SYLLABLE STRUCTURE AS COUPLED OSCILLATOR MODES: EVIDENCE FROM GEORGIAN VS. TASHLHIYT BERBER

Louis Goldstein[a], *Ioana Chitoran*[b] *& Elisabeth Selkirk*[c]

[a]Haskins Laboratories & Yale U., [b]Dartmouth College, [c]U. of Massachusetts Amherst
[a]`goldstein@haskins.yale.edu`, [b]`ioana.chitoran@dartmouth.edu`, [c]`selkirk@linguist.umass.edu`

ABSTRACT

A theory is presented that claims the basis for syllable structure is to be found in the modes of a system of coupled oscillators that control intergestural timing in speech. Onsets correspond to the in-phase mode and codas to the anti-phase mode. Articulatory data from Georgian and Tashlhiyt Berber are presented that support the association of onsets with in-phase mode.

Keywords: syllable structure, gesture timing, coupled oscillators, Georgian, Tashlhiyt Berber.

1. THE BASIS OF SYLLABLE STRUCTURE

The organization of phonetic units into syllables with an internal constituency of onset consonants, a nucleus (vowel), and coda consonants, is a fundamental property of the phonology of human languages. Several universal properties of phonological structure (markedness) require reference to syllable constituency, for example:

1. CV syllables are the only type found universally (although cf. recent work [2] claiming that the language Arrernte lacks syllables with onset consonants).
2. Onsets combine relatively freely with nuclei, while combination is likely to be more constrained within onsets, within codas, and between nuclei and codas.
3. Coda consonants are frequently weight-bearing (moraic) and can therefore influence metrical patterning, while onset consonants only rarely bear weight.

Syllable constituency also helps shape the phonetic properties of gestural magnitude, overlap, and variability in speech production [6], and it plays a key role in speech errors [9].

Despite its central status, there has been relatively little theorizing about the biological basis of syllable constituency. One outstanding exception is the frame-content theory [14], which posits that the source of the syllable (the 'frame') is in the oscillatory behavior of the jaw, which can be evolutionarily traced to its role in mastication. The unmarked CV structures that emerge in canonical infant babbling are seen, in this theory, as simple jaw oscillation cycles, with little or no control over individual consonant and vowel constrictions. While this theory may explain the source of units with the approximate size of syllables, it does little to illuminate the robust properties of syllable constituency, such as (1-3) above. It comes closest in the case of (1), but even here is not clear why the cyclic jaw oscillation of the infant is transcribed by adults as composed of CV rather than VC syllables, and whether this can be attributed to the infant's behavior rather or to the adult perception of that behavior.

A more recently developed theory [10], to be outlined in the next section, views syllable structure as emerging from the planning and control problem of establishing stable patterns of relative timing among phonetic units (articulatory gestures). Relating syllable structure to relative timing control makes predictions about differences in the timing of the articulatory events associated with a phonetic sequence when that sequence is syllabified differently (as for example in different languages). One of these predictions is tested in this paper by comparing the timing of comparable sequences in Georgian and Tashlhiyt Berber.

2. COUPLED OSCILLATOR MODEL OF PLANNING GESTURAL TIMING

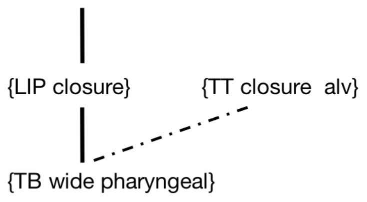
When we produce an utterance, the gestural units that compose it exhibit stable and reliable patterns of relative timing. This must be so, because differences in the relative timing of gestures are informational and can be used to distinguish utterances. For example, the English words *ban* and *mad* are composed of the same sequence of oral contrictions (lip closure—tongue body constriction—tongue tip closure), but they differ in the relative timing of the velum lowering with respect to oral constrictions. How can the relative timing of these gestural events be planned in such a way as to preserve the coherence of the informational structure, but at the same time to allow flexibility in timing as a function of speaking rate, prosodic context, and transient perturbations?

