Cognitive Template for a Phonetic Correlate of Syllable Structure

Melissa A. Redford
University of Oregon, USA
E-mail: redford@darkwing.uoregon.edu

ABSTRACT

The study was designed to test a coarticulatory origin for the long-short segment duration pattern typical of syllable onset clusters. The hypothesis was that the internal members of these clusters are shortened due to coproduction with the following vowel. Specifically, the internal consonants are terminated when the downward-moving jaw makes for inefficient consonantal articulation. To test the hypothesis, native-English speakers produced nonsense words with intervocalic consonant sequences in four speech production conditions: (1) normal speech, (2) bite-block speech without auditory feedback, (3) clenched-jaw speech without auditory feedback; (4) normal speech without auditory feedback. Overall, the results show that speakers preserve consonant duration patterns that conform to intended syllable structures despite articulatory and perceptual disruptions to the system. Such results do not support or exclude a coarticulatory origin for syllable-related consonant duration patterns. Instead, they suggest that these patterns are important to the mental representation of syllables, at least in English.

1. INTRODUCTION

Segment duration varies with syllable structure [1], resulting in word-level segment duration patterns. These patterns provide important phonetic cues to syllable boundary perception in English (e.g., [2], [3], [4]). The patterns themselves, however, seem arbitrary. For instance, why is the internal member of a cluster shorter than the external member and not vice-versa? One compelling answer is that these patterns emerge from coarticulatory constraints. The long-short pattern typical of onset clusters [5] may emerge because internal consonants are coproduced with the following vowel. Whereas an external consonant is articulated during the most closed portion of the jaw cycle, an internal consonant is articulated while the jaw continues in its downward trajectory towards the vowel target [6]. At some point the jaw may be too open to sustain efficient consonantal articulation, and so the internal consonant is “truncated” by the articulation of the vowel.

Such a hypothesis assumes that syllables are units of the speech plan, but that syllable-related segment duration patterns are not explicitly encoded. Instead these duration patterns are explained to emerge from intergestural dynamics. The present hypothesis is therefore consistent with coproduction and task-dynamic theories of speech production [7], and strikingly similar to explanations that have been advanced for vowel shortening in closed syllables [8] [9].

The current study was designed to provide a strong test of the hypothesis that the long-short, onset-cluster duration pattern is due to a constraint on consonantal articulation imposed by a downward-moving jaw. Speakers produced intervocalic consonants as syllable-onset clusters or as singleton offset/onset sequences normally and with a fixed jaw. The fixed jaw conditions included a bite-block condition and a clenched-jaw condition. It was predicted that typical duration patterns would be disrupted in the fixed jaw conditions, but not in the normal speech conditions.

2. METHODS

2.1. Design

Two young-adult, native-English speakers produced intervocalic consonant sequences embedded in nonsense words under four speaking conditions: (1) normal speech; (2) bite-block speech; (3) clenched-jaw speech; and (4) a normal speech control. Auditory feedback was eliminated in conditions 2 and 3 to avoid perception-induced compensation. Condition 4 controlled for the effect of no auditory feedback on normal production. Conditions 1 and 4 were always completed first and last, respectively. Speaker 1 completed condition 2 before 3, and Speaker 2 completed condition 3 before 2. The nonsense word tokens were produced in the same random order by both speakers and in all four speaking conditions.

Auditory feedback was eliminated in conditions 2–4 by having the speakers listen to a continuous stream of pink noise. Speakers listened to the noise over headphones, and had control over its level. They were instructed to adjust the level so that they could not hear themselves speak. This instruction appeared to be effective in that the experimenter often had to indicate to the speakers during the experiment that they were speaking either too loudly or too softly.

2.2. Stimuli and Materials

A pilot study indicated that speakers may compensate for the speech production manipulations in order to preserve the duration patterns that cue syllable boundary location. To control for the same possibility in the present study, the stimuli were intervocalic obstruent+sonorant or sonorant+ obstruent sequences that were word-internal or word-peripheral. For example, the sequences /-kl-/ and /-lk-/ were internal to the disyllabic nonsense words [sɛˈklat] and [sɛlˈkat], and divided by a word boundary in the nonsense word pairs [mɛˈsɛk ˈlano] and [mɛˈsɛl ˈkano]. There were 8 nonsense words for each consonant sequence order (8 x 2 = 16), and at each of the word-internal and -peripheral locations (16 x 2 = 32). All disyllabic nonsense words and nonsense word pairs were spoken in the frame sentence "Say ___ eight times," and repeated 3 times for a total of 96 tokens per speech production condition.

The bite blocks were 15 milimeters (mm) sections of rubber belt, which was 10 mm thick and 12 mm wide. Speakers clenched two bite blocks (one for each side of the mouth) between their premolars. If the bite blocks had been clenched closer to the front of the mouth, as in other such experiments (e.g., [10], [11]), they would have directly impeded the anterior consonantal articulations under study.

