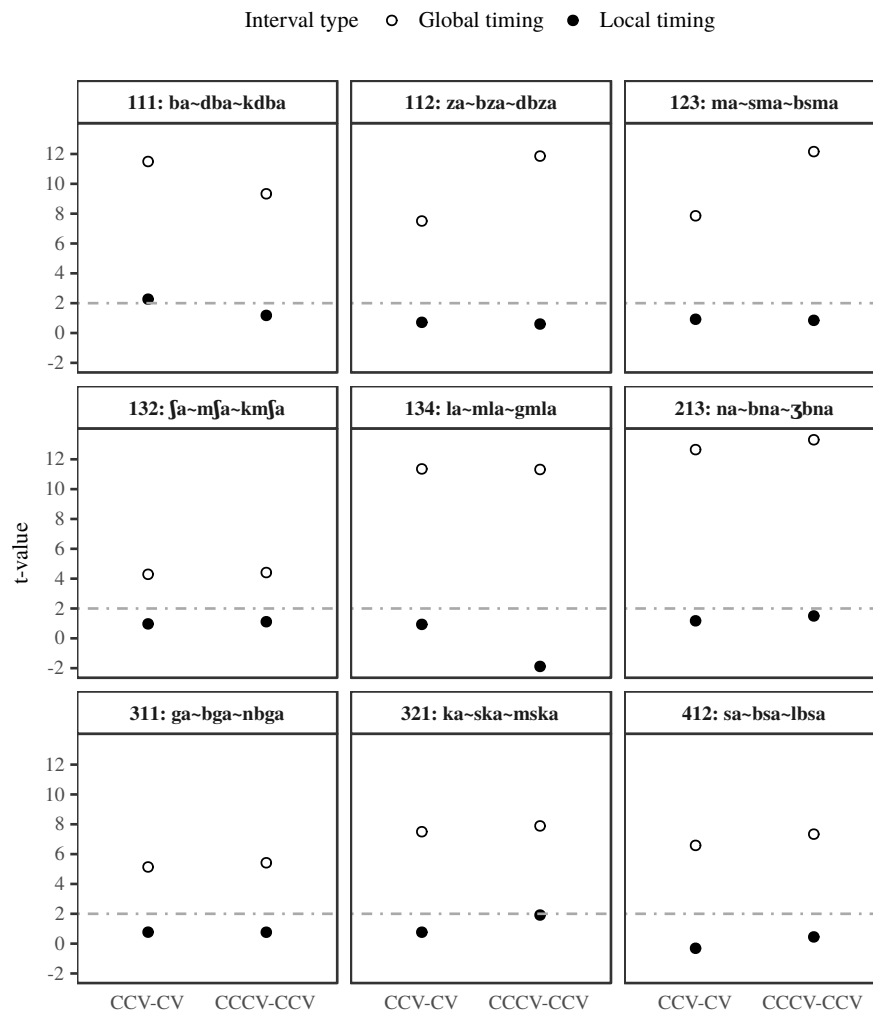


Fig. 12: Interval duration change as a function of interval type and triplet. For each triplet, the  $t$ -values for the change in interval duration between CCV-CV and between CCCV-CCV are shown for each interval type (local and global timing). The dash-dotted horizontal line indicates a  $t$ -value of two, the lower bound corresponding to a significance level of  $\alpha = 0.05$ .

A more precise test of whether sonority profile affects the stabilities of our intervals can be constructed by considering minimal pairs of consonant sequences wherein for one sequence sonority is monotonically rising whereas for the other it is monotonically falling. One pair of triplets illustrates this comparison in our data, the monotonically rising profile  $\text{hama} \sim \text{hasma} \sim \text{mabsma}$  (123) versus the monotonically falling profile  $\text{baka} \sim \text{maska} \sim \text{mamska}$  (321). Fig. 13 shows interval durations for these. The data show no stability difference between these two sonority profiles.

To assess this statistically, we note again that if sonority affects the interval duration differently across the falling and rising sonority profiles, one would expect cluster differences by-interval type to be different across the two contrasting profiles.

