2012

Compression Effects in English

Jonah Katz
West Virginia University, katzlinguist@gmail.com

Digital Commons Citation
Katz, Jonah, "Compression Effects in English" (2012). *Faculty Scholarship*. 1150.
https://researchrepository.wvu.edu/faculty_publications/1150

Compression effects in English

Jonah Katz[a]
CNRS – Institut Jean-Nicod
Institut d'Etude de la Cognition, Ecole Normale Supérieure

Pavillon Jardin
29, rue d'Ulm
75005 Paris
France
katzlinguist@gmail.com

Abstract

This paper reports the results of an English experiment on vowel-shortening in different contexts. The data concern compression effects, whereby, in syllables with a greater number of segments, each one of the segments is shorter than in syllables with fewer segments. The experiment demonstrates that the amount of vowel compression found in English monosyllabic words depends in part on which consonants occur adjacent to the vowel in that word, how many consonants occur, and in which position they occur. Consonant clusters drive more vowel shortening than singletons when they involve liquids, but not when they involve only obstruents. Clusters involving nasals drive shortening relative to singletons only in onset position. We suggest that the results cannot be reduced to general principles of gestural overlap and coordination between consonants and vowels, but instead require a theory with overt representation of auditory duration.

Keywords: vowel duration, compression, compensatory shortening, closed-syllable vowel shortening, c-center

The author wishes to thank Edward Flemming, Adam Albright, and Donca Steriade for extensive feedback and discussion, and Kai von Fintel and Norvin Richards for comments on an earlier draft. Thanks as well to Adam Albright, Louis Goldstein, Michael Kenstowicz, John Kingston, Jelena Krivokapic, and audiences at the MIT-UMass Meeting, the LSA Annual Meeting, the Ludwig-Maximilians-Universität Institute of Phonetics and Speech Processing, the University of Southern California, and the Université Paris Descartes Psychology of Perception Laboratory for helpful discussion. This research was made possible in part by an NSF Graduate Research Fellowship Award and an ESF European Young Investigator Award.

1.0 Introduction

This paper is concerned with the timing of segmental sequences and the factors in speech perception and production that influence that timing. Previous studies have shown that vowels tend to be shorter in syllables with more segments than syllables with fewer segments, whether those segments appear in onset or coda position (Fowler, 1983; Munhall, Fowler, Hawkins, & Saltzman, 1992; van Santen, 1992; Clements & Hertz, 1996; Shaiman, 2001 for English; Lindblom & Rapp, 1973 for Swedish; Farnetani & Kori, 1986 for Italian; Waals, 1999 for Dutch; Maddieson, 1985 for cross-linguistic survey of closed-syllable shortening). We refer to these phenomena collectively as *compression effects*.

Compression effects are of interest, among other reasons, because they have been explained both as an emergent property of independent principles of articulatory coordination (Fowler, 1981; Nam, Goldstein, & Saltzman, 2009) and as a product of constraints that directly regulate acoustic duration (Campbell & Isard, 1991; Clements & Hertz, 1996; Flemming, 2001; Lindblom & Rapp, 1973; Maddieson, 1985; Myers, 1987). As such, clarifying the situation with regard to compression may allow us to shed some light on the general properties of phonetic representations of timing, a foundational question in the field.

In particular, approaches grounded in general gestural principles and those grounded in auditory constraints make different predictions about the distribution of vowel compression. A gestural account predicts that, to the extent there are useful generalizations to be had about consonant-vowel timing that are independent of consonant manner, there will also be generalizations about vowel compression that are independent of consonant manner. In a theory where constraints on vowel duration are stated in auditory terms, however, the presence and amount of vowel shortening in a given context may depend upon the perceptual relationship between the vowel and the consonants in that context. To the extent that consonants show differences in the degree to which they mask an adjacent vowel, for instance, there may also be differences in how much vowel shortening is observed. Previous studies suggest that there may indeed be differences in vowel compression depending on the adjacent consonant, as we will see in the next section, although the evidence is not yet conclusive.

In this paper, we report the results of an English production study that examines whether and where compression effects are observed. The experiment is a pseudo-word production study that tests for vowel compression within English monosyllabic words. We limit our focus largely to vowel compression and to complexity-driven shortening because these are phenomena that can help tease apart various views on what drives compression. The finding is that patterns of compression depend on the number, manner, and syllabic position of adjacent consonants. We argue, in agreement with Marin & Pouplier (2010), that the results are not accounted for straightforwardly by general principles of gestural overlap between consonants and vowels; some additional theoretical mechanism is required to explain the data. We suggest that the additional mechanism may be perceptual in nature, and suggest an approach to testing this hypothesis.

1.1 Compression

Following Munhall *et al*. (1992), we refer to compression effects that are driven by increasing the number of segments in a string as *compensatory shortening* (henceforth CS). Because the current study examines compression in several contexts, it will be useful to introduce some terminology to describe those contexts.

We distinguish between CS driven by the addition of segments to the onset of a syllable from CS driven by coda segments: *onset CS* vs. *coda CS*. We label CS observed in the comparison of syllables that contain one (consonantal) segment at the relevant periphery of the syllable (onset or coda) to syllables that contain no segments at the relevant periphery as *simplex CS*. For example, if we observe that the vowel is shorter in a CVC syllable than in a comparable CV syllable, it would be classified as simplex coda CS. Another case is CS observed in the comparison of syllables that contain one (consonantal) segment at the periphery to syllables that contain more than one: *incremental CS*. For example, if we observe that the vowel is shorter in a CCVC syllable than a comparable CVC syllable, it would be classified as incremental onset CS. Many of the comparisons in this study examine incremental CS for pairs of items that involve the same consonant adjacent to a vowel, and differ in the presence or absence of an additional consonant to the other side of the original consonant. This includes pairs such as /nod/-/snod/ and /don/-/donz/. In cases where CS does obtain between such pairs, we say that the innermost consonant 'drives' incremental CS, as a terminological shortcut.

1.2 Previous studies

The current experiment is designed to clarify and extend the description of the acoustic manifestations of CS for vowels across a range of consonants and contexts in English. The specific questions addressed are whether CS for vowels exists in English, and whether the existence and size of CS effects differ between onset and coda position, by the number of consonants in a sequence, or by the manner of those consonants. Some of these questions have been addressed in previous studies. All findings related below pertain to vowel duration unless otherwise indicated.

Simplex coda CS, often referred to as *closed-syllable vowel shortening*, is widely attested cross-linguistically. Maddieson (1985) contains an extensive review of languages where the phenomenon has been attested. For the general question of whether CS exists in English, Fowler (1983), van Santen (1992), Katz (2008), and Marin & Pouplier (2010) report CS across a variety of contexts. Lindblom & Rapp (1973) find simplex and incremental CS in both onset and coda position in Swedish, while Waals (1999) finds coda CS in Dutch. Farnetani & Kori (1986) find simplex coda CS and limited incremental onset CS in Italian. One series of studies fails to find evidence for any compression effects in English (Crystal & House 1982, 1988, 1990). This discrepancy is explained by the methodological critique of van Santen (1992).

There is less agreement in the literature on whether there exist differences in CS by syllable position or consonant manner. Fewer studies have examined these questions, and there may well be interactions between the two dimensions, i.e., manner-dependent differences between onset and coda.

Most studies agree that obstruents drive vowel CS in onset position, while studies disagree as to whether obstruents do so in coda position. Clements & Hertz (1996) report simplex CS for segments in what they call 'the extended nucleus' of an English syllable, which includes voiced transitions preceding and following the vowel, following glides, and following liquids. Munhall *et al*. (1992), Byrd (1995),[1] and Shaiman (2001) all find that obstruents drive incremental coda CS; Katz (2008) and Marin & Pouplier (2010) find no effect.

Even the facts involving onset obstruents are not entirely clear. This is because clusters of onset obstruents in English necessarily consist of /s/ followed by a voiceless stop. Comparing these sequences to singleton aspirated stops is not a straightforward test of CS, because the two sequences differ in both number of phones *and* presence of acoustic aspiration. Marin & Pouplier (2010) include aspiration as part of vowel duration and conclude that these sequences induce CS for vowels. In gestural frameworks, however, the glottal abduction gesture associated with voicelessness and aspiration is either anchored to a consonantal gesture (e.g. Browman & Goldstein 1992) or treated as an independent gestural component in a consonant cluster (e.g. Goldstein *et al*. 2006); it is thus unclear why the acoustic interval associated with this gesture (aspiration) should form part of the unit which is expected to shorten due to compression (the vowel). A more conservative test for CS in this context would use as a baseline only the voiced portion of the vowel following an aspirated stop.

Some studies report differences in CS across different manners of consonant. Van Santen (1992) finds differences in preceding vowel duration depending on the following consonant; for instance, /r/ is preceded by extremely short vowels. He also finds a small but significant incremental onset CS effect (about 10 ms) for obstruent-liquid clusters alone. Farnetani & Kori (1986) report incremental onset CS in Italian driven by obstruent-/r/ clusters but not obstruent-obstruent clusters. Waals (1999) finds that, in onset position in Dutch, CS for consonants disproportionately affects higher-sonority segments relative to lower-sonority ones. For vowels, she finds incremental coda CS with all consonant types. For phonologically long vowels, the effect is largest for those preceding liquids, intermediate with /n/, and smallest with obstruents. Katz (2008) finds finds that /l/ drives incremental CS in both onset and coda position, but obstruents do not in either position, if aspiration is not treated as part of the vowel. /n/ drives incremental CS in onset position, but the effect varies between subjects and is only marginally significant. Marin & Pouplier (2010), in the most comprehensive study published on the topic, find that all onset consonants drive articulatory C-center effects and incremental CS, and that the addition of a voiceless stop drives shortening of /Vl/ sequences in coda position, but finds no incremental CS driven by obstruents or /m/ in coda position. These results suggest, like the other two English studies, that incremental coda CS differs by consonant manner. However, vowel duration is not measured separately in the liquid series, due to the great difficulty of drawing a boundary between a vowel and following /l/; this means that the observed effect could be entirely due to shortening of the /l/ rather than the vowel.

---

[1] To clarify: Byrd finds a small shortening effect in two-obstruent clusters relative to singleton obstruents; the statistical significance of this comparison is not reported. Byrd also finds a significant *lengthening* effect on the preceding vowel from three-obstruent codas relative to two-obstruent codas. These comparisons, however, are inherently confounded with morphological differences, because sequences such as /sks/ and /skt/ are unambiguously polymorphemic in English. Polymorphemic coda sequences are elongated relative to monomorphemic ones (Sugahara & Turk 2009). For this reason, three-consonant clusters are not considered here.

Furthermore, comparisons within Marin & Pouplier's (2010) liquid series are confounded with differences in voicing; the study compares singleton voiced sonorants to clusters involving voiceless stops. It is well known that vowels preceding voiceless stops are far shorter than those preceding voiced stops in English (Peterson & Lehiste 1960), and that this difference contributes to the perception of the voicing contrast in coda position (Denes 1955). This means that the differences between singleton and cluster stimuli in this condition may be entirely due to an independently observed durational asymmetry used to enhance a phonological contrast, rather than having any bearing on sub-phonemic timing principles *per se*. Put differently, vowel shortening driven by the addition of a voiceless obstruent following a liquid is a necessary but not sufficient condition for demonstrating CS effects in this context. The same objection may hold for *onset* stimuli involving /l/ and voiceless stops, although less is known about whether and how duration of a following vowel contributes to the perception of a preceding voicing contrast.

If /l/ truly is an exception to a general lack of incremental coda CS for vowels, it is an interesting question whether this exception extends to the other English liquid, /r/, or whether it is a *sui generis* property of /l/. This could also shed light on whether the peculiar durational properties associated with /l/ are due to the large gestural and temporal differences between /l/ in onset and in coda position (Sproat & Fujimura 1993); such differences have not been reported for /r/.

Finally, we know very little about whether these patterns generalize across vowels. All of the studies testing incremental coda CS examined a single vowel; the Marin & Pouplier study, which is our most abundant source of prior data, considers only a single vowel in each consonant series. Although there are no theoretical reasons why we would expect patterns of CS to differ across different vowels, we do know that different vowels have different properties with regard to inherent duration (some of which contribute to phonological contrasts), and it seems prudent to at least check that there are no large differences between vowels in 'compressability'.

