Syllable as a synchronization mechanism

Yi Xu

Dept of Speech, Hearing and Phonetic Sciences, University College London, UK

Abstract

Despite being highly intuitive and widely recognized, syllable continues to be a controversial notion. It is argued here that a resolution may lie in recognizing that speech is a highly skilled motor activity with a core problem shared with other motor skills: how to reduce degrees of freedom (DOF) to the extent that makes its central nervous control possible. The most effective way of reducing DOF is to *synchronize* multiple articulatory movements, and the syllable serves exactly this function for speech. This synchronization hypothesis also offers resolutions to coarticulation and many other unsettled problems, and has implications for motor control in general.

Key words: synchronization, target approximation, edge synchronization, tactile anchoring, degrees of freedom

Introduction

The nature of the syllable remains a mystery to this day: "*although nearly everyone can identify syllables, almost nobody can define them*" (Ladefoged 1982: 220). Existing proposals (e.g., Browman & Goldstein 1992, MacNeilage 1998) are not able to answer some of the most basic questions about the syllable: a) *Why are there syllables*? b) *Do syllables have clear phonetic boundaries*? c) *Do segments have definitive syllable affiliations*? Without clear answers to these questions, many other issues about speech also remain unresolvable, including, in particular, coarticulation.

Syllable as a synchronization mechanism

This article is a brief introduction of a *synchronization hypothesis* that can address all three questions mentioned above. The overarching proposal is that syllable is a temporal coordination mechanism whose function is to synchronize multiple articulatory movements so as to make speaking possible. The coordination involves three basic mechanisms: *target approximation*, *edge synchronization* and *tactile anchoring*.According to this hypothesis, speech encodes information by generating variations in phonetic (segmental, tonal and phonational) properties in quick succession, which requires concurrent articulatory movements toward multiple underlying targets (*target approximation*). The central nervous control of the concurrent movements is made possible by synchronizing the onsets and offsets of the movements (*edge synchronization*) to critically reduce degrees of freedom

(DOF). And tactile sensation during the closed phase of each syllable provides alignment references for the synchronization of movements (*tactile anchoring*).

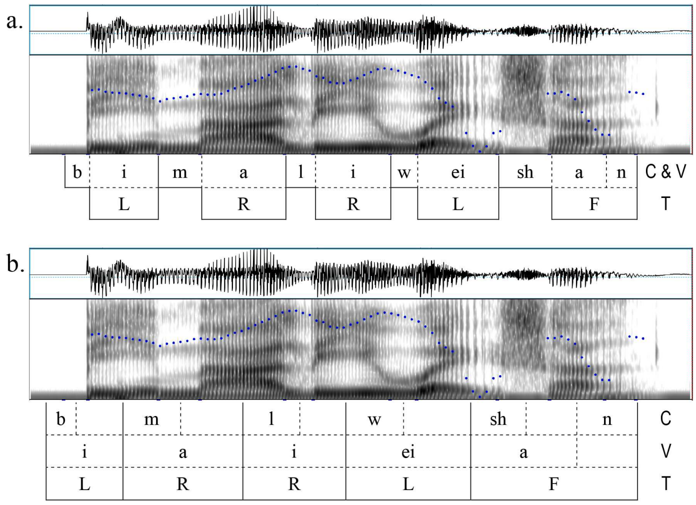

Figure 1. Spectrogram of the Mandarin phrase "比麻黎偽善" /bǐ má lǐ wěi shàn/ [more hypocritical than Ma Li], with broad phonetic transcriptions. In both panels, C, V and T stands for consonant, vowel and tone. The segmentation in a. is conventional, while that in b. is based on the synchronization hypothesis.

A direct consequence of synchronization hypothesis is a major change in the acoustic segmentation of the syllable, as illustrated in Figure 1. The conventional segmentation is shown in Figure 1a, where syllable onsets are aligned to points of abrupt spectral shift corresponding to the moments of complete oral closure (with the exception of /w/, for which there is not even a widely-agreed segmentation). The segmentation based on the synchronization hypothesis is shown in Figure 1b, where all the segmental boundaries are shifted leftward from those in Figure 1a. These *conceptual* shifts are based on the *target approximation* principle that the onset of a segment is when the spectral pattern *starts to move toward* its prototypical configuration. For /l/, for example, the onset is in the middle of the "preceding" vowel, where F1 starts to drop toward the oral closure. The leftward shift of vowel onsets is even more extensive. For example, the first /a/ now begins from the middle of the first conventional /i/ interval where F2 starts to drop toward its prototypical level; and the second /i/ begins from the middle of the conventional /a/ interval where F2 starts to rise toward its prototypical level. Thus each vowel onset is shifted leftward, across the conventional consonant interval, and well into the conventional interval of the preceding vowel.