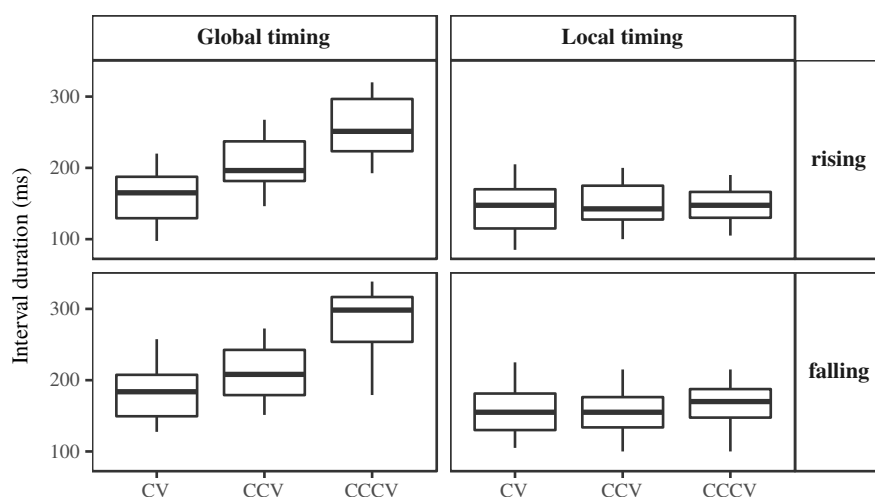


Fig. 13: Interval durations in monotonically rising (123) and falling (321) sonority profiles. The rising profile corresponds to  $\text{hama} \sim \text{hasma} \sim \text{mabsma}$  (123); the falling to  $\text{baka} \sim \text{maska} \sim \text{mamska}$  (321).

Fig. 14 shows the  $t$ -values for each cluster size comparison. Across both sonority profiles, the local timing interval was found to be least affected by cluster size changes. This is shown by the fact that consistently across both profiles  $t$ -values for this interval are below 2, corresponding to a significance level of  $\alpha = 0.05$ . For the global timing interval, consistently across both sonority profiles,  $t$ -values are well above 2.

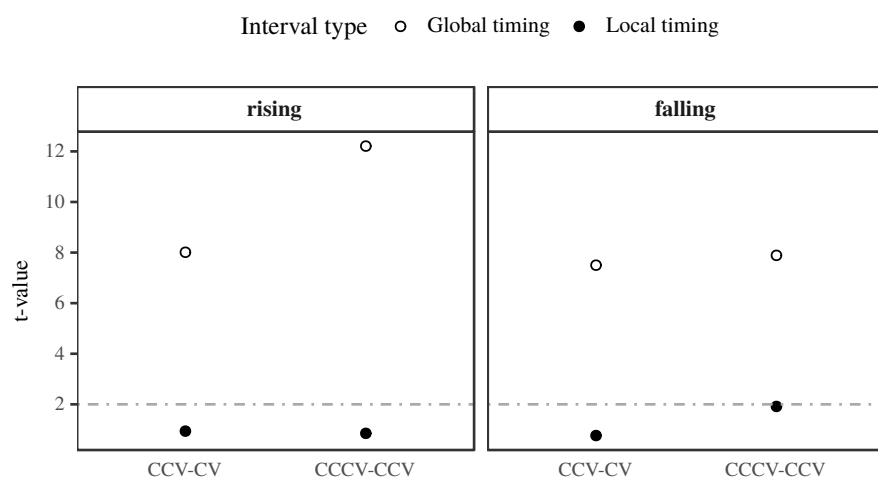


Fig. 14: Interval duration change as a function of interval type and strictly monotonically rising (123) versus monotonically falling (321) sonority profiles. The rising profile corresponds to the triplet  $\text{hama} \sim \text{hasma} \sim \text{mabsma}$  (123); the falling to  $\text{baka} \sim \text{maska} \sim \text{mamska}$  (321). For each profile, the  $t$ -values for the two comparisons CCV-CV and CCCV-CCV are shown for each interval type. The dash-dotted horizontal line indicates a  $t$ -value of two, the lower bound corresponding to a significance level of  $\alpha = 0.05$ .

This corroborates our main analyses on the stability advantage for the local timing interval.

Finally, we considered sonority profile differences across CV, CCV pairs only, excluding the CCCV, as this permits a better contrast of sonority profiles (see also discussion on the VCCCV stimuli in Sect. 5). In doing so, the available clusters can be binned into three sonority profiles; falling: baka~maska, kaʃa~kamʃa; level: daba~nadba, naga~sabga; and rising: hama~hasma, ʒana~ʒabna, qasa~qabsa, taza~dabza, hala~hamla. Again, the hypothesized syllabification for all CCVs is C.CV regardless of sonority profile. Fig. 15 illustrates the interval durations for these.