One further factor that hasn't received much attention in the previous literature is the question of which portion of the speech stream is acoustically compressed in complexity-driven CS. The answer may be important because it can help us uncover what is driving the acoustic effect: if transitions between segments are shortening, for instance, it suggests global shortening of articulatory gestures (possibly corresponding to an increase in *stiffness* in some articulatory frameworks); if only the relatively steady parts of segments are shortening, on the other hand, it would be consistent with increased overlap or truncation of gestures.

The current study attempts to shed further light on patterns of CS, in particular those that have been understudied or led to conflicting findings in the previous literature. We examine voiced obstruents-liquid clusters in onset and coda position, to ensure that any putative incremental CS effects are not driven by the voicing contrast. We use acoustic landmarks to segment the speech stream into sub-phonemic regions that can then be tested for shortening, in order to confirm that any putative differences across manners of consonant and syllable positions really do pertain to *vowel* shortening. We separate the duration of aspiration from that of vowels, to perform a more conservative test of CS for onset stops. Finally, we consider /r/ in addition to /l/, and consider multiple vowels for each consonantal series, attempting to ensure that our results generalize.

The specific hypotheses to be tested are as follows:

(H1) Vowels are shorter in the presence of preceding/following singleton consonants than in comparable sequences without preceding/following singleton consonants.
(H2) Vowels are shorter following onset clusters than following comparable singletons.
(H2a) The presence of this pattern differs by consonant manner.
(H2b) The presence of this pattern differs by vowel.
(H3) Vowels are shorter preceding coda clusters than preceding comparable singletons.
(H3a) The presence of this pattern differs by consonant manner.
(H3b) The presence of this pattern differs by vowel.
(H4) Shortening is present primarily on segmental steady states, rather than transitions.

2 Methods

2.1 Materials

The ‘target’ materials consisted of every phonotactically legal combination of the vowels {ɪ, ɑ, o} with: the consonants{r, l, n, ∅} in onset and coda position; /p/ in onset position; /s/ in coda position; the clusters {br, gl, sn, sp} in onset position; and the clusters {rb, lb, nz, sp} in coda position. Each item contained a ‘fixed’ consonant /d/ at the opposite edge of the syllable/word from the one being manipulated. The number of logically possible combinations was 54. Three of these (/dɪ/, /dɪr/, and /dɪrb/) were phonotactically illegal in English. This left a total of 51 target syllables/words. Of these, 24 corresponded to existing English words (if the slang word *diss/dis* is counted); the remaining 27 were nonce-words.

[TABLE 1 HERE]

Wherever possible, items were assigned orthographic representations that do not correspond to English words. The only exceptions were *rode*, *lode*, *don*, possibly *diss* (meaning ‘disrespect’) and possibly *doh* (an exclamation of dismay associated with Homer Simpson). Some of the words unavoidably were assigned unusual or ambiguous orthographic representations. The pronunciation of nine such words was demonstrated to subjects before the experiment.

In addition to the target items, 39 filler words were included in the reading session. These were also monosyllables, with different consonants and vowels than the target items, including some consonant clusters. *Freave*, *skay*, and *jeg* are examples of filler words used in the experiment.

The experiment included 17 target carrier sentences and 13 filler carrier sentences. The target carrier sentences were nine syllables long, of the form [[*X*] [*Y the Z W*]], where: *X* is a trochaic first name; *Y* is a past tense monosyllabic verb; *Z* is the target item; and *W* is a four-syllable modifier, beginning with a preposition and containing one noun. *Thomas bought the dore at a yard sale* and *Dustin got the snid off of E-bay* are examples of target sentences used in the study. The filler sentences varied in length, syntactic structure, and illocutionary force. They included questions, statements of opinion, and commands. *The yeam is poisonous, right?* and *This jutch wouldn’t be a bad thing to buy* are examples of filler sentences used in the study.

The filler items and sentences were used in order to decrease the rhythmic repetitiveness of the experiment. The target words were all monosyllables and were embedded in carrier sentences that were nearly identical in prosodic terms. Reading many such sentences in sequence might reasonably be supposed to induce a level of rhythmic evenness that is not generally characteristic of natural speech. Rhythmic evenness, in turn, is likely to alter the durational characteristics of segments, which is the focus of the study. The filler sentences used here were prosodically, syntactically, and semantically diverse; the phonological makeup of the nonce-words in these sentences were also quite different from the target words.

There were 90 total experimental items (target and filler) to pair with 30 carrier sentences; each experimental block of 30 trials included one third of the experimental items and each carrier sentence. Pairings of experimental item and carrier sentence were randomized, as was the order of trials inside each block of 30. The randomized sentences were presented to subjects on a computer screen. They were asked to 'read each sentence in as natural a manner as possible' before pressing a button to move to the next sentence. They were given the opportunity to take breaks after each block of 90. There were 4 repetitions of each experimental item for a total of 360 utterances. The experiment ran between 30 and 45 minutes for all subjects.

2.2 Subjects

Subjects were 6 native speakers of American English, 4 female, 2 male, all between 21 and 31 years old. None reported being diagnosed with any speech, reading, or hearing disorders. Three were from Massachusetts; the other three from New York, North Carolina, and Minnesota.

2.3 Recording

Subjects were recorded in a sound-attenuated booth inside the MIT phonetics laboratory. They were outfitted with a head-mounted condenser microphone. The utterances were recorded in mono at 44.1 kHz and saved to .aiff files.

2.4 Measurement

The problem of segmentation is a vexing one. It is clearly difficult to find a set of objective criteria for drawing a boundary between vowels and liquids, for instance. The same uncertainty also arises in cases that would seem to be relatively clear, such as vowel-obstruent boundaries. Because segments are coarticulated, there is almost never in principle a clear point in the signal where one segment ends and the next begins.

The solution adopted here is in the spirit of the *phone-and-transition* model advocated by Hertz (1991) though it differs in some details. In a sequence of two segments, the acoustic signal was segmented into the steady state of the first segment, the transition between the two segments, and the steady state of the second segment. Each boundary was selected using a particular acoustic landmark or combination of landmarks. Even this model is an idealization; there was often no clear point in the signal where a formant switched from some slope to no slope, for instance. The experimenter identified a short portion of the signal as containing the boundary; within that portion, exact boundary selection was guided by the Praat (Boersma & Weenink) formant tracker. Although we refer to some of the measured intervals as 'steady states', this is not intended as a claim that acoustic parameters remained completely flat inside these intervals. The term simply refers to the portion of the acoustic stream delimited by points where the slope of some acoustic parameter either approaches zero, suddenly and visibly gets shallower, or reverses direction.

The recordings were annotated for duration by hand using the Praat software (Boersma & Weenink). For all words, the fixed consonant (always /d/) measurement in onset position extends from an abrupt drop in (or cessation of) energy in the preceding schwa to the onset of periodic voicing in the vowel of the target word; for codas, from an abrupt drop in (or cessation of) energy to just after the following release burst. In cases where /d/ was realized as a tap, the offset was marked after the abrupt drop in energy and formants around the tap. The vowel steady-state region extended from the innermost edge of the fixed consonant /d/ to the first point where acoustic values begin to slope noticeably toward characteristic values for the target consonant. For target /r/, the value of F3 was used in all cases. For the other consonants, a combination of F1, F2, and large changes in overall spectral energy were used, depending on the particular string in question. For instance, most boundary marks pertaining to /ɪ/ were placed with reference to F2, while most boundary marks pertaining to /ɑ/ were placed with reference to F1. The C-V transition measurement extended from the vowel steady-state to the steady-state of the adjacent consonant. Consonant steady-states were defined in terms of low F3 for /r/; transitions from periodic to aperiodic noise for obstruents; combinations of F1, F2, and large changes in energy for /l/ and /r/, depending on the particular string in question. Full details about all boundary criteria are provided in Appendix A.

The objective for the segmentation strategy was to delimit intervals that are comparable across items that differ in the number of target consonants. Although the boundaries may not correspond to any psychologically real boundary between two symbols, they at least give us an acoustic landmark that can be compared across items. If we find that the interval of vowel with relatively steady F1 in /glɑ/ is shorter than that in /lɑ/, for instance, it entails that there is incremental onset CS for /l/. On the other hand, this boundary won't be strictly comparable to the one in /rɑ/, where the comparable boundary tracks F3 rather than F1. If we find that the intervals delimited by such boundaries differ in duration, the most we can say is that the interval of vowel with steady F3 in /rɑ/ is shorter than the interval of vowel with steady F1 in /lɑ/. Similarly, the marked boundaries in /lɑ/ are not strictly comparable to those in /lɪ/, which track F2, or in /p^{h}ɑ/, which track aspiration. For this reason, we limit our investigation to differences between items within the same consonantal series, and higher-order differences between these differences.

2.5 Analysis

Separate models were constructed for each of two dependent variables: duration of the steady-state vowel and duration of the CV/VC transition.[2] The data were analyzed with linear mixed effects regression models. This type of model offers several advantages over the repeated measures ANOVA models that are common in speech and language research (Quené & van den Bergh, 2004; Baayen, Davidson, & Bates, 2008). The models were implemented using the lme4 package (Bates 2007) in the statistical environment R.

Statistical analysis proceeded by a hierarchical backward elimination procedure. The process began with a 'baseline' model that included a separate fixed effect for each item in the

---

[2] A third model examined the duration of consonant steady states. Those data are not reported here for reasons of space, but are consistent with compensatory shortening.

experiment. Such a model corresponds to a theory of temporal coordination in which each lexical item (or perhaps bigram) is stored in memory with its own idiosyncratic timing pattern, and there are not necessarily any useful generalizations to be made about similarities in timing between words with similar segments. This is an extremely weak hypothesis, in the sense that it makes fewer predictions than a theory that includes equivalence classes (e.g. segments, features, cues) internal to lexical items. Successively stronger theories were then tested by removing parameters or blocks of parameters from the model. This corresponds to modifying our hypothesis to include ever more general equivalence classes. Checking how much these removals decrease model fit tells us how much empirical coverage we lose by strengthening the hypothesis.

The baseline model included *random effects* of subject identity and carrier sentence. The model included *fixed effects* of two kinds: level-defining effects that were manipulated to create the different experimental conditions; and normalizing effects that attempted to control for differences in speech rate, prosodic structure, allophony, and any other phenomena that might differ between utterances. Note that this distinction into two types of fixed effect is purely for expository purposes; the variables were treated exactly the same by the model.

The normalizing effects pertained to several different aspects of the materials. Lexical status ({word, non-word}) and frequency (natural logarithm of values from the CELEX database)[3] pertained to the familiarity of each item. Trial (how far along in the experiment the item was uttered) pertained to possible changes in speech rate, familiarity, and concentration as the experiment progressed. For items in the onset condition, the allophonic status of word-final fixed /d/ ({flapped, non-flapped}) pertained to speech rate and prosodic phrasing,[4] as did the duration of the fixed consonant in both onset and coda positions. Two further fixed effects were related to allophonic properties of VC words discussed in section 3.1.1.

The level-defining effects were vowel ({ɪ, ɑ, o}), C1 quality ({rhotic, lateral, nasal, obstruent}), syllabic position of the target consonant(s) ({onset, coda}), and number of target consonants ({0, 1, 2}). The baseline model, then, included all 4-way interactions between these variables. Removing higher-order interactions from the model generalizes across classes of item, creating a stronger hypothesis. The statistical significance of the higher-order interactions is a metric of how much we've damaged the empirical coverage of our hypothesis by making it more general.

We followed Baayen (2008) in testing the significance of fixed effects with Markov chain Monte Carlo (MCMC) sampling. Baayen *et al*. (2008) give a detailed description of this procedure. Non-significant fixed effects were removed level by level if they included a term for number of target consonants. These were the parameters that tested whether CS is present and whether it

---

[3] Two English words included in this study, *rid* and *diss*, have no listing in CELEX. They were assigned the mean log frequency of the other existing English words in the experiment. This solution was adopted because we don't believe that these words are vanishingly rare. We suspect that the omission of *rid* is some type of an editing or compilation error, and that *dis/diss* is either too recent a coinage or too rare in written language to appear in CELEX.