In the model developed in [10,15,16,18,20], this is accomplished by associating each gesture of an utterance with a nonlinear (limit-cycle) planning oscillator (or 'clock') that triggers the production of that gesture during speech production. The oscillators are coupled to one another in *graphs* that encode the informationally significant aspects of relative timing. This method of controlling intergestural timing can be related to generic recurrent connectionist network architectures used to control serial behavior [1, 12]. As argued in [10] an advantage of an architecture in which each gesture is associated with its own clock and in which gestures are coordinated in time by coupling their clocks is that such networks exhibit hallmark behaviors of coupled nonlinear oscillators—entrainment, multiple stable modes, and phase transitions, all of which appear relevant to speech timing. It is also argued that these behavioral phenomena form the basis for syllable structure.

2.1. Coupling Graphs and Planning

Coupling graphs are hypothesized to be part of speakers' phonological knowledge of particular word forms. In such a graph, oscillators (corresponding to gestural units) are coupled in pair-wise fashion. Each coupling link is associated with a target relative phase for that pair of oscillators. For example, Fig. 1 shows the coupling graph for the word *mad*. The oscillator for the {`LIP closure`} gesture is coupled to the vowel {`TB wide pharyngeal`} oscillator and also to the velum lowering {`VEL wide`} oscillator.

Figure 1: Coupling graph for the English word *mad*. Solid lines represent in-phase coupling, dashed lines represent anti-phase coupling.



The solid lines indicate that these oscillators are coupled with a target relative phase of 0°, or in-phase coupling. The {`TT closure alv`} oscillator is also coupled to the vowel, but its line is dashed, indicating that its target relative phase with respect to the vowel is 180°, or anti-phase.

During the planning simulation (as described in [18,20]), the nonlinear planning oscillators are set into motion at arbitrary phases, and they settle into stable patterns of relative phase due to coupling forces in the individual oscillators' equations of motion that arise from the relative phases targets for the linked pairs of gestures in the coupling graph (and their coupling strengths). In the example in Fig. 1, there is no competition in the coupling graph, and therefore the final relative phases will be identical to the target ones. At the end of planning, gestural activations forming a gestural score are determined using the stabilized oscillator time functions (with the onset of a gesture's activation at phase 0° of its planning oscillator), and the score is input to the task-dynamic model of interarticulator coordination [19] to yield motion of model speech articulators.

2.2. Intrinsic Modes and Syllable Structure

Research on coordinating rhythmic action of multiple limbs [11,21] has shown that subjects can successfully use either of two modes without any learning or training, *in-phase* and *anti-phase*. The in-phase mode has been shown to be more stable than the anti-phase mode. Other phase relations between the limbs can be produced, but only with training.

For an activity like speech that requires the coordination of multiple actions (gestures) and yet can be acquired by all members of the species with (approximately) equal facility, it is reasonable to suppose that these intrinsic modes are employed in the coordination of speech gestures. The coupled oscillator model of syllable structure hypothesizes that onset and coda represent the two possible intrinsic modes of coordinating a consonant (C) and a vowel (V) gesture, in-phase, and anti-phase respectively. Thus, in the coupling graph in Fig. 1, the lip closure gesture and the vowel gesture are specified in the coupling graph with target relative phase of 0°, which will result in their activations beginning triggered at the same time. Data on the timing of C- and V-related activity in CV syllables [13] supports this hypothesis.

As discussed in [10], the coupling mode hypothesis can provide a basis for the signature properties of syllable structure outlined in (1-3). The unmarked status of CV syllables (1) can be understood as a result of the fact that in-phase is the most stable mode of coupling. The ability of onset Cs to combine relatively freely with Vs (2) can also be understood as resulting from the stability of this mode—any action can be combined with any other action as long as the most stable mode of coordination is being employed. There is no need to learn how to produce each combination. The explanation of property (3) has been discussed in [15].

2.3. Competitive Coupling in Complex Onsets

Now let us consider onsets composed of consonant clusters. In-phase coupling with the vowel gesture defines an onset consonant, which suggests that both Cs in an onset cluster should bear that coupling relation to the vowel. However, that would also result in the Cs being synchronous with one another, which would make it likely that one or the other might not be perceptually recoverable. In addition, onset clusters in some languages (e.g., Georgian) may contrast two possible orders of consonants in an onset, which would not be possible if the Cs in an onset were all synchronous. Therefore, the coupling model has hypothesized [4] that the Cs in an onset are coupled to each other (C-C) in anti-phase mode, as well as being *both* coupled in-phase to the vowel (C-V). The result is a competitive coupling graph. The final relative phases, at the end of planning, exhibit a compromise between the competing targets [20].