We assessed interval duration differences (how each interval changes as the string goes from CV to CCV) for each sonority profile. If interval durations change differently from CV to CCV across the falling, level and rising sonority profiles, one would expect interval duration differences not to be uniform across the three contrasting profiles. Fig. 16 shows the  $t$ -values for each interval and sonority profile. Across all three sonority profiles, the locally timed interval type was found to be least affected by cluster size changes. This is shown by the fact that consistently across profiles  $t$ -values for this interval are below 2, corresponding to a significance level of  $\alpha = 0.05$ . For the globally timed interval, consistently across all sonority profiles,  $t$ -values are well above 2. This corroborates our main analyses on the stability advantage for the locally timed interval.

To sum up, using different ways of grouping our stimuli in sonority profiles, we find no crucial evidence that sonority plays a role in modulating the timing of the consonants with the subsequent vowel. Timing patterns in our data are local in the

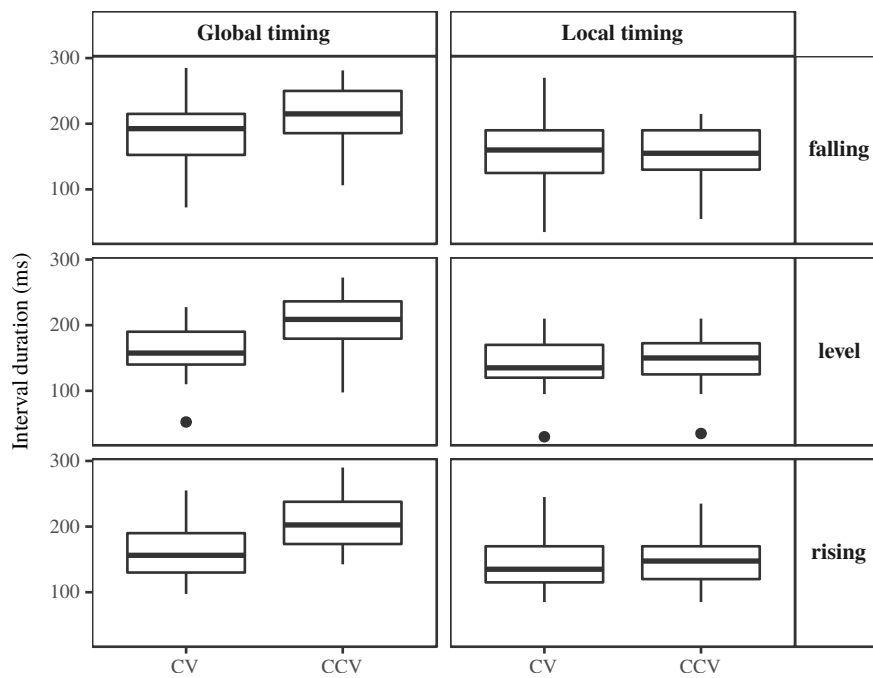


Fig. 15: Stability patterns illustrated for falling, level, and rising sonority profiles across CV, CCV pairs only (excluding CCCV). Falling: *baka*~*maska*, *kaja*~*kamja*; level: *daba*~*nadba*, *naga*~*sabga*; rising: *hama*~*hasma*, *zana*~*zabna*, *qasa*~*qabsa*, *taza*~*dabza*, *hala*~*hamla*.

following sense. It is the timing relation between the vowel and its immediately preceding C that remains most stable as the number of consonants changes from CV to CCV to CCCV and this is so regardless of the sonority profile of the consonant sequence.

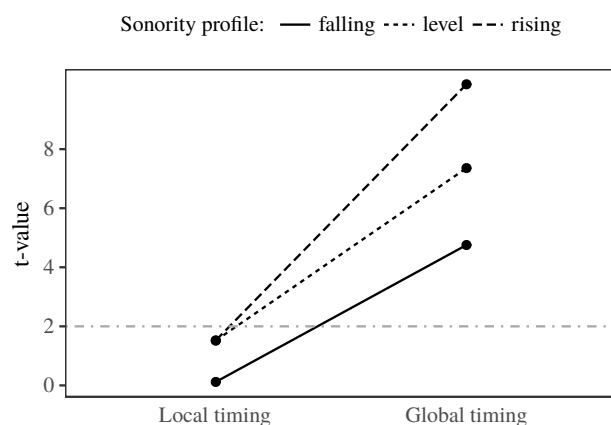


Fig. 16: Interval duration changes across CV, CCV pairs only (excluding CCCV) by interval type and sonority profile contrasting between falling, level and rising. Falling: baka~maska, kaja~kamja; level: daba~nadba, naga~sabga; rising: hama~hasma, ʒana~ʒabna, qasa~qabsa, taza~dabza, hala~hamla. For each of the three profiles, the  $t$ -values for the interval change across CCV-CV are shown for each interval type. The dash-dotted horizontal line indicates a  $t$ -value of two, the lower bound corresponding to a significance level of  $\alpha = 0.05$ .