[4] This variable was coded 0 if there was a visible or audible burst in the realization of /d/, 1 otherwise; this may not correspond exactly to intuitions about what is and is not a flap, but it is at least a concrete and replicable criterion.

varies from one context to another. All fixed effects were retained if they did not include a number-of-consonants term; even if these effects are non-significant, retaining them can only increase the accuracy of the estimated CS effects. After each step, MCMC simulation was repeated for the reduced model.

The significance of random effects was assessed with a likelihood-ratio test. After the fixed effects in the model had been reduced by the procedure described above, by-subject random slopes were tested.

In what follows, fixed effects are reported with an effect size and p-value from MCMC sampling. Random effects are reported with an effect size, chi-squared statistic, and p-value from the likelihood-ratio test.

Before statistical analysis, the data were centered around 0 and normalized with a z transformation for each subject. This transformation characterizes data points by how many standard deviations they lie above (positive values) or below (negative values) a subject's mean. Effect sizes, then, are in standard deviations; in the text, they are translated back into a range of ms values for ease of comprehension. These ms values represent the range obtained by multiplying the z-transformed effect size by the smallest and largest subject standard deviations.

3 Results
3.1 Simplex CS
3.1.1. VC syllables

The VC syllables in the experiment were realized with substantial variation pertaining to the presence and nature of a glottal constriction at the beginning of the item. Some tokens included a realization of *the* as /ði/, with a modally-voiced transition between /i/ and the target vowel; other tokens included full glottal closure following a schwa in *the*, with near-immediate modal voicing of the target vowel upon release; most tokens fell in between these two extremes. For instance, some tokens included a creaky-voiced transition between the two vowels. In some cases, this was preceded or punctuated by fairly long closures; in some cases glottal pulses were irregular but more or less continuous. Illustrative examples of various realizations are shown in figure 1.

[FIGURE 1 HERE]

This variability raises the question of what 'counts' as vowel duration, both psychologically and for practical purposes. Investigating various metrics of duration is useful for what it reveals about the timing of these items and for discovering the most consistent metric to use in comparisons with other items. Three different metrics were investigated:

- **m1**: only the portion of the vowel with modal voicing and steady formants
- **m2**: the portion of the vowel with steady formants, regardless of glottalization
- **m3**: the entire portion of the vowel with visible formant structure

The third metric produced more consistent results than the other two. Data for m3 had the smallest standard deviations and the most similar means across different types of realization, indicating less variability. This is despite the fact that the absolute numbers for m3 are the largest of the three metrics. As a purely practical matter, m3 was adopted as the measure for VC items in all further statistical modeling. The lesser variability under this metric will make it easier to compare these items to others in the experiment.

The nature of the relationship between duration and onset quality in these items is also of theoretical interest, however. Comparisons within this class of item reveals something about the nature of CS. VC materials were split into five rough classifications of onset quality, corresponding to the tokens in figure 1 (a-c, e-f). The classes were defined with reference to whether the transitions contained only modally-voiced formant transitions (1a, referred to as *glide*), visibly creaky formant transitions (1b, *creaky transition*), closure and following creaky vowel (1c, *creaky steady state*), creaky portion with transitions and followed by steady formants (1e, *2-part creak*), or only closure (1f, *stop*). Figure 2 shows the average duration of each part of the VC syllables by class.

[FIGURE 2 HERE]

The most obvious pattern here is that modal, steady-state vowel duration is shorter when it is preceded by formant transitions (right 3 bars), longer when it is preceded by closure (left 2 bars). This suggests that the transitions count at least partially as vowel duration in whatever sense is relevant to temporal coordination.

3.1.2 Comparison to CVC syllables

For all comparisons, vowel steady states in CV and VC words were significantly longer than in CVC words, confirming H1. This is shown below. The effects ranged in size from 0.56 standard deviations (25-35 ms) for /od/ vs. /pod/ to 2.79 sds (122-175 ms) for /da/ vs. /dar/. All effects were significant below the $p = 0.0001$ level.[5]

[FIGURE 3 HERE]

To compare the size of the simplex CS effect in various contexts, we focused on items with /r/. This is because boundary marks between /r/ and vowels track F3 movement in all items, so landmarks are comparable across word-pairs. By this test, the simplex CS effect was not significantly different across vowels, nor across onset and coda position. There was one significant 3-way interaction term: the effect was much larger in coda than in onset position for the vowel /ɑ/, relative to the vowel /o/ (46-65 ms greater difference between onset and coda for /ɑ/; $p < 0.0001$).

The magnitude of the simplex CS effect differed between subjects (although the direction of the effect did not), and adding that variation to the model resulted in a significantly better fit: $\chi^2(2) = 26.3$; $p < 0.0001$. The size of the effect differed for various subjects by up to 34 ms from the mean effect, but all subjects displayed simplex CS. Subjects also differed significantly with regard to the relative size of the simplex CS effect in onset and coda position ($\chi^2(4) = 25.9$; $p < 0.0001$). Three subjects showed more shortening from VC to CVC items, two showed more

---

[5] An anonymous reviewer suggests that the comparison of CVC syllables to VC syllables with glottal stop epenthesis is not really a test of CS, because the glottal stop can be considered an onset. The statistical analysis was rerun excluding the VC tokens with glottal stops; all effects were still highly significant, if not quite as large as in the previous model.

shortening from CV to CVC items, and one subject showed no difference between onset and coda. If there are differences in simplex CS depending on context, they vary in their direction and presence from subject to subject, unlike the main effect.

3.2 Incremental CS

Patterns of incremental CS differ by consonant manner, and they differ between onset and coda for some consonants. These findings relate to H2, H3, and their sub-hypotheses. Patterns are shown in the boxplot below, which compares CVC words to comparable CCVC or CVCC words.

[FIGURE 4 HERE]

Note that none of the interactions between incremental CS and vowel quality came out significant. This means that, broadly speaking, patterns of CS do not differ between vowels. These data thus provide no support for H2b or H3b. Including by-subject random slopes for incremental CS did not significantly improve the model: for all variables representing incremental CS effects, χ^2 statistics ranged from 2 to 9 on 7 Df; $p > 0.3$. This means that subjects did not differ with regard to incremental CS.

Laterals and rhotics drive significant incremental CS in onset position (liquids: 11-15 ms; rhotics: 16-22 ms; $p < 0.01$ for both), as seen in the leftmost and rightmost panels in the upper row of figure 4. There is even more incremental CS in coda position: 9-13 ms more for laterals, 2-3 ms more for rhotics (left and right panels, bottom row). When the distinction between incremental CS with laterals and rhotics is collapsed, creating the class ‘liquids’ (the difference between the two is not significant), the onset-coda asymmetry is significant: 8-11 ms; $p < 0.05$.

Nasals drive incremental CS in onset position (2nd panel, upper row). It is not significantly different from the CS observed for laterals in onset position (1-2 ms difference between nasals and laterals). There is a small incremental CS effect for nasals in coda position (3-5 ms, 2nd panel, bottom row), which is not significant. In conjunction with the results for liquids, this confirms H3a.

/p/ in /spVd/ is followed by a shorter steady-state vowel than /p^h/in /p^hVd/ (3rd panel, top row). The effect is significantly larger than the onset effect for /l/ (14-20 ms larger than /l/; $p = 0.0046$). The effect is reversed in coda position for /dVs/ and /dVsp/ (3rd panel, bottom row), producing a significant interaction between number of consonants, obstruent manner, and syllable position (29-42 ms difference between /s/ in coda position and /p/ in onset, $p < 0.0001$). This coda effect, 4-5 ms in magnitude, is not significantly different from zero.

3.3 Other effects

Several other effects besides those related to the experimental hypotheses were present in the data. Words ending in /d/ had vowels that were significantly shorter when the /d/ was flapped (8-11 ms, $p < 0.0001$). This is presumably an effect of increased speech rate or smaller prosodic junctures, both of which could lead to shorter vowels and make flapping more likely.

Vowels in VC items that were preceded by creaky transitions were significantly longer than those that were not (29-41 ms, $p < 0.0001$). This reflects the issues with metric m3 discussed in section 3.1. The results may suggest that only part of the preceding formant transitions ‘count’ as vowel when compared to a word with initial glottal closure.

There was a significant acclimation effect over the course of the experiment: vowels got shorter by 0.03-0.04 ms in every successive item in the experiment, on average ($p < 0.0001$). This would average out to a shortening of about 3-4 ms. between successive utterances of a single item. Subjects differed in the presence/absence and magnitude of this effect. Including this variation significantly improved the fit: $\chi^2(6) = 30.2$; $p < 0.0001$.

Some items are homophonous with existing English words. Neither the existence of a homophonous word nor frequency differences between existing homophones had a significant effect on steady-state vowel duration.

3.4 Transition effects

A separate model investigated how the duration of the transition between vowel and adjacent consonant changes depending on syllable structure and consonant manner. /$\mathrm{p^h}$Vd/ and /spVd/ words were excluded from this model, because their transitions (aspiration and formant transitions, respectively) are not comparable to one another. Data are shown below.

[FIGURE 5 HERE]

Transitions in CVCC words are shorter than their counterparts in CVC words by less than 2 ms on average. This effect is not significant, confirming H4. The (lack of a) shortening effect does not interact significantly with syllable position or vowel quality. There is one significant interaction involving consonant quality and shortening: the transitions between /s/ and the adjacent vowel show significantly more shortening from /dVs/ to /dVsp/ words than the other consonant manners show (7-11 ms. more shortening, $p = 0.0027$).

Subjects do not differ significantly for any transition effects.

Transitions have a tendency to be longer in coda position than in onset position. This effect differs by vowel, however, and is not observed for /ɑ/. The onset/coda asymmetry is significant for /o/: 8-13 ms difference, $p = 0.002$. It is even larger for /ɪ/: 16-26 ms larger difference, $p < 0.0001$. The difference is reversed for /ɑ/: transitions are somewhat longer in onset position, but not significantly so.

Of the other effects examined, only acclimation was significant: transitions got shorter by 0.01-0.02 ms in every successive item in the experiment, on average ($p < 0.05$). This averages out to a about 1-2 ms. between successive utterances of a single item.

4 Discussion

4.1 Summary of results

Returning to our experimental hypotheses, H1 is confirmed: all consonants drive simplex CS in both onset and coda position. Examining the segment with the most consistent measurement criteria across contexts, /r/, codas do not appear to drive significantly greater CS than onsets.

The incremental CS results for steady-state vowels, on a first pass, are summarized as follows:

[TABLE 2 HERE]

This partially confirms H2 and H3: vowels do have a general tendency to be shorter adjacent to clusters than adjacent to singletons. We have no evidence that this pattern differs by vowel (H2b and H3b). The presence or absence of incremental coda CS does differ by consonant, confirming H3a. There is no clear evidence for analogous asymmetries in onset position (H2a). Recall, however, that the onset obstruent items are not minimal pairs; it is not obvious what the best comparison is. While the period of *steady-state* vowel is shorter in /spVd/ than in /p^{h}Vd/, /spVd/ also contains a period of formant transitions into the vowel that /p^{h}Vd/ does not. When that period is included in the vowel measurement, there is a small effect in the opposite direction, which does not reach statistical significance (6-9 ms, $p = 0.079$). Because aspirated stops don't occur as the second consonant in English clusters, and voiceless unaspirated stops don't occur as singleton onsets, this is the best comparison we can manage, but it is not a straightforward singleton-cluster pair. Comparing the duration of *modally-voiced* vowels, the results are as follows:

[TABLE 3 HERE]

Not reflected in the table is one finding about the magnitude of incremental CS: liquids induce a slightly larger incremental shortening effect in coda than in onset position, particularly /l/.

For transitions, results were somewhat more variable. In general, transitions do not shorten between singleton and cluster words, which tends to confirm H4. There is one exception to this, /dVs/-/dVsp/. Transitions tend to be longer in coda position than in onset position. This last pattern shows idiosyncratic reversals across vowels, however, and should be interpreted with caution. The design of the experiment does not allow for completely and strictly controlled comparisons between boundaries in onset and coda position; indeed, there may not exist such a design.[6]

4.2 Comparison with previous studies

The majority of the results that can be directly compared to previous studies replicate those studies. The existence of incremental coda CS for obstruents, however, has been the subject of conflicting findings in previous studies; we briefly review the situation here.