The new *offset* of a segment is where its prototypical configuration is best approach, but not necessarily attained. For /i/, for example, it is at the peak of F2 and F3, for /a/ it is at the peak of F1 and valley of F2, and for /w/ it is at the valley of F2, where intensity is also at a minimum. For the obstruent consonants, the offset is no longer at the end of its prototypical spectral pattern (e.g., closure gap in /b/, nasal or lateral formants in /m/ and /l/, and the frication in /ʂ/), but in the middle of those intervals.Furthermore, as shown in the bottom tier of Figure 1b, the new tone interval fully coincides with that of the entire syllable, which consists of not only the initial consonant and the nuclear vowel, but also the coda consonant, as in the case of /shàn/.

Coarticulation and DOF

A major impact of the new syllable segmentation is on the understanding of coarticulation. Because the initial movement toward a vocalic target is now viewed as the vowel proper rather than its anticipation, it is no longer considered as evidence of either long-distance anticipatory V-to-V coarticulation (Öhman 1966) or local V-to-C anticipatory coarticulation. Instead, because initial consonant and the first vowel start at the same time, they are considered as fully *coproduced* or coarticulated for the duration of the consonant. Also no longer needed is the notion of carryover coarticulation. Due to inertia, an articulatory gesture toward a target has to be one of moving away from its initial state, which is the end result of approaching the preceding target. Overcoming inertia therefore necessarily carries the influence of the preceding target that is no longer being executed. So, what is carried over is only its remnant effect rather than its continued articulation.

This view of coarticulation differs from both the assimilation account and gestural overlap account (Saltzman & Munhall 1989), as illustrated in Figure 2. In panel *a* we can see that smooth surface trajectories can be generated by strictly sequential target approximation movements. In panel *b* the first movement is much shortened from that in panel *a*, resulting in an undershoot of the first target. Because the undershoot is due to a premature termination of the first movement, which is effectively *truncated* by the second movement, there is neither *assimilation* nor *anticipation*. In panel *c*, instead of truncation, the first two movements are partially overlapped via *gestural blending*. The resulting trajectory, however, is not very different from the one in panel *b*. More importantly, such a blending requires *more degrees of freedom* than sequential target approximation, as the amount of overlap and its exact location both have to be specified.

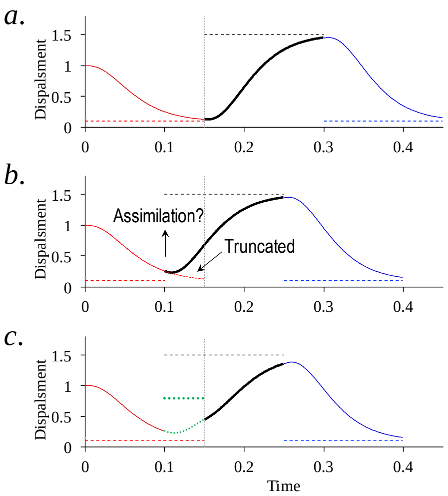

Figure 2. Sequential vs. blending target approximation. In *a*, the three movements are strictly sequential, and the vertical line divides the first two movements. In *b*, the vertical reference remains at time 0.15, but the first movement is shortened by 0.05 unit. All the movements remain sequential, so that the first movement is *truncated* by the second. In *c*, the first and second movements overlap with each other by 0.05 units. The overlap is implemented by applying a blended target (horizontal green dotted line), which is the average of the first two targets. All trajectories generated by the qTA model (Prom-on et al. 2009).

Tactile anchoring

This mechanism is the final piece that completes the syllable puzzle, because it provides solutions to three critical issues. First, it explains how synchronization is achieved: by tactile feedback during the closure phase of consonants. Second, it points out that it is the edges, rather than the center of the syllable (where sonority is the highest, cf. review by Ohala 1992), that is the essence of the syllable. Finally, it predicts that syllable onset, where a maximum number of gestures can be coproduced, is a better synchronization site than syllable offset, which in turn explains why CV is the more prevalent syllable structure than VC and CVC.

References

Browman, C.P., Goldstein, L. 1992. Articulatory phonology. Phonetica 49, 155-180.

MacNeilage, P.F. 1998. The frame/content theory of evolution of speech production. Behavioral and Brain Sciences 21, 499–546.

Mechsner, F., Kerzel, D., Knoblich, G., Prinz, W., 2001. Perceptual basis of bimanual coordination. *Nature* 414: 69-73.

Ladefoged, P. l982. A Course in Phonetics. University of California, Los Angeles.

Ohala, J.J., 1992. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In *Papers from the Parasession on the Syllable*. (eds.) Chicago: Chicago Linguistic Society**:** 319-338.

Prom-on, S., Thipakorn, B., Xu, Y. 2009. Modeling tone and intonation in Mandarin and English as a process of target approximation. JASA 125.

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