## 8 Conclusion

We have provided evidence that the temporal organization patterns between consonants and their following vowel in MA so far documented for the word-initial context generalize to the word-medial context. If the temporal coordination patterns found in the word-initial context in MA concern inter-segmental prosody in general as opposed to just word-initial cluster timing, then the same patterns should also be found in word-medial clusters. Indeed, we find that just as the facts about word-initial con-

sonants show, when consonants are added between vowels in a word-medial VCV, no temporal re-organization can be seen in the CV substring. The added consonants are simply left-adjoined to the CV. Kiparsky (2003) uses the term “adjunction” on how these extra segments are incorporated into the prosodic parse of the string.

These results indicate that there is potential in seeking correspondences between abstract phonological analyses and patterns in phonetic data. We have specifically seen that aspects of the organization posited by phonological analysis are reflected in aspects of the temporal organization in our data. The organization brought out in the data concerns temporal stability patterns and holds over a range of segmental sequences. The data patterns are therefore at a level of generality that qualifies them as candidates for establishing links to theoretical claims which are correspondingly general. The timing patterns found hold regardless of the sonority sequencing of the segments involved. One way to describe the resulting state of affairs then is to say that the theoretical position on the lack of complex onsets in MA, despite the plethora of consonant clustering possibilities, is not merely a purely abstract property of some phonological analysis. Rather, that theoretical position also has concrete correlates in the phonetic record.

Let us clarify where our measurements stand in relation to theory and data. Our data are analogous to transcriptions as used in most work that has fueled phonological theorizing since the 1960s, with the exception that they are records of individual acts of talking obtained rigorously and quantified replicably. By quantification, one can derive (non-impressionistic) empirical generalizations which speak to theoretical constructs and their predictions. By empirical generalizations we mean precise

statements about patterns in the data (not vague statements about detailed data; see Smolensky 2006). The stability of local timing regardless of sonority profile in MA is an empirical generalization. The corresponding theoretical construct is the idea that the syllable structure in CCV sequences is uniformly C.CV.

One fairly standard interpretation of our evidence would be that the measures we have offered merely provide candidate downstream “physical correlates” of what are primarily mental aspects of the organization of sound structure. We could furthermore qualify this by saying that physical correlates may not and need not always be identifiable – Kahn (1976, pp. 16–17) does this in a landmark contribution crucial in resurrecting the construct of the syllable which Chomsky and Halle (1968) saw no need to assume – and phonological argumentation for the existence of phonological constructs can proceed untethered from phonetic record considerations.<sup>10</sup> It could serve as evidence for the above interpretation that we rely on measuring gestures and “gestures are movements” which may exist in vocal tracts but not in minds.

However, there is an alternative which points to a more foundational level of convergence. That alternative posits that the vocal tract is as much a home to phonological structure as the mind-brain is. Putting this in other words, the organization in the vocal tract preserves the (language-specific) linguistic properties of the units produced. For both mind-brain and vocal tract, the organization is phonological. Linguistic structure lives in both systems. When the organization of interest is manifest

---

<sup>10</sup> A consequence is that analytical options multiply fast. To wit, Boudlal (2001, p. 71) takes the form [k.təb] to be iambic, with a second heavy syllable. Dell and Elmedlaoui (2002, p. 295) assign the same syllabic form [k.təb] but with a second light syllable as we have seen earlier with the FinL templatic constraint.

in the vocal tract, as quantified in the individual actions of our speakers, it does not become less abstract. There is no descending of levels of abstraction. There is also no abandoning of discreteness as we explain below. Overall, in the conventional view sketched in the previous paragraph, one thinks *vertically* about phonological organization in the mind-brain versus its downstream fate in the vocal tract. In its alternative, one thinks *horizontally*.