---

[6] Because previous studies tend to test duration of vowel and transition together, we conducted a second statistical analysis using this measure. The results, qualitatively, match those from separate consideration of steady states and transitions: where and only where one component shows a significant shortening effect, the combined measure also does.

The finding that obstruents do not condition incremental coda CS is in agreement with Marin & Pouplier (2010), but conflicts with the findings of Munhall *et al.* (1992) and Shaiman (2001): they found that there is a small incremental coda CS effect for obstruent clusters. Byrd (1995) found a small shortening effect, but neither the magnitude nor a statistical comparison is reported. The current results may also be taken as a contradiction of Fowler's (1983) similar findings; that study, however, reported no statistical analysis and made no distinctions by consonant manner.

There are several possible reasons why the findings might differ. Marin & Pouplier (2010) suggest that the discrepancy may be due to their use of real words exclusively; the current experiment, however, used nonsense monosyllables that were a mix of real and fake words, and obtained the same result. A second possibility is that the speech elicited in the other studies was different from the Marin & Pouplier and the current study. In particular, it was probably more rhythmically constrained, due to the repetition of a single carrier sentence in the Munhall *et al.* and Shaiman studies, and the repetition of words to a metronome beat in the Fowler study. Both Marin & Pouplier and the current study used a variety of carrier phrases, which may help 'break up' rhythmic patterns that extend across sentences. An extra-linguistic, task-specific constraint enforcing isochrony might lead to extra compression effects beyond those encountered in less isochronous speech. A third possibility is that the subjects in the other studies did indeed show the relevant effect, but the effect doesn't generalize to the population of English speakers. Fowler's data is based only on utterances from the author. Munhall *et al.* test three speakers, while Shaiman tests five; but these authors perform separate analyses of variance for each subject. Neither of these procedures allows one to generalize results to the broader population of English speakers; the statistical issues are discussed in length by Max & Onghena (1999). Marin & Pouplier and the current study, in contrast, treat subject as a random variable and attempt to generalize across levels of this variable. In essence, this explanation says that the shortening effects found in these experiments would have been meaningless if between-subject variability were taken into account.

A final possibility is that there is a very small incremental effect in obstruent clusters that Marin & Pouplier (2010) and the current study were unable to detect. The effects in the Munhall *et al.* study are generally rather small (the largest is 36 ms but most comparisons are on the order of 3-10 ms) and vary between subjects. Standard deviations are also extremely small: reconstructing from the standard error terms given in the paper, they seem to be on the order of 5-25 ms. This is a fraction of the variance observed in the current experiment, which would make small effects easier to detect. This difference in variance is presumably due to the rhythmic factors mentioned above.

4.3 Implications

In this section, we ask what factor or factors may account for the asymmetries in vowel compression reported here. We argue that, in addition to principles of gestural coordination and overlap, a proper analysis of the data will require perceptual constraints (whose effects may, of course, be instantiated by altering patterns of gestural overlap). We briefly sketch such an account.

The *emergent* approach to compression treats it as an epiphenomenon of articulatory gestural coordination (Fowler, 1981 *et seq.*; Browman & Goldstein, 1990 *et seq.*; Nam *et al.*, 2009). This class of theories includes no mechanism for actively controlling the acoustic duration of a vowel, for instance. Rather, there is a small set of articulatory gestural coupling relations as primitives, and facts about acoustic duration emerge from those gestural relationships. Essentially, shortening happens when part of an articulatory gesture is acoustically masked by an overlapping gesture.

This framework offers an elegant and simple account of acoustic shortening effects. It explains durational patterns as a consequence of independently-attested gestural principles; there is no additional mechanism for regulating auditory/acoustic duration. In the articulatory phonology framework, for instance, vowel compression will obtain when a series of consonants compete to be coupled to an adjacent vowel in phase (so the onsets are aligned), as in onset position. The fact that compression effects are not limited to onset position means that either patterns of acoustic shortening are not limited to cases of competitive phase alignment, or that competitive coupling (corresponding to the pattern known as the C-center effect) is not limited to onset position, but is present for certain consonants in coda position. Either way, we will require some addition or revision to the hypothesis that patterns of *general* consonant-vowel gestural alignment explain patterns of acoustic compression. The results reported here suggest that such an additional mechanism is indeed required.

One key finding in this regard is that liquids in coda position drive incremental compression of an adjacent vowel in coda position. This result would not be predicted by the hypothesis that consonants are coupled sequentially, rather than competitively in a C-center configuration, in coda position. One possibility is that the hypotheses regarding gestural overlap are correct, but acoustic compression corresponds to a different phenomenon, namely shortening or truncation of gestures. If this were the case, however, we would expect every part of the vowel's acoustic signature to be shortened. However, we found that these particular compression effects pertain to the acoustic steady state of the vowel, and we have no evidence that the vowel-consonant transition is shortening as well. De Jong (1991) offers several articulatory criteria for attempting to distinguish global shortening from local overlap. Marin & Pouplier (2010) show that the acoustic rime shortening they find in /Vl/-stop sequences (which may be related to the effects shown here) is accompanied by a leftward shift of the /l/ tongue-tip constriction relative to the singleton item; this may in principle be consistent with either greater overlap between the vowel and /l/ or global shortening of one or both segments.

Regardless of whether the acoustic shortening is due to gestural overlap or gestural shortening, the question is why it affects some sequences but not others. Marin & Pouplier (2010) suggest that the answer may lie in the extent to which a following consonant perceptually masks a preceding vowel: due to the more vowel-like leading gesture of coda /l/ (Sproat & Fujimura 1993), it perceptually masks the preceding vowel less. The pressure from parallel transmission to compress gestures here, then, is unopposed by recoverability concerns, and we find compression.

We concur with this general conclusion, and note that it can be straightforwardly incorporated into the approach involving auditory constraints discussed in section 1. This view emerges from investigations of closed-syllable vowel shortening (henceforth CSVS). In this phenomenon, vowels in closed syllables are observed to be shorter than vowels in open syllables. CSVS, then, is a specific kind of compression effect. The most frequent analysis of this pattern, whether explicit or implicit, is that it involves conflict between duration targets for smaller units such as segments and larger units such as moras, rimes, or syllables. Maddieson (1985), after arguing that CSVS is widespread enough to be considered a near-universal, suggests that the phenomenon may itself be an argument for treating the syllable rime as a unit of timing. Myers (1987) invokes this trading approach in a phonological analysis of English closed-syllable vowel shortening. Fujimura's (1987) model, although not explicitly concerned with compression, characterizes the tradeoff between syllable duration and segment duration in precisely this manner. Flemming (2001) makes use of weighted constraints to characterize competing pressures on segment and syllable duration. The exact tradeoff between elongating a syllable and compressing a segment is expressed by assigning weights to each constraint. This approach is part of a larger class of theories that view segment duration as wholly- or partially-dependent upon the constraints imposed by a higher-level duration frame (Kozhevnikov & Chistovich, 1965; Lehiste, 1970; Witten, 1977; Campbell & Isard, 1991; Campbell, 2000; *inter alia*). In the particular version sketched here, the grammar favors long segments for perceptibility; it also enforces a duration target on larger units, to foster rapid and efficient communication (Liberman *et al*. 1967, Mattingly 1981) and at least a tendency towards isochrony. More isochronous syllables may facilitate perception by inducing strong temporal expectations (Quené & Port, 2005).

The general idea behind this approach can be captured with the metaphor of fitting malleable objects into a malleable container. Each object and the container has some inherent volume, its volume when not acted upon by external forces. As the number of small objects inside the container increases, the size of the objects and the size of the container come into conflict. We must either compress the small objects, or stretch the container, or both. This can result in compression of all the objects in the container.

In this approach, the duration of lower-level and higher-level linguistic units is directly manipulated by the grammar.[7] Compression results from goal-oriented constraints: preserving segmental perceptibility, on the one hand, and relatively fast or evenly-timed syllables, on the other.[8] In contexts where vowel recoverability is relatively robust to overlap with an adjacent consonant, the pressure from recoverability constraints to maintain a long vowel will be less stringent, and more shortening will be observed. This general property allows the framework to

---

[7] An anonymous reviewer suggests that these constraints need not be part of the grammar, but may emerge 'in a self-organizing way'. In the absence of criteria for deciding which compression effects are attributable to grammatical constraints on gestural coordination, and which are due to non-grammatical principles, however, this seems needlessly complex.

[8] An anonymous reviewer suggests that the relevant higher-level constraints may be imposed by articulatory efficiency of jaw movement. This would have the effect of compressing clusters that obey the sonority sequencing principle more than those that violate it, along the lines suggested by Lindblom (1983) and Redford (2004 *et seq.*). This may be a promising approach, although without further elaboration it can not explain the results for clusters involving nasals. Articulation of these clusters should be affected by jaw dynamics in much the same way as obstruent clusters, yet the compression facts are different.

capture the connection between consonant manner and presence of vowel shortening, provided that the perceptual facts do in fact mirror patterns of vowel compressability. Formally, the model would be identical to Flemming's (2001) weighted constraint system, with the additional proposal that evaluation of the segmental duration constraint depends on the duration of a segment itself as well as the amount of information about that segment contributed by adjacent segments (such as the amount of information about a vowel provided by an adjacent liquid as opposed to an adjacent obstruent).

If overlap with an obstruent, for instance, harms vowel recoverability more than a comparable amount of overlap with a sonorant, it points to an explanation of why incremental vowel compression appears adjacent to sonorants in the current study but not obstruents. The greater magnitude of vowel compression adjacent to coda as opposed to onset liquids may be explained by perceptual asymmetries between liquids in the two positions; this may in turn be driven by the temporal asymmetry in coordination of the tongue-body and tongue-tip gestures for /l/ mentioned above. However, it is not known if the same temporal asymmetry holds for onset and coda /r/ in English. Whatever the source, this approach predicts that vowels should be easier to recover from coda than from onset liquids, corresponding to the asymmetry in production observed in these contexts. Similarly, this approach predicts that overlap with a preceding nasal consonant must harm vowel recoverability less than overlap with a following nasal. This may be due to anticipatory nasalization (Beddor 1993, Krakow 1999). Predictions of these kind can and should be tested in perceptual experiments.

The perceptual theory predicts that the auditory duration of a vowel is actively regulated by the grammar, but makes no particular prediction about the gestural means used for this shortening. Another area for future research is to determine directly which types of compression are accounted for by overlap and which by gestural shortening. This would require paired articulatory and acoustic data for a wide range of consonants and clusters in onset and coda position. Such an investigation would help establish with greater certainty the facts on the ground for which any theory of compression must account.

The picture that emerges of acoustic compression in English involves differences by consonant manner. We've argued that, regardless of the articulatory implementation of this shortening, it is best accounted for at some level by constraints on the perception and recoverability of segments. This is interesting because it is a case where the details of speech production seem seem to call for an analysis that makes crucial use of constraints that involve facts about perception. It thus contributes to the idea that both articulatory and perceptual facts must interact in a fairly intricate way in the grammar (Nearey 1997).

Appendix A Segmentation criteria used in the study

The table below lists the acoustic criteria used for segmentation. The columns represent the boundaries between vowel proper and transition, transition and C1, C1 and C2 in cluster items, and C1 and the adjacent word in singleton items, respectively. Abbreviations are *high plateau* (HP), *low plateau* (LP), *onset* (on), *offset* (off), *abrupt rise in energy above the 1st formant* (ER), *abrupt drop in energy above the 1st formant* (ED).

[FIGURE A1 HERE]

Table captions

Table 1 Phonetic and orthographic representations of the words elicited in the experiment.

Table 2 Presence of incremental CS effect for steady-state vowel as a function of C1 manner and syllable position; measurement criterion excludes formant transitions of /sp/ clusters.

Table 3 Presence of incremental CS effect for steady-state vowel as a function of C1 manner and syllable position; measurement criterion includes formant transitions of /sp/ clusters.

Figure Captions

Figure 1. Utterances of ‘the odd’ illustrating variability in the VC condition. a) modally-voiced transition. b) creaky-voiced transition. c) closure followed by creaky onset. d) intermittent glottal pulsing followed by creaky onset. e) creaky transition and creaky steady state followed by modal voicing. f) full glottal stop.

Figure 2. Duration, portion-by-portion, of VC syllables, separated into onset classes. ‘Trans2’ refers to creaky steady state; ‘Trans1’ refers to a transition with moving formants.