2.3. Acoustic Measures

The utterances were recorded in an acoustically-insulated experiment room using a Shure BG 5.1 microphone, and saved directly into a computer. They were later displayed as oscillograms and spectrograms, and the acoustic durations of the intervocalic obstruent and sonorant were measured. Obstruent boundaries were defined by a sudden drop/rise in the amplitude of a periodic waveform and by aperiodicity. Sonorant boundaries were defined by amplitude and frequency changes in the periodic waveform on one side, and by the obstruent boundary on the other. Ambiguity in defining sonorant boundaries was resolved by repeated listening to different sections of the waveform.

3. RESULTS and DISCUSSION

The aim of the experiment was to evaluate a coproduction explanation for the long-short consonant duration pattern of syllable onset clusters. By hypothesis, jaw movement is the key contributor to this duration pattern, and so was fixed in two of the four speech production conditions. The prediction was that fixed-jaw productions of intervocalic consonant sequences would not exhibit typical duration patterns. To test this prediction, we turn first to results from the production of word-internal, obstruent-sonorant and sonorant-obstruent sequences.

3.1 WORD-INTERNAL CONSONANT SEQUENCES

The disyllabic words with intervocalic obstruent-sonorant sequences were produced so that the sequence formed an onset cluster to the second syllable. The words with intervocalic sonorant-obstruent sequences were produced so that the sonorant was an offset to the first syllable, and the obstruent an onset to the second. In normal speech, the different syllabifications are characterized by long-short and long-long or short-long pattern (*cf.* [5], [2], and [12] who show that phrase-medial, syllable-final consonants are ofter shorter than syllable-initial consonants). In the present study, these patterns were captured by a single value, namely, the duration difference between the first and second consonant (C1-C2). The typical C1-C2 duration value in normal speech will be positive when the sequence is an onset cluster, and close to zero or negative when split.

A 4-way analysis of variances (ANOVA) was performed on C1-C2 duration to test the effect of jaw movement on the consonant duration patterns. The factors were (1) Speaker, (2) Speech Production Manipulation, (3) Consonant Order, and (4) Segment Identity. The highest order significant interaction was a 3-way interaction between factors 1, 2, and 3 [$F(3, 348) = 3.97$, $p < 0.01$]. However, post-hoc tests within speaker showed that the overall pattern for the two speakers did not vary systematically with the speech production manipulation. For Speaker 1, C1-C2 durations were significantly different from normal speech (i.e., condition 1) in the bite block and control conditions. For Speaker 2, C1-C2 durations were not significantly different from normal speech in any condition, but the durations in the clenched-jaw and control conditions were significantly different from one another.

Apart from the unsystematic differences just noted, the results for the two speakers were very similar: The consonant sequences produced as onset clusters were associated with a long-short duration pattern, those that were split were associated with a short-long pattern. The interaction between speech production manipulation and consonant order was not significant, as is evident from Figure 1.

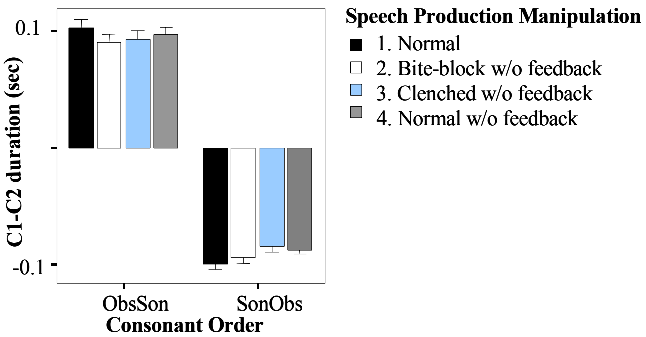

Figure 1: Consonant duration patterns of word-internal, intervocalic consonant sequences. The speech production manipulation had no effect on the duration patterns associated with the different consonant orders.

The result that speakers produce the same consonant duration patterns with a fixed-jaw as they do in normal, unconstrained speech might be explained in one of two ways. The first explanation is suggested by inspection of Figure 1. The duration difference between obstruents and sonorants appears to be the same whether obstruents are ordered first or second in the sequence. So, it could be the case that sonorants are intrinsically shorter than obstruents. If this is true, then the long-short pattern typical of onset clusters just reflects the fact that the best and most common onset clusters across languages are composed of an obstruent followed by a sonorant [13].

To examine this possibility, the difference between obstruents and sonorants durations were calculated without respect to the order in which they appeared. A within speaker analysis showed that consonant order was significant for Speaker 2 [$F(1, 173) = 20.92$, $p < 0.01$]. So, for Speaker 2 obstruent and sonorant duration differed by syllable position. This result weakens the possibility that syllable-related consonant duration patterns are explained by intrinsic consonant duration alone.