A key implication of this alternative view is the abstract but also actual nature of phonological organization. As argued elsewhere and reviewed in our Background section, there is evidence that the grammar is involved in the arrangement of spatio-temporal units and specifically in ways that refer to the temporal nature of these units. Consider, for example, that in MA homorganic sequences are coordinated differently at affixal junctures than within templates (see Fig. 2). The theory required to express this systematicity must inescapably admit units with a temporal dimension (but as emphasized elsewhere not in the sense of “real” time; see Gafos 2002). A-temporal theories of phonology cannot express such generalizations.

What further good can the study of abstract organization in the vocal tract do for phonological theorizing? Defining and understanding vocal tract organization, we argue, is worth pursuing because it augments the explanatory resources for theoretical constructs needed in other areas of phonology. Consider, for example, the notion of the segment. Notational conventions for describing segments and grammar formalisms for referring to these are in widespread use. However, there is no theory of the segment (but see Byrd 1994; Garvin et al. 2018; Steriade 1993, 1994). There are, in contrast, candidate theories for the spatio-temporal units in the vocal tract (see

Guenther 1995, Saltzman 1995 and references therein as well as Sorensen and Gafos 2016 for a review of model updates since then and a more recent proposal). In stating this juxtaposition, the intention is not to hint at replacing segments by gestures and no reductionist attempt is at play. Segments are gestural constellations; that they may be necessary is neither incompatible nor antagonistic to gestures (see Gafos 2002 for an analysis that crucially invokes both) or potentially also other units of an auditory nature. Rather, the point is made to illustrate that evidence for foundational units such as the segment in phonological theory can derive from sources that take seriously spatio-temporal organization in the vocal tract and its acoustic consequences (as the attempts cited above do). Furthermore, just as there is no descending in levels of abstraction when sources of evidence are sought in vocal tract action and corresponding acoustics, there is also no (necessary) loss of discreteness in doing so. When properly understood, spatio-temporal units of vocal tract action are no less discrete than symbols. Gestures are discrete and stable in the following sense. Discreteness and qualitiveness inheres in the fact that the dynamical systems formalizing these units admit fixed points or attractors which endow the system with a non-axiomatic stability (because, in contrast to the axiomatic stability in symbolic views, reproducibility of modes in behavior, regardless of initial conditions, is a direct consequence of the formal model in dynamics).<sup>11</sup>

---

<sup>11</sup> Any theory that makes reference to sequences of speech units (and predicates over them) must necessarily make explicit how such sequences unfold in time but are stable enough to be reproducible in behavior. The problem is unrecognizable in conventional theories precisely because stability in representations is an axiom in these theories. But the challenge is fundamental to a viable theory of phonology. The tools to address this challenge do not derive from the theory of formal languages (a subfield of the math-

The view that the essential units of linguistic form are abstract and actual is not novel. As early as in the formative period of generative linguistics, Fowler et al. (1980) were explicitly concerned with “reconciling the units that a speaker/hearer is assumed to know with those that he uses,” the vision being the development of a theory in which these units are both abstract and actual, where the “know” in the above excerpt is explicitly used in the sense of “competence” (Chomsky 1965). Theoretical developments since then as well as the increasing body of work in the area of spatio-temporal coordination, enabled by the availability of techniques for obtaining quantitative records of articulator motion during speech and corresponding acoustics, encourages a renewed interest in this perspective.

**Acknowledgements** Thanks to Astrid Assmann and Nomiki Koutsoumpari for help with data measurements. Philip Hoole acknowledges support from German Research Council Grant HO3271/3-1. Adamantios I. Gafos, Jens Roeser and Stavroula Sotiropoulou acknowledge support from ERC Advanced Grant 249440 and, for AIG and SS, German Research Council Grant SFB 1287, Project C04.

## References

- Angermeyer, Philipp. 2003. Copying contiguous gestures: an articulatory account of Bella Coola reduplication. In *University of Pennsylvania working papers in linguistics*, eds. Elsi Kaiser and Sudha Arunachalam, Vol. 9, 17–30. Philadelphia: University of Pennsylvania.
- Baayen, R. Harald. 2008. *Analyzing linguistic data – A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.

---

ematical Theory of Computation, Hopcroft et al. 2007), from which Chomsky and Halle’s (1968) “Sound Pattern of English” and its lineage derive. Rather, the formal foundation for addressing this fundamental challenge is Dynamical Systems Theory, which concerns itself with recurrence in patterning over time.

- Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68 (3): 255–278.
- Bat-El, Outi. 1994. Stem modification and cluster transfer in Modern Hebrew. *Natural Language & Linguistic Theory* 12 (4): 571–596.
- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015a. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67 (1): 1–48. doi:10.18637/jss.v067.i01.
- Bates, Douglas, Reinhold Kliegl, Shravan Vasishth, and Harald Baayen. 2015b. Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.
- Benhallam, Abderrafi. 1980. Syllable structure and rule types in Arabic. PhD diss, University of Florida.
- Benhallam, Abderrafi. 1990. Moroccan Arabic syllable structure. *Langues et littératures* 8: 177–191.
- Benkirane, T. 1998. Intonation in Western Arabic (Morocco). In *Intonation systems*, eds. D. Hirst and A. Di Cristo, 345–359. Cambridge: Cambridge University Press.
- Benus, S., Smorodinsky I., and A. Gafos. 2004. Gestural coordination and the distribution of English “geminates”. In *University of Pennsylvania working papers in linguistics*, eds. Sudha Arunachalam and Tatjana Scheffler, Vol. 10, 33–46. Philadelphia: University of Pennsylvania.
- Bernstein, N. A. 1967. *The co-ordination and regulation of movements*. London: Pergamon Press.
- Bloomfield, Leonard. 1933. *Language*. New York: Holt, Rinehart and Winston.
- Bombien, Lasse, and Philip Hoole. 2013. Articulatory overlap as a function of voicing in French and German consonant clusters. *The Journal of the Acoustical Society of America* 134: 539–550.
- Boroff, Marianne. 2007. A landmark underspecification account of the patterning of glottal stop. PhD diss, Stony Brook University.
- Boudlal, Abdelaziz. 2001. Constraint interaction in the phonology and morphology of Casablanca Moroccan Arabic. PhD diss, Université Mohammed V, Facult des Lettres, Rabat, Morocco.
- Box, G. E. P., and D. R. Cox. 1964. An analysis of transformations. *Journal of the Royal Statistical Society* 26 (2): 211–252.
- Bradley, Travis G. 2002. Gestural timing and derived environment effects in Norwegian clusters. In *Wccfl 21 proceedings of the 21st west coast conference on formal linguistics*, eds. Line Mikkelsen and Christopher Potts, 43–56. Somerville, MA: Cascadilla Press.

- Bradley, Travis G. 2006. Spanish complex onsets and the phonetics-phonology interface. In *Optimality-theoretic studies in Spanish phonology*, eds. Fernando Martínez-Gil and Sonia Colina, 15–38. Amsterdam: John Benjamins.
- Bradley, Travis G. 2007. Morphological derived-environment effects in gestural coordination: A case study of Norwegian clusters. *Lingua* 117: 950–985.
- Browman, Catherine P., and Louis M. Goldstein. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica* 45 (2–4): 140–155.
- Byrd, Dani. 1994. Articulatory timing in English consonant sequences. PhD diss, UCLA. Published in UCLA Working Papers in Phonetics, 86.
- Byrd, Dani. 1995. C-centers revisited. *Phonetica* 52 (4): 285–306.
- Byrd, Dani. 1996. Influences on articulatory timing in consonant sequences. *Journal of Phonetics* 24 (2): 209–244.
- Byrd, Dani, and Susie Choi. 2010. At the juncture of prosody, phonology, and phonetics: the interaction of phrasal and syllable structure in shaping the timing of consonant gestures. *Laboratory Phonology* 10: 31–59.
- Byrd, Dani, and Elliot Saltzman. 2003. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31 (2): 149–180.
- Byrd, Dani, Sungbok Lee, Daylen Riggs, and Jason Adams. 2005. Interacting effects of syllable and phrase position on consonant articulation. *The Journal of the Acoustical Society of America* 118 (6): 3860–3873.
- Cassery, Elizabeth D. 2012. Gestures in optimality theory and the laryngeal phonology of Faroese. *Lingua* 122: 41–65.
- Catford, John Cunnison. 1977. *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.
- Catford, John Cunnison. 1988. *A practical introduction to phonetics*. Oxford: Clarendon Press.
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.
- Clements, George N. 1990. The role of the sonority cycle in core syllabification. In *Papers in laboratory phonology I: Between the grammar and physics of speech*, eds. John Kingston and Mary E. Beckman,