Figure 3. Average steady-state vowel duration for zero and one target consonants, across subjects and vowels, in standard deviations from the mean. The leftmost item in each plot is the item with no target consonant. ‘Lat’ = lateral, ‘Nas’ = nasal, ‘Obs’ = obstruent, ‘Rho’ = rhotic, ‘on’ = onset, ‘co’ = coda. Inside each plot, the boxes indicate the inter-quartile range (IQR), the range between the first and third quartile. The solid dot indicates the median. The whiskers indicate the range, up to 1.5 times the IQR away from the median. Open dots outside the whiskers lie more than 1.5 times the IQR away from the median and are potential outliers.

Figure 4. Average steady-state vowel duration across subjects and vowels, in standard deviations from the mean. Each plot represents one manner of consonant in onset or coda position; the left bar in each plot represents duration in the singleton item, the right bar duration in the cluster item. ‘Lat’ = lateral, ‘Nas’ = nasal, ‘Obs’ = obstruent, ‘Rho’ = rhotic, ‘on’ = onset, ‘co’ = coda. For instance, the left and right bars inside the box labeled ‘co’ and ‘Rho’ show mean durations for /Vr/ and /Vrb/ items, respectively. Inside each plot, the boxes indicate the inter-quartile range (IQR), the range between the first and third quartile. The solid dot indicates the median. The whiskers indicate the range, up to 1.5 times the IQR away from the median. Open dots outside the whiskers lie more than 1.5 times the IQR away from the median and are potential outliers.

Figure 5. Average transition duration across subjects, in standard deviations from the mean. Each plot represents one manner of consonant in onset or coda position; the left bar in each plot represents duration in the singleton item, the right bar duration in the cluster item.

References

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.

Bates, D. (2008). lme4: An R package for fitting and analyzing linear, nonlinear and generalized linear mixed models.

Beddor, P. (1993). The perception of nasal vowels. In M. Huffman & R. Krakow (eds.), *Nasals, nasalization, and the velum* (171-196). New York: Academic Press.

Boersma, P. & Weenink, D. Praat: doing phonetics by computer. Software program.

Broselow, E., Chen, S., & Huffman, M. (1997). Syllable weight: convergence of phonology and phonetics. *Phonology,* 14, 47-82.

Browman, C., & Goldstein, L. (1986). Towards an Articulatory Phonology. *Phonology,* 3, 219-252.

Browman, C., & Goldstein, L. (1988). Some notes on syllable structure in Articulatory Phonology. *Phonetica* , 45, 140–155.

Browman, C., & Goldstein, L. (1990). Tiers in Articulatory Phonology, with some Implications for Casual Speech. In J. Kingston & M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech* (341-397). Cambridge, UK: Cambridge University Press.

Browman, C., & Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica,* 49, 155-180.

Browman, C., & Goldstein, L. (1995). Gestural Syllable Position Effects in American English. In Bell-Berti & Raphael (eds.), *Producing Speech: Contemporary Issues* (19-33). New York: AIP Press.

Byrd, D. (1995). C-Centers revisited. *Phonetica*, 52, 263-282.

Campbell, N. (2000). Timing in Speech: a multi-level process. In Horne (ed.), *Prosody: Theory and Experiment* (281-334). Dordrecht: Kluwer academic publishers.

Campbell, N., & Isard, S. (1991). Segment durations in a syllable frame. *Journal of Phonetics*, 19(1), 37-48.

Clements, G., & Hertz, S. (1996). An integrated model of phonetic representation in grammar. In Lavoie and Ham (eds.), *Working Papers of the Cornell Phonetics Laboratory No. 11* (43-116). Ithaca, NY: CLC Publications.

Crystal, T., & House, A. (1982). Segmental durations in connected speech signals: Preliminary results. *Journal of the Acoustical Society of America,* 72(3), 705-716.

Crystal, T., & House, A. (1988). Segmental duations in connected-speech signals: Current results. *Journal of the Acoustical Society of America,* 83(4), 1553-1573.

Crystal, T., & House, A. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America,* 88(1), 101-112.

Denes, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America,* 27(4), 761-764.

Farnetani, E., & Kori, S. (1986). Effects of syllable and word structure on segmental durations in spoken Italian. *Speech Communication*, 5 (1), 17-34.

Flemming, E. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology,* 18 7-44.

Fowler, C. (1981). A Relationship between Coarticulation and Compensatory Shortening. *Phonetica,* 38, 35-50.

Fowler, C. (1983). Converging Sources of Evidence on Spoken and Perceived Rhythms of Speech: Cyclic Production of Vowels in Monosyllabic Stress Feet. *Journal of Experimental Psychology,* 112(3)*,* 386-412.

Fujimura, O. (1987). A Linear Model of Speech Timing. In Channon & Shockey (Eds.), *In Honor of Ilse Lehiste* (109-124). Dordrecht: Foris Publications.

Goldstein, L., D. Byrd & E. Saltzman. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In Arbib (Ed.), *From Action to Language: The Mirror Neuron System* (215-249). Cambridge, UK: Cambridge University Press.

Hertz, S. (1991). Streams, Phones, and Transitions: Toward a New Phonological and Phonetic Model of Formant Timing. *Journal of Phonetics,* 19, 91-109.

Honorof, D., & C. Browman. (1995). The center or edge: how are consonant clusters organised with respect to the vowel? In Elenius and Branderud (eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences* (552-555). Stockholm: KTH and Stockholm University.

De Jong, K. (1991). An articulatory study of consonant-induced vowel duration changes in English. *Phonetica*, 48, 1-17.

Katz, J. (2008). English compensatory shortening and phonetic representations. Presented at *Consonant Clusters and Structural Complexity*, Institute of Phonetics and Speech Processing, Munich, July-August 2008.

Kozhevnikov, V., and Chistovich, L. (1965). *Speech: Articulation and Perception*. Washington, DC: US Dept. of Commerce.

Krakow, R. (1999). Physiological organization of syllables: a review. *Journal of Phonetics,* 27, 23-54.

Lehiste, I. (1970). *Suprasegmentals*. Camrbidge, Mass.: MIT Press.

Liberman, A.M., Cooper, F.S., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74, 431–461.

Lindblom, B. (1983). Economy of speech gestures. In MacNeilage (ed.), *The Production of Speech* (217-246). New York: Springer-Verlag.

Lindblom, B., & Rapp, K. (1973). *Some Temporal Regularities of Spoken Swedish*. Publication 21 of Papers in Linguistics from the University of Stockholm (PILUS).

Maddieson, I. (1985). Phonetic cues to syllabification. In Fromkin (ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged* (203-221). NY: Academic Press.

Marin, S., & Pouplier, M. (2010). Temporal Organization of complex onsets and codas in American English: Testing the Predictions of a Gestural Coupling Model. *Motor Control*, 14, 380-407.

Mattingly, I.G. (1981) Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver and J. Anderson (eds.), *The Cognitive Representation of Speech,* (415-420). North Holland Publishing Company.

Max, L., & Onghena, P. (1999). Some issues in the statistical analysis of completely random and repeated measures designs for speech, language, and hearing research. *Journal of Speech, Language, and Hearing Research,* 42, 261-270.

Munhall, K., Fowler, C., Hawkins, S., & Saltzman, E. (1992). “Compensatory Shortening” in monosyllables of spoken English. *Journal of Phonetics*, 20, 225-239.

Myers, S. (1987). Vowel shortening in English. *Natural Language and Linguistic Theory,* 5, 485-518.

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure. In Pellegrino, Marisco, & Chitoran (eds.), *Approaches to phonological complexity* (297-328). Berlin: Walter de Gruyter.

Nearey, T. (1997). Speech perception as pattern recognition. *Journal of the Acoustical Society of America,* 101 (6), 3241-3254.

Peterson, G.E. & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America,* 32, 693-703.

Quené, H., & van den Bergh, H. (2004). On Multi-Level Modeling of data from repeated measures designs: A tutorial. *Speech Communication*, 43(1-2), 103-121.

Quené, H. & Port, R. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica,* 62(1), 1-13.

Redford, M. (2004). Origin of consonant duration patterns. In Agwuele, Warren, & Park (eds.), *Proceedings from the 8th Annual Texas Linguistics Society Conference* (54-61). Somerville, Mass.: Cascadilla Press.

Shaiman, S. (2001). Kinematics of compensatory shortening: the effect of speaking rate and coda composition on intra- and inter-articulatory timing. *Journal of Phonetics*, 29, 89-107.

Sugahara, M., & Turk, A. (2009). Durational correlates of English sublexical constituent structure. *Phnoology,* 26, 477-524.

van Santen, J. (1992). Contextual effects on vowel duration. *Speech Communication,* 11, 513-546.

Waals, J. (1999). *An Experimental View of the Dutch Syllable*. Utrecht University PhD Dissertation.

Witten, I.H. (1977). A flexible scheme for assigning timing and pitch to synthetic speech. *Language and Speech,* 20(3), 240-260.

Zhang, J. (2004). The role of language specific durational patterns in contour tone distribution. In B. Hayes, R. Kirchner, & D. Steriade (eds.), *Phonetically-Based Phonology* (157-190). Cambridge, UK: Cambridge University Press.

Table 1

		Onset		*Coda*	
Manner	**Consonants**	**IPA**	**Orth.**	**IPA**	**Orth.**
Null	**0**	ɑd, od, ɪd	od, oad, idd	dɑ, do	dah, doh
/l/	**1**	lɑd, lod, lɪd	lod, lode, lidd	dɑl, dol, dɪl	dall, dole, dil
	2	glɑd, glod, glɪd	glod, gload, glid	dɑlb, dolb, dɪlb	dalb, dolb, dilb
/r/	**1**	rɑd, rod, rɪd	rodd, rode, ridd	dɑr, dor	dar, dore
	2	brɑd, brod, brɪd	brod, brode, brid	dɑrb, dorb	darb, dorb
/n/	**1**	nɑd, nɪd, nod	nodd, nid, noad	dɑn, dɪn, don	don, dinn, doan
	2	snɑd, snɪd, snod	snod, snid, snoad	dɑnz, dɪnz, donz	donz, dinz, doanze
obstruent	**1**	pɑd, pɪd, pod	podd, pid, poad	pɑd, pɪd, pod	podd, pid, poad
	2	spɑd, spɪd, spod	spod, spid, spoad	spɑd, spɪd, spod	spod, spid, spoad