The consequences of the predicted compromise have been observed in English in the timing of C and V gestures [3,6]. As Cs are added to an onset, the timing lag between the rightmost consonant and the vowel gets shorter (i.e., the consonant shifts to the right with respect to the vowel) in order to accommodate the fact that each of the consonants is also coupled in-phase with the vowel (also known as the 'c-center' effect).

3. COMPETITIVE COUPLING AS DIAGNOSTIC FOR SYLLABIFICATION

If the coupling model is correct in hypothesizing that complex onsets have a competitive coupling graph, then it should be possible to use the consequences of the competition (e.g., the rightward shift of the final C) as a diagnostic that a C sequence is syllabified as an onset. To test this, we measured (using EMMA) the timing of C gestures to the following V in words beginning with sequences of one, two or three consonants in two languages, Georgian and Tashlhiyt Berber. In Georgian, the initial sequences are analyzed as complex onsets [22]. In Tashlhiyt, complex onsets are not allowed [8], and only the rightmost of the consonants in the sequence bears the onset relation to the following vowel.

3.1. Georgian

Two speakers of Georgian produced words like those in Table 1 (frame: **Sit'q'va _ gamoit^{h}k^{h}mis ordʒer**).

Table 1: Mean lag (ms) between achievement of Tongue Tip target for /r/ and Tongue Body target for the vowel /i/.

		S1	S2
rial-i	'commotion'	85	47
k'rial-i	'glitter'	48	46
ts'k'rial-a	'shiny clean'	22	33

The time from the target attainment of the rightmost C gesture to the target attainment of the V gesture was measured, using a velocity threshold to determine target attainment (Onsets were too difficult to measure reliably, particularly for the V gestures).

Results for speaker S1 are typical of those found for other word sets and are consistent with those of syllable onsets in English. As consonants are added to the onset, the lag between the rightmost consonant and vowel becomes shorter. Speaker S2 failed to show the effect in this set. In listening to the forms produced, it was clear that this speaker produced a (somewhat low) epenthetic vowel between the /k/ and /r/ in all forms. Since the Cs do not form an onset cluster for this subject, it follows from the model that no rightward shift would be predicted, and none is observed. Both speakers exhibited epenthesis in some forms, though not always the same ones. Rightward shifts were never seen whenever there was epenthesis, but were observed elsewhere. Epenthesis was only ever observed between two Cs if C_1 had a more posterior constriction than C_2. This is consistent with the results in [7], which showed greater temporal lags between stops when they are ordered back-to-front. Similar results were obtained for complex onsets consisting only of stops.

3.2. Tashlhiyt Berber

A speaker of Tashlhiyt Berber produced words like those in Table 2 (frame: **inna ___ʁassad**). C target to V target times were measured as in Georgian.

Table 2: Mean lag (ms) between achievement of Lip Aperture target for /m/ and Tongue Body target for the vowel /u/.

mun	'accompany'	45
s-mun	'caus-accompany'	52
t-s-mun	'3fs- caus-accompany'	48

The results for Berber are strikingly different from Georgian and English. Adding additional consonants to the sequence at beginning of the word has no effect on the timing of the rightmost C to the V. This is consistent with the hypothesis that the rightward shift is diagnostic of the competitive coupling graph associated with a complex onset. In Berber, this string is not an onset and therefore it would not be predicted to exhibit the rightward shift. (/tsmun/ would be analyzed as two syllables [ts.mun], with [s] functioning as the nucleus of the first; /smun/ would be syllabified as [s.mun] with the [s] either a syllable on its own, or a coda to the preceding syllable).