- 283–333. New York: Cambridge University Press.
- Davidson, Lisa. 2003. The atoms of phonological representation: Gestures, coordination, and perceptual features in consonant cluster phonotactics. PhD diss, Department of Cognitive Science, Johns Hopkins University.
- Davidson, Lisa. 2006. Schwa elision in fast speech: segmental deletion or gestural overlap? *Phonetica* 63: 79–112.
- Dell, François, and Mohamed Elmedlaoui. 1985. Syllabic consonants and syllabification in Imdlawn Tashlhiyt Berber. *Journal of African Languages and Linguistics* 7: 105–130.
- Dell, François, and Mohamed Elmedlaoui. 1988. Syllabic consonants in Berber: Some new evidence. *Journal of African Languages and Linguistics* 10: 1–17.
- Dell, François, and Mohamed Elmedlaoui. 1996. On consonant releases in Imdlawn Tashlhiyt Berber. *Linguistics* 34: 357–395.
- Dell, François, and Mohamed Elmedlaoui. 2002. Syllable structure in Moroccan Arabic. In *Syllables in Tashlhiyt Berber and in Moroccan Arabic*, Vol. 2, 227–290. Dordrecht, Netherlands, and Boston, MA: Kluwer Academic Publishers. doi:10.1007/978-94-010-0279-0\_8.
- Elmedlaoui, Mohamed. 2014. What does the Moroccan Malhun meter compute, and how. *The Form of Structure, the Structure of Form: Essays in honor of Jean Lowenstamm* 12: 139.
- Elson, Ben. 1947. Sierra Popoluca syllable structure. *International Journal of American Linguistics* 13 (1): 13–17.
- Fougeron, Cécile, and Patricia A. Keating. 1997. Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America* 101 (6): 3728–3740.
- Fowler, Carol A., Philip Rubin, Robert E. Remez, and M. E. Turvey. 1980. Implications for speech production of a general theory of action. *Language Production, Speech and Talk* 1: 373–420.
- Frank, Harry, and Steven C. Althoen. 1995. *Statistics: concepts and applications*. Cambridge, UK: Cambridge University Press.
- Gafos, Adamantios I. 2002. A grammar of gestural coordination. *Natural Language & Linguistic Theory* 20 (2): 269–337.
- Gafos, Adamantios I., Philip Hoole, Kevin Roon, and Chakir Zeroual. 2010. Variation in overlap and phonological grammar in Moroccan Arabic clusters. In *Laboratory phonology*, eds. Cécile Fougeron,

- Barbara Kühnert, Mariapaola D'Imperio, and Nathalie Valleé, Vol. 10, 657–698. Berlin, Germany, and New York, NY: Mouton de Gruyter.
- Garvin, Karee, Myriam Lapierre, and Sharon Inkelas. 2018. A Q-theoretic approach to distinctive subsegmental timing. *Proceedings of the Linguistic Society of America* 3 (9): 1–13.
- Goldstein, Louis M. 2011. Back to the past tense in English. In *Representing language: Essays in honor of Judith Aissen*, eds. R. Gutiérrez-Bravo, L. Mikkelsen, and E. Potsdam, 69–88. Santa Cruz, CA: Digital Library eScholarship Depository, Linguistic Research Center, University of California.
- Guenther, Frank H. 1995. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review* 102 (3): 594–621.
- Hall, N. E. 2003. Gestures and segments: Vowel intrusion as overlap. PhD diss, University of Massachusetts, Amherst.
- Hardcastle, William J. 1985. Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication* 4: 247–263.
- Heath, Jeffrey. 1987. *Ablaut and ambiguity: Phonology of a Moroccan Arabic dialect*. New York: State University of New York Press.
- Hoberman, Robert. 1995. Current issues in Semitic morphology. In *Handbook of phonological theory*, ed. John Goldsmith, 839–847. Oxford: Blackwell.
- Hoole, Philip, and Andreas Zierdt. 2010. Five-dimensional articulography. In *Speech motor control: New developments in basic and applied research*, eds. Ben Maassen and Pascal H. H. M. van Lieshout, 331–349. Oxford, UK: Oxford University Press.
- Hoole, Philip, Andreas Zierdt, and Christian Geng. 2003. Beyond 2D in articulatory data acquisition and analysis. In *Proceedings of the 15th International Conference of Phonetic Sciences (ICPhS)*, 265–268.
- Hopcroft, John E., Rajeev Motwani, and Jeffrey D. Ullman. 2007. *Introduction to automata theory, languages, and computation*, Vol. 3. Boston: Pearson.
- Kahn, Daniel. 1976. *Syllable-based generalizations in English phonology*. Indiana, MI: Indiana University, Linguistics Club Bloomington.
- Katz, Jonah. 2012. Compression effects in English. *Journal of Phonetics* 40 (3): 390–402.
- Keegan, John M. 1986. The role of syllable structure in the phonology of Moroccan Arabic. In *Current*