A second explanation for why the speech production manipulations failed to disrupt the consonant duration patterns is that speakers consciously preserve these patterns in production. This possibility is especially likely given that relative consonant duration provides a disambiguating phonetic cue to boundary location [2].

To examine the possibility that the speakers in the present study overcame the constraint of a fixed-jaw in order to preserve boundary information, we turn now to results from the production of obstruent-sonorant and sonorant-obstruent sequences that crossed a word boundary.

3.2 CONSONANT SEQUENCES ACROSS WORD BOUNDARIES

Speakers produced the same consonant sequences as before, but in these tokens the sequences were divided by a word boundary. This meant that the obstruent-sonorant sequences, previously syllabified as syllable-onset clusters, were produced so that the obstruent was word-final, and the sonorant word-initial. As before, the sonorant-obstruent sequences were split, but at a word boundary instead of word-internally.

Once again, a 4-way analysis of variances (ANOVA) was performed on C1-C2 duration to test the effect of jaw movement on the consonant duration patterns. In this analysis, the significant highest order interaction was not with speaker, but with segment identity [speech production manipulation x consonant order x segment identity, $F(3, 351) = 5.49$, $p < 0.01$]. Segment identity affected the duration pattern because word-final /t/ was often flapped or omitted in speech production conditions 2–4. Indeed, when the analysis was repeated with only non-zero C1 durations, the 3-way interaction disappeared [$F(3,294) = 0.65$, $p = 0.58$]. However, the 2-way interaction between speech production manipulation and consonant order, shown in Figure 2 for all C1-C2 durations, remained significant [$F(3, 294) = 13.68$, $p < 0.01$].

The figure shows, contrary to the original hypothesis, that fixed-jaw speech did not affect consonant durations patterns. Instead, the duration patterns of normal speech were different than those of any other condition when the consonant order was obstruent-sonorant. Word-final obstruents were longer than word-initial sonorants in unconstrained speech. Although one might conclude from this result that auditory feedback is important to the production of consonant duration patterns, we believe that such a conclusion is premature for the following reasons: (1) the speakers always completed the normal speech condition first; and (2) speakers reduced word-final /t/ in conditions 2–4. These two facts suggest that obstruent-sonorant duration patterns were different in the normal speech condition because speakers hyperarticulated the word-final obstruents. Once speakers were more familiar with the stimuli, these obstruents were hypoarticulated.

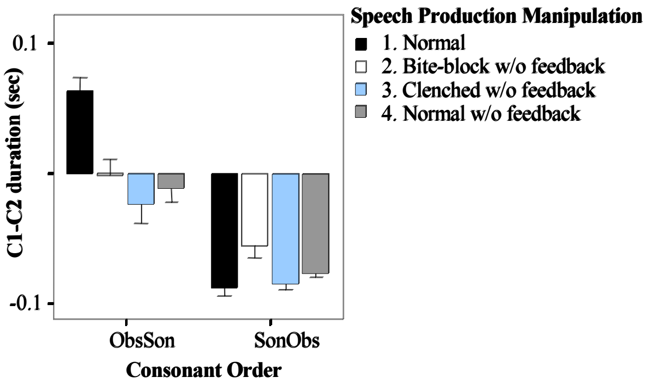

Figure 2: Consonant duration patterns of consonant sequences divided by a word boundary. The interaction between speech production manipulation and consonant order was significant, but probably due to hyperarticulation in the normal condition (see text).

Setting aside the confounding effect of hyperarticulation, we see evidence in these results for the possibility that speakers compensate for articulatory and perceptual disruptions to preserve boundary information. Whereas the duration patterns of obstruent-sonorant and sonorant-obstruent sequences are still different from one another in Figure 2, it is clear that the obstruent-sonorant pattern is not the long-short pattern typically associated with onset clusters. The results from an overall analysis within consonant order, shown in Figure 3, confirms that the duration patterns of obstruent-sonorant sequences syllabified as onset clusters are distinct from those that are split by a word boundary [$F(3, 366) = 9.08$, $p < 0.01$].

Additional analyses on obstruent and sonorant durations (4-way MANOVA) indicated that obstruent duration varied with word boundary and consonant order [$F(1,690) = 43.64$, $p < 0.01$]. Obstruents that were external members of an onset cluster were longer than singleton obstruents. Word-final obstruents were shorter than either word-initial or word-internal syllable-final obstruents. Sonorant duration varied especially with word boundary [$F(1,690) = 132.65$, $p < 0.01$]. Word-initial and word-final sonorants were longer than word-internal sonorants, whether they preceded or followed an obstruent consonant.