Table 2

	Onset	Coda
Obstruent	Y	N
Nasal	Y	N
Liquid	Y	Y

Table 3

	Onset	Coda
Obstruent	N	N
Nasal	Y	N
Liquid	Y	Y

Figure 1a

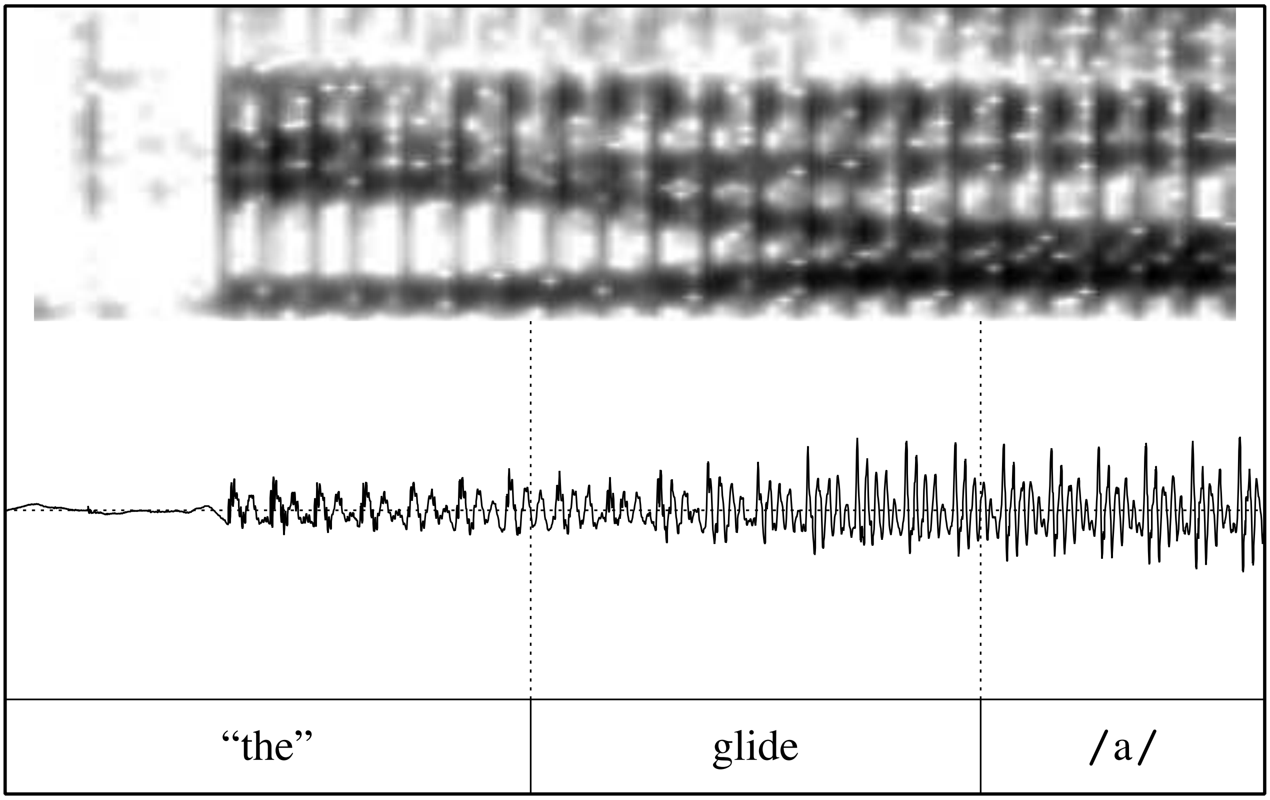


Figure 1b

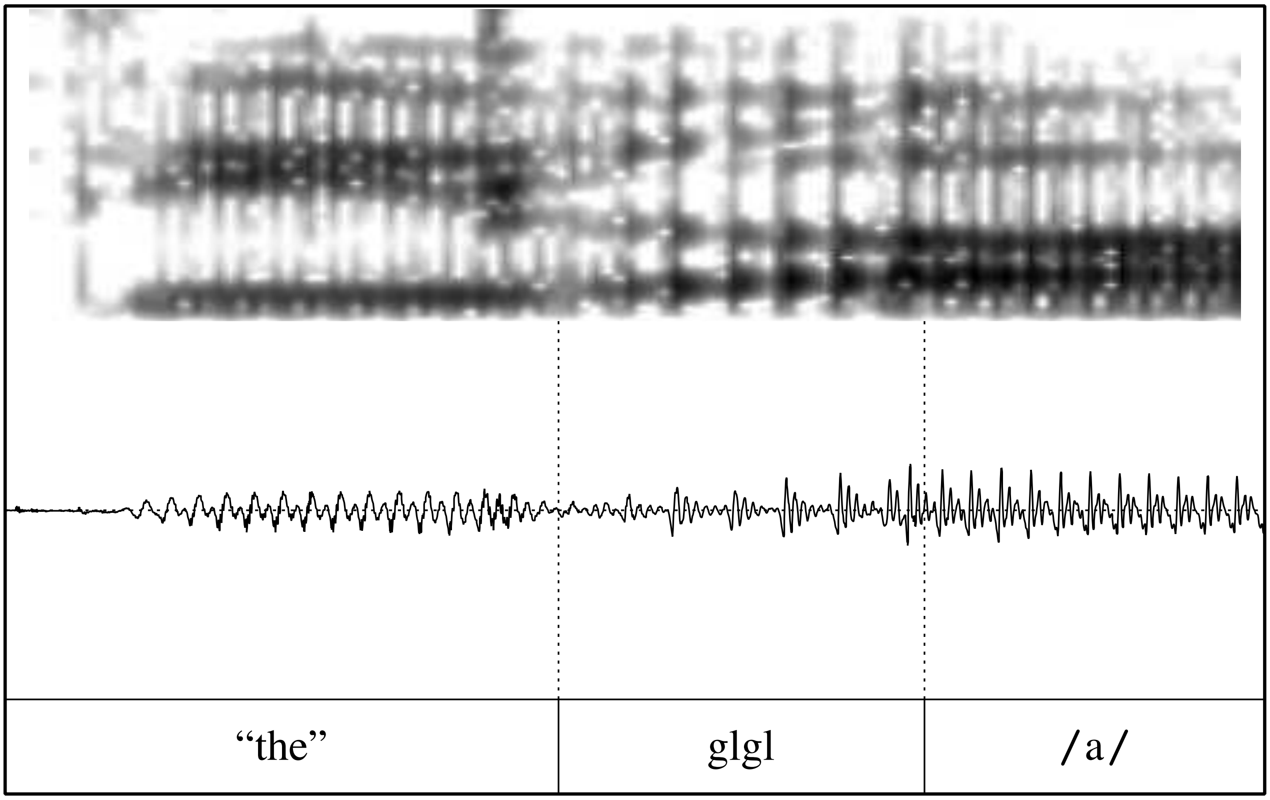


Figure 1c

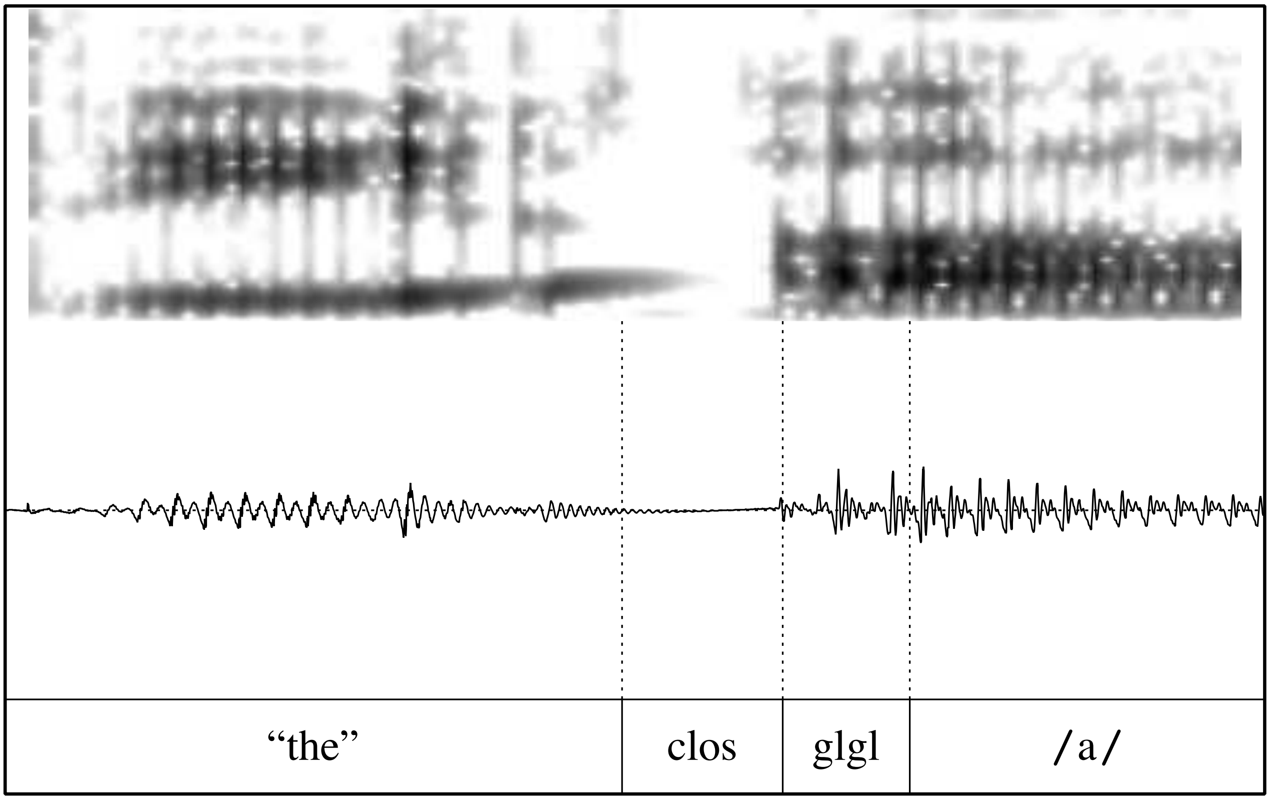


Figure 1d

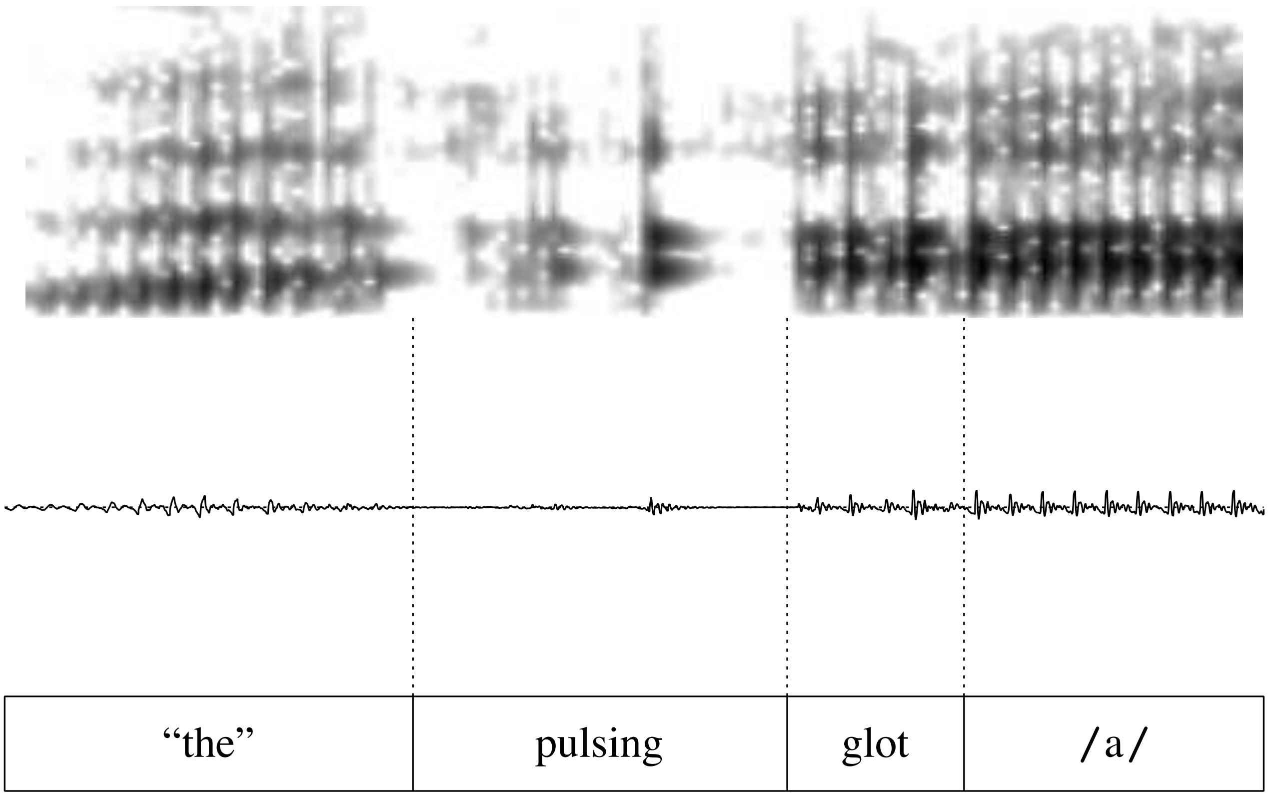


Figure 1e