- approaches to African linguistics*, ed. G. Dimmendaal, Vol. 3, 209–226.
- Kiparsky, Paul. 2003. Syllables and moras in Arabic. In *The syllable in optimality theory*, eds. C. Féry and R. F. H. E. van der Vijver, 147–182. Cambridge, UK: Cambridge University Press.
- Kuznetsova, Alexandra, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen. 2013. *lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)*. R package version 2.0-3. <http://CRAN.R-project.org/package=lmerTest>.
- McCarthy, John J. 1984. Prosodic structure in morphology. In *Language sound structure*, eds. M. Aronoff and R. Oehrle, 299–317. Cambridge, MA: MIT Press.
- Perkell, Joseph S., Marc H. Cohen, Mario A. Svirsky, Melanie L. Matthies, Iñaki Garabieta, and Michel T. T. Jackson. 1992. Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *Journal of Acoustical Society of America* 92: 3078–3096.
- Pouplier, Marianne. 2011. The atoms of phonological representations. In *The Blackwell Companion to Phonology*, eds. Marc van Oostendorp, Keren Rice, Beth Hume, and Colin Ewen, 107–129. Oxford: Wiley-Blackwell.
- Prince, Alan, and Paul Smolensky. 2004. *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell.
- R Core Team. 2013. *R: A Language and Environment for Statistical Computing*. Vienna, Austria. R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Ridouane, Rachid, Anne Hermes, and Pierre Hallé. 2014. Tashlhiyt’s ban of complex syllable onsets: phonetic and perceptual evidence. *STUF-Language Typology and Universals* 67 (1): 7–20.
- Saltzman, Elliot L. 1995. Dynamics and coordinate systems in skilled sensorimotor activity. In *Mind as motion: Explorations in the dynamics of cognition*, ed. Timothy Van Gelder Robert F. Port, 149–173. Cambridge, MA: MIT Press.
- Saltzman, Elliot L., and Kevin G. Munhall. 1989. A dynamical approach to gestural patterning in speech production. *Ecological psychology* 1 (4): 333–382.
- Shaw, Jason A., and Adamantios I. Gafos. 2015. Stochastic time models of syllable structure. *PLoS ONE* 10 (5): 0124714. doi:10.1371/journal.pone.0124714.
- Shaw, Jason A., Adamantios I. Gafos, Philip Hoole, and Chakir Zeroual. 2009. Syllabification in Moroccan Arabic: Evidence from patterns of temporal stability in articulation. *Phonology* 26: 187–215.

- Smolensky, Paul. 2006. On theoretical facts and empirical abstractions. In *Wondering at the natural fecundity of things: Essays in honor of Alan Prince*, eds. E. Bakovic, J. Ito, and J. McCarthy, 271–291. Santa Cruz, CA: Linguistics Research Center [<http://repositories.cdlib.org/lrc/prince/13>].
- Sorensen, Tanner, and Adamantios I. Gafos. 2016. The gesture as an autonomous nonlinear dynamical system. *Ecological Psychology* 28 (4): 188–215.
- Steriade, Donca. 1993. Closure, release, and nasal contours. In *Nasals, nasalization and the velum*, eds. Marie Huffman and Loren Trigo. Vol. 5 of *Phonetics and Phonology*, 401–470. Elsevier.
- Steriade, Donca. 1994. Complex onsets as single segments: the Mazateco pattern. 51: 203–291.
- Steriade, Donca. 1999. Alternatives to syllable-based accounts of consonantal phonotactics. In *Proceedings of LP*, eds. O. Fujimura, B. Joseph, and B. Palek, 205–245. Prague: The Karolinum Press.
- Wright, Richard Albert. 1996. Consonant clusters and cue preservation in Tsou. PhD diss, University of California, Los Angeles.
- Zierdt, Andreas, Philip Hoole, and Hans G. Tillmann. 1999. Development of a system for three-dimensional fleshpoint measurement of speech movements. In *Proceedings of the 14th International Conference of Phonetic Sciences (ICPhS)*, eds. John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville, and Ashlee C. Bailey, 735–738. Berkeley, CA: Department of Linguistics, University of California.