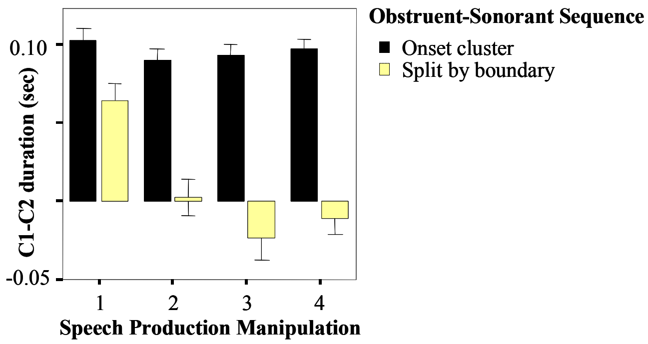

Figure 3: Consonant duration patterns of obstruent-sonorant sequences that were produced as onset clusters (dark bars) or were split by a word boundary (light bars). The duration patterns are distinct across all speech production conditions.

The analysis of obstruent and sonorant durations strongly suggest that speakers compensate for the articulatory and perceptual disruptions of the speech production manipulations to achieve duration patterns that cue boundary location. Obstruents of word-internal, intervocalic obstruent-sonorant sequences are significantly lengthened to achieve the long-short pattern typical of syllable-onset clusters. Sonorants that occur as single onsets or offsets at the edges words are significantly lengthened to achieve the long-long or short-long pattern that signals a boundary between consonants.

4. CONCLUSIONS

The results contradict the prediction that syllable-related consonant duration patterns would be disrupted in fixed-jaw speech. Instead, speakers produced duration patterns that distinguished onset clusters from singleton onset/offset sequences in spite of the speech production manipulations. From this we conclude that the segment duration patterns that are correlated with syllable structure in production [1] and provide phonetic cues to syllabification in perception [2] [3] [4] are important to the mental representation of syllables, at least in English.

The findings of the present study are inconsistent with a strict coproduction explanation for syllable-related duration patterns, which requires that such patterns emerge spontaneously from intergestural dynamics during each production. The findings do not, however, rule-out the possibility that these seemingly arbitrary patterns *originate* from coproduction constraints, and are phonologized only in languages that make a distinction between obstruent-sonorant sequences as onset clusters and as singleton onset/offset sequences. Ongoing work is testing this origins hypothesis by examining the effect of a fixed jaw on the production of intervocalic obstruent-sonorant sequences in Finnish where such sequences are produced with the typical long-short pattern, but are always syllabified—according to speakers—as a singleton onset/ offset sequence [14].

REFERENCES

[1] I. Lehiste, *Suprasegmentals*, Cambridge, MA: MIT Press, 1970.

[2] W. Christie, "Some cues for syllable juncture perception in English," *Journal of the Acoustical Society of America*, vol. 55, pp. 819-821, 1974.

[3] V. Boucher, "A parameter of syllabification for VstsopV and relative timing invariance," *Journal of Phonetics*, vol. 16, pp. 299-326, 1988.

[4] B. Tuller and J. Kelso, "The production and perception of syllable structure," *Journal of Speech and Hearing Research*, vol. 34, pp. 501-508, 1991.

[5] M. Haggard, "Correlations between successive segment durations: Values in clusters," *Journal of Phonetics*, vol. 1, pp. 111-116, 1973.

[6] M. Redford, *An Articulatory Basis for the Syllable*, PhD thesis, University of Texas at Austin, 1999.

[7] C. Fowler and E. Saltzman, "Coordination and coarticulation in speech production," *Langauge and Speech*, vol. 36, pp. 171-195, 1993.

[8] K. Munhall, C. Fowler, S. Hawkins, and E. Saltzman, "Compensatory shortening in monosyllables of spoken English," *Journal of Phonetics*, vol. 20, pp. 225-239, 1992.

[9] S. Shaiman, "Kinematics of compensatory vowel shortening: the effect of speaking rate and coda composition on intra- and inter-articulatory timing," *Journal of Phonetics*, vol. 29, pp. 89-107, 2001.

[10] C. Fowler and M. Turvey, "Immediate compensation in bite-block speech," *Phonetica*, vol. 37, pp. 306-326, 1980.

[11] B. Lindblom, J. Lubker, and T. Gay, "Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation," *Journal of Phonetics*, vol. 7, pp. 147-161, 1979.

[12] M. Redford and R. Diehl, "The relative perceptibility of initial and final consonants in CVC syllables," *Journal of the Acoustical Society of America*, vol. 106, pp. 1555-1565, 1999.

[13] A. Bell and J. Bybee-Hooper, "Issues and evidence in syllabic phonology," in *Syllables and Segments*, A. Bell and J. Bybee-Hooper, Eds., pp. 3-22. Amsterdam: North Holland Publishing.

[14] M. Redford, "The relationship between syllable structure and segment duration patterns in Finnish," Paper presented at the annual meeting of the Linguistics Society of America, Atlanta, GA, Jan. 2003.