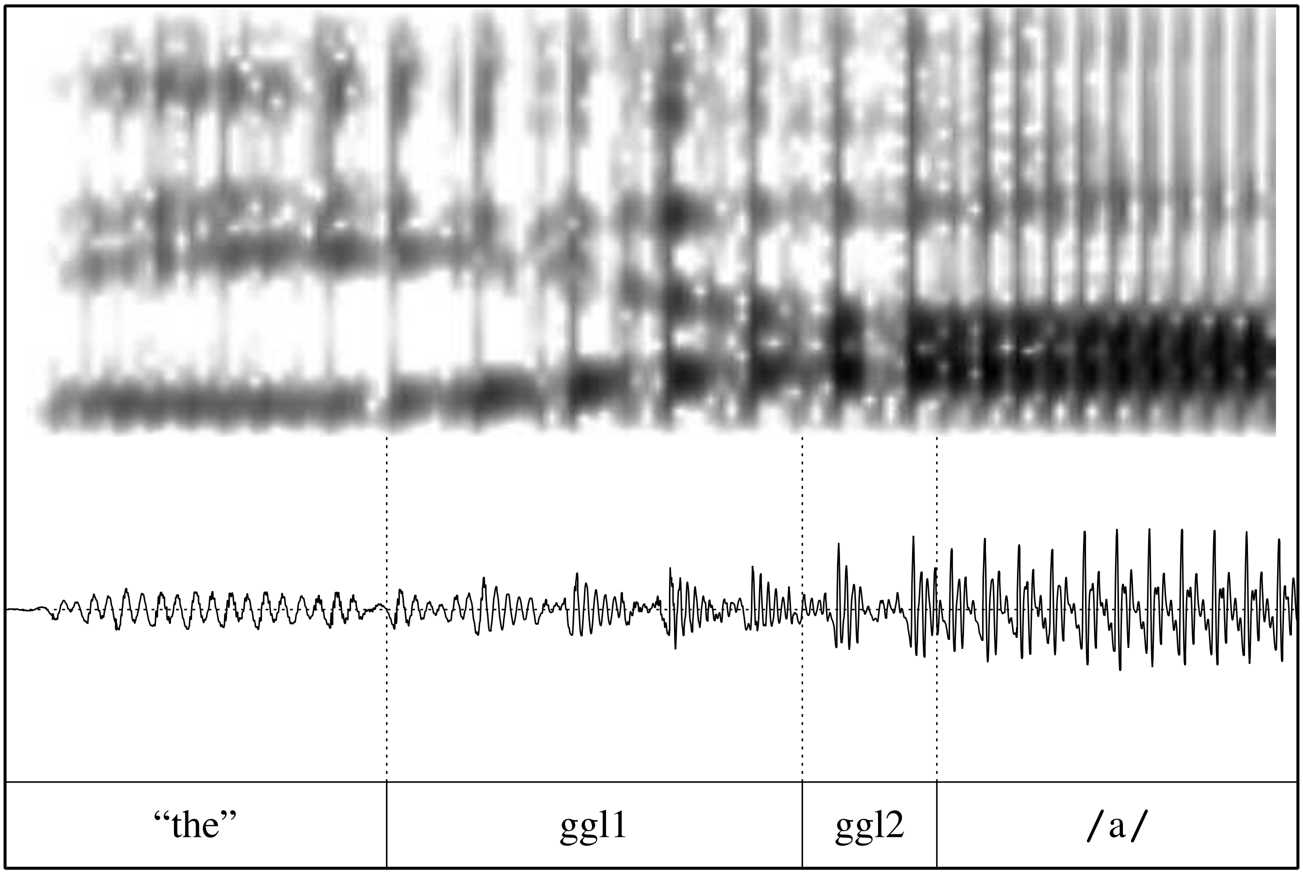


Figure 1f

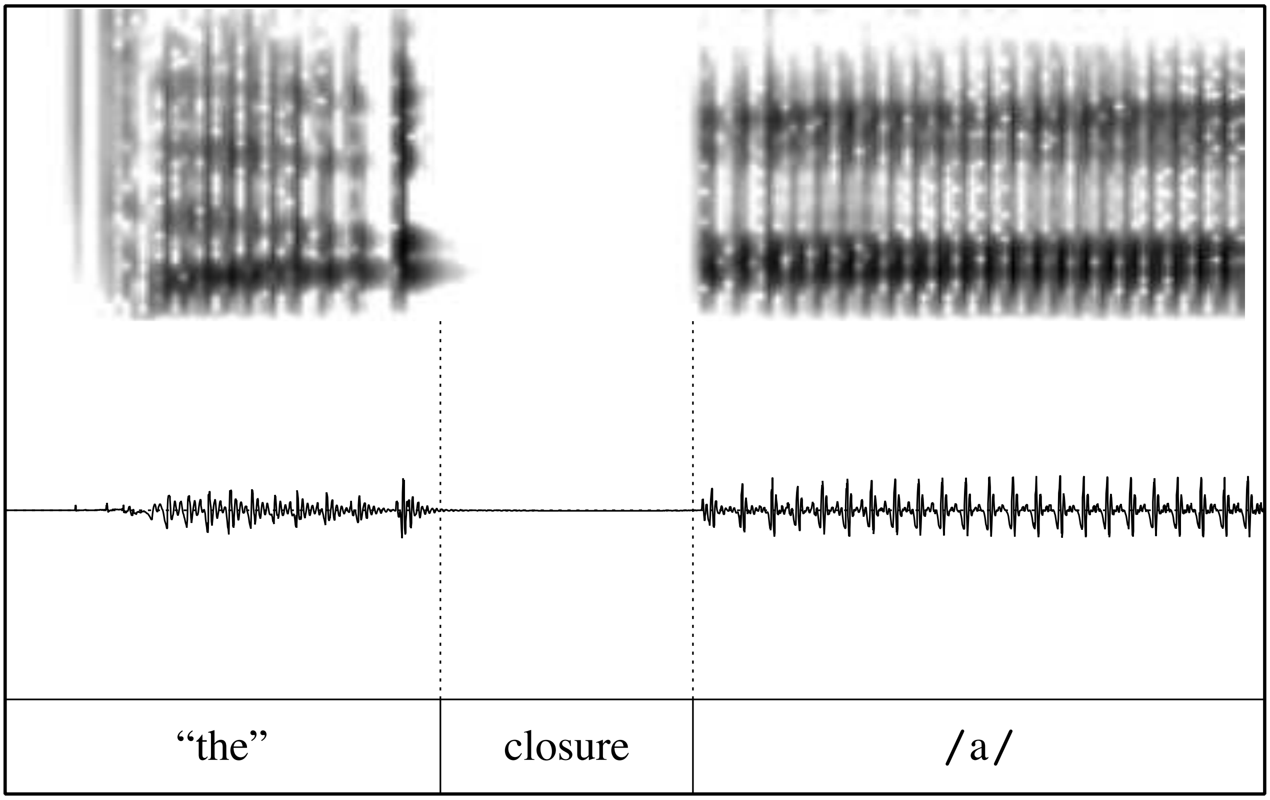


Figure 2

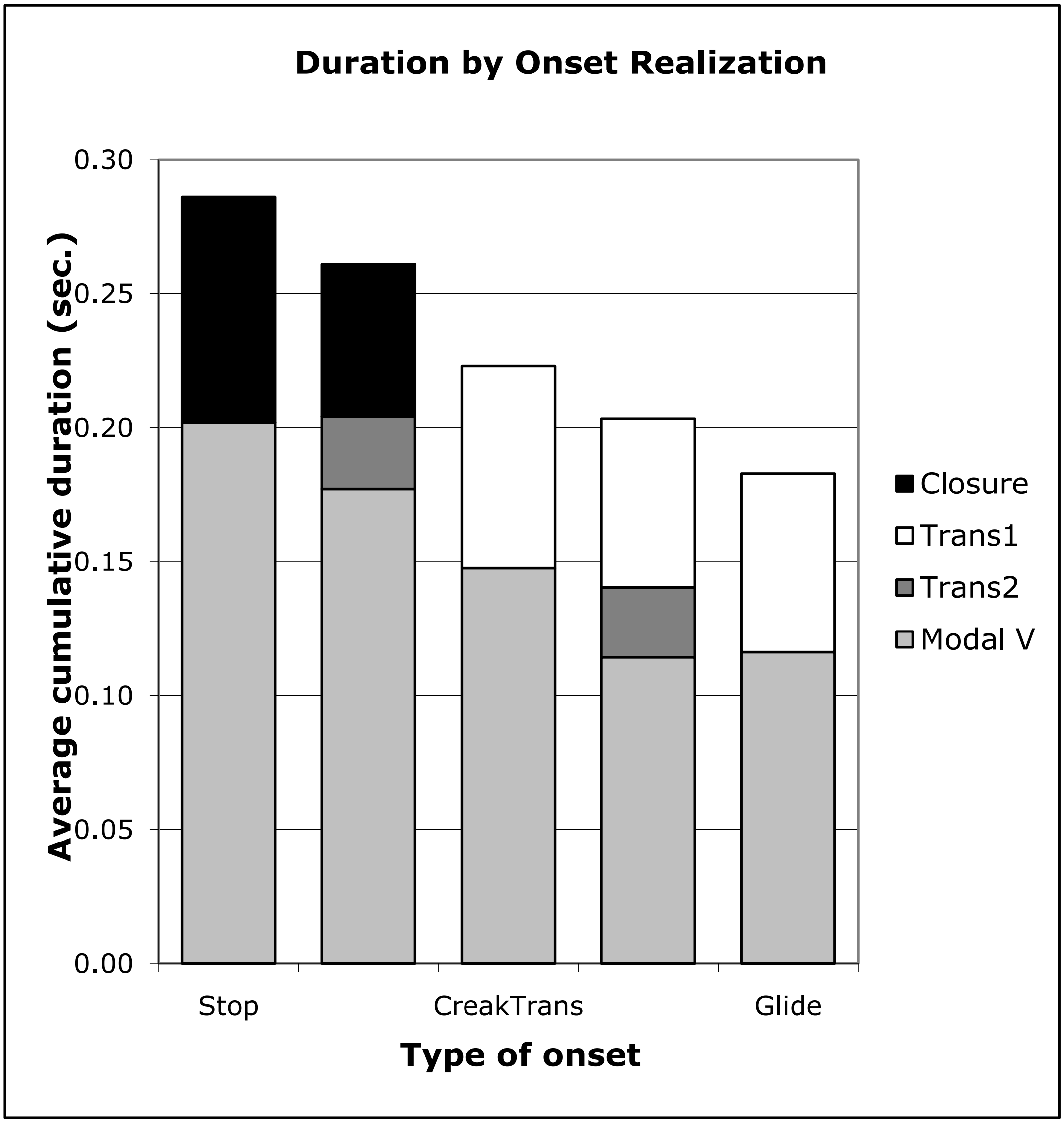

Duration by Onset Realization
Average cumulative duration (sec.)
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Stop
CreakTrans
Glide
Type of onset
■ Closure
□ Trans1
■ Trans2
■ Modal V