4. CONCLUSIONS

The data from Georgian and Berber have several implications. First, they provide support for the general theory that syllable structure has its basis in the modes of coupled timing oscillators with distinct syllable constituents corresponding to distinct modes. More specifically, it supports the hypothesis that complex onsets have a coupling graph in which these modes are in competition. Secondly, it demonstrates that a relatively new observable, rightward shift in the timing of onset C gestures with respect to the following V gesture, could be used as a diagnostic for a complex onset. This would be a very valuable addition to the phonologist's toolkit, although obviously much more testing is required. Finally, the results are consistent with the unconventional syllabification proposed for Tashlhiyt Berber [8,17], in which onsets are restricted to single Cs and all Cs can function as syllable nuclei.

5. REFERENCES

[1] Bailly, G., Laboissière, R. and Schwartz, J. L., 1991, Formant trajectories as audible gestures: An alternative for speech synthesis. *Journal of Phonetics 19*, 9-23.

[2] Breen, G., Pensalfini, R. 1999. Arrernte: A language with no syllable onsets. *Linguistic Inquiry 30*, 1-26.

[3] Browman, C. P., Goldstein, L. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica 45,* 140-155..

[4] Browman, C.P., Goldstein, L. 2000 Competing constraints on intergestural coordination and self-organization of phonological structures, Les Cahiers de l'ICP, *Bulletin de la Communication Parlée 5*, 25-34.

[5] Byrd, D. 1995. C-Centers revisited. *Phonetica 52*, 263-282.

[6] Byrd, D., 1996. Influences on articulatory timing in consonant sequences. *Journal of Phonetics 24*, 209-244.

[7] Chitoran, I., Goldstein, L. and Byrd, D., 2002. Gestural overlap and recoverability: Articulatory evidence from Georgian. In Laboratory Phonology VII. (C. Gussenhoven and N. Warner, editors), pp. 419-448. Berlin: Walter de Gruyter.

[8] Dell, F., Elmedlaoui, M., 1985. Syllabic Consonants and Syllabification in Imdlawn Tashlhiyt Berber, *Journal of African Languages and Linguistics 7*, 105-130.

[9] Dell, G., Juliano, C., Govindjee, A. 1993. Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science 17*, 149-195.

[10] Goldstein, L., Byrd, D., Saltzman, E. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In: Arbib, M. (ed), *From action to language: The mirror neuron system*. Cambridge: Cambridge University, pp. 215-249.

[11] Haken, H., Kelso, S., Bunz, H. 1985. A theoretical model of phase transitions in human hand movements. *Biological Cybernetics 51*, 347-356.

[12] Jordan, M. I., 1986. *Serial order in behavior: A parallel distributed processing approach* (Tech. Rep. No. 8604). San Diego: University of California, Institute for Cognitive Science.

[13] Löfqvist, A., Gracco, V. 1999. Interarticulator programming in VCV sequences: lip and tongue movements. *Journal of the Acoustic Society of America 105*, 1854-1876.

[14] MacNeilage, P.F., Davis, B.L. 2000. Origin of the internal structure of word forms. *Science 288*, 527-531.

[15] Nam, H. in press. A competitive, coupled oscillator model of moraic structure: Split-gesture dynamics focusing on positional asymmetry. In Cole, J. and Hualde, J. (eds). *Papers in Laboratory phonology 9*. Berlin: Mouton.

[16] Nam, H., Saltzman, E. 2003. A competitive, coupled oscillator of syllable structure. *Proceedings of the XIIth International Congress of Phonetic Sciences*, 2253-2256.

[17] Ridouane, R. (accepted). Voiceless, vowel-less syllables in Tashlhiyt Berber: phonetic and phonological evidence. *Phonology*.

[18] Saltzman, E. and Byrd, D., 2000, Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. Human *Movement Science 19*, 499-526.

[19] Saltzman, E. L., and Munhall, K. G., 1989, A dynamical approach to gestural patterning in speech production. *Ecological Psychology 1*, 333-382.

[20] Saltzman, E., Nam, H., Goldstein, L. (submitted). Intergestural timing in speech production: the role of graph structure. *Human Movement Science*.

[21] Turvey, M. 1990. Coordination. *American Psychologist 45*, 938-953.

[22] Vogt, H. 1971. *Grammaire de la langue géorgienne*. Oslo: Universitetsforlaget.