Figure 3

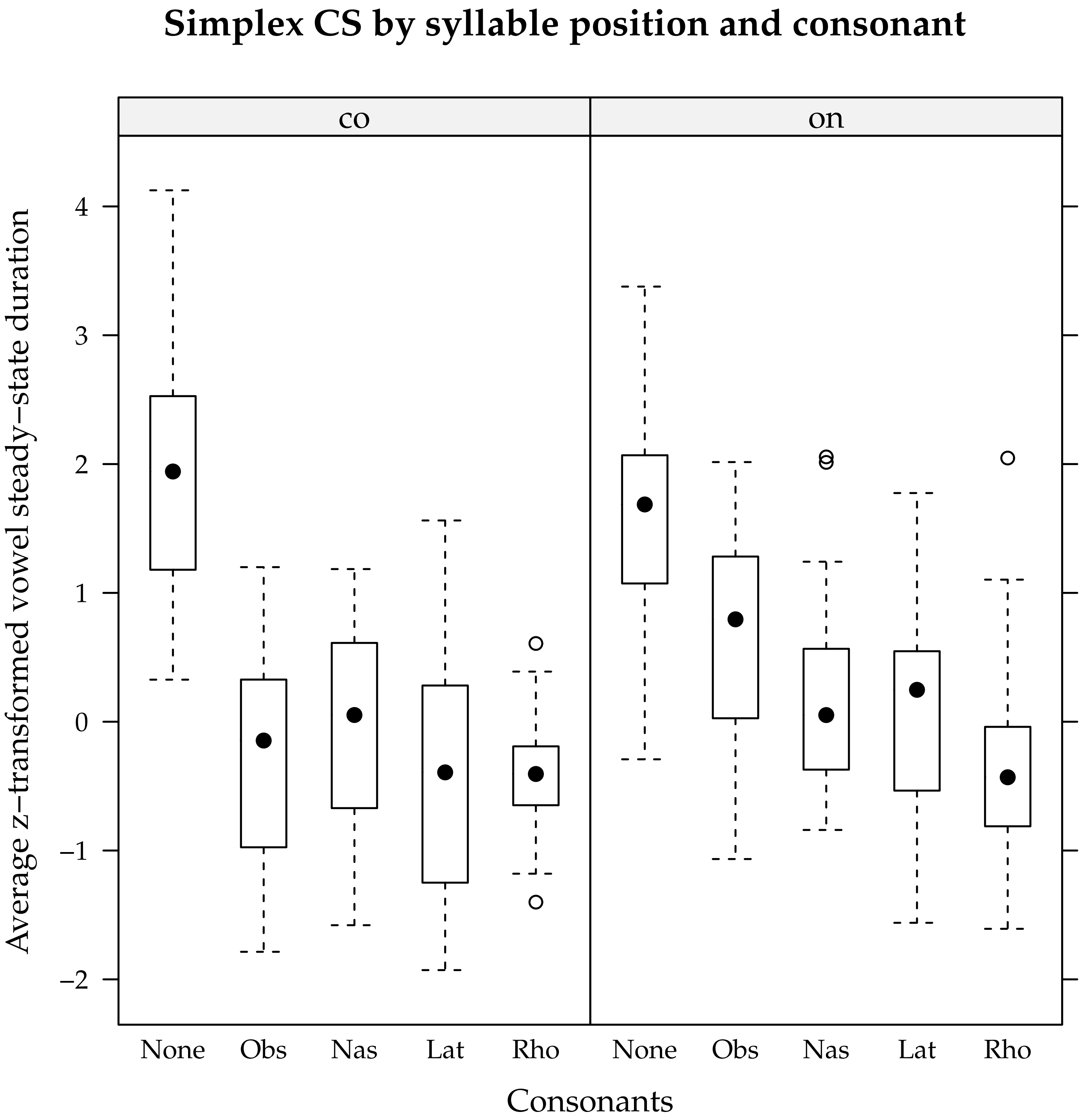
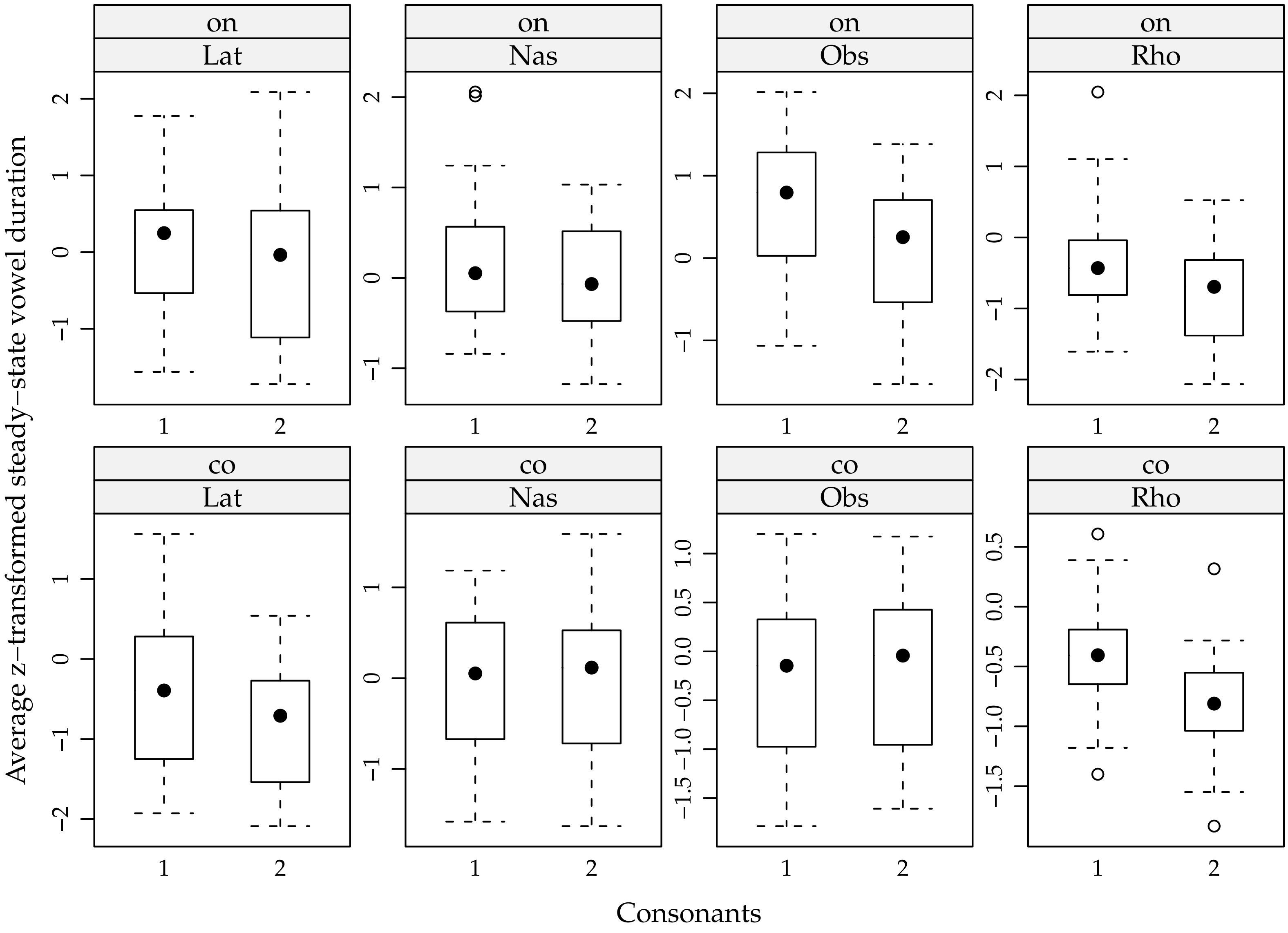
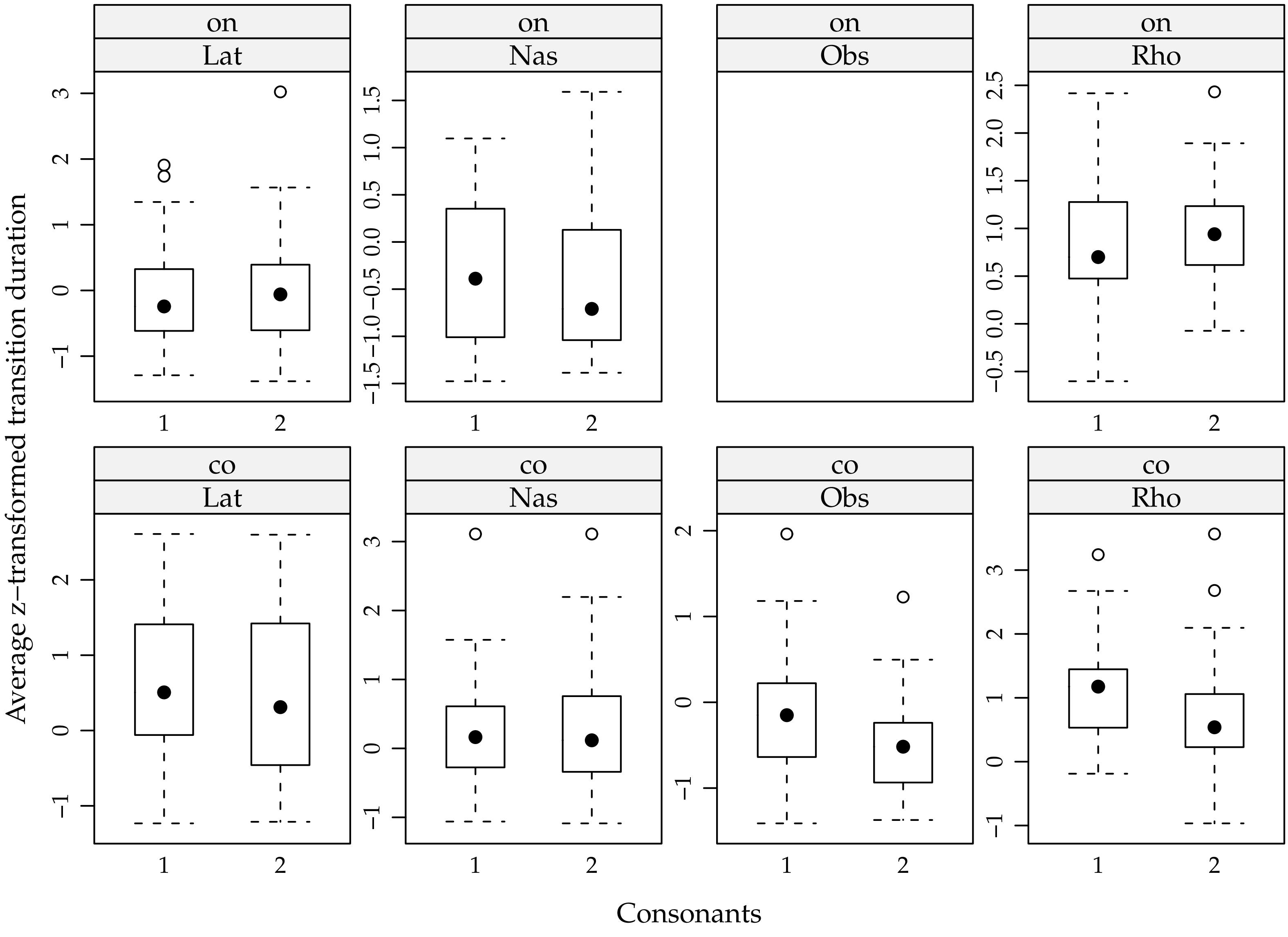

Simplex CS by syllable position and consonant
co
on
Average z−transformed vowel steady−state duration
4
3
2
1
0
−1
−2
None
Obs
Nas
Lat
Rho
None
Obs
Nas
Lat
Rho
Consonants

Figure 4



on
Lat
on
Nas
on
Obs
on
Rho
co
Lat
co
Nas
co
Obs
co
Rho
Average z−transformed steady−state vowel duration
Consonants
1
2
−2
−1
0
1
2
−1.5
−1.0
−0.5
0.0
0.5
1.0

Figure 5



on
Lat
on
Nas
on
Obs
on
Rho
co
Lat
co
Nas
co
Obs
co
Rho
Average z−transformed transition duration
Consonants
1
2
−1
0
1
2
3
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5

Figure A1

		Boundary			
		V-trans	**trans-C**	**C-C**	**C-#**
segment / context	l/_ɑ	F1 HP-on	F1 LP-off, ER	ER	F1 LP-on, ED
	l/_ɪ	F2 HP-on	F2 LP-off or F1 LP-off, ER		
	l/_o	F1+F2 HP-on	F1+F2 LP-off, ER		
	l/ɑ_	F1 HP-off, ED	F1 LP-off	ED	F1 LP-off or F2 LP-off
	l/ɪ_	F2 HP-off, ED	F2 LP-off		F1 LP-off
	l/o_				
	r/_ɑ	F3 HP-on	F3 LP-off	ER	F3 LP-on
	r/_ɪ				
	r/_o				
	r/ɑ_			ED	F3 LP-off
	r/ɪ_				
	r/o_				
	n/_ɑ	F2 LP-on	F1 LP-off, ER	End of silence or onset of voicing	F1 LP-on, ED
	n/_ɪ	F2 HP-on			
	n/_o	F1 HP-on			
	n/ɑ_	F2 LP-off	F1 LP-on, ED	1st appearance of aperiodic noise, ED	F1 LP-off, ER
	n/ɪ_	F2 HP-off			
	n/o_	F2 LP-off			
	pʰ/_ɑ	Onset of energy around F1	Onset of aperiodic noise following burst	--	ED
	pʰ/_ɪ				
	pʰ/_o				
	p/_ɑ	F1 HP-on	Onset of energy around F1	Offset of HP of energy above 5 kHz	--
	p/_ɪ	F2 HP-on			
	p/_o	F1 HP-on or F2 LP-on or F2 HP-on			
	s/ɑ_	F2 LP-off	HP of energy above 5 kHz		Offset of HP of energy above 5 kHz
	s/ɪ_	F2 HP-off			
	s/o_	F2 LP-off			
	d/#_	--	Onset of energy around F1	--	ED
	d/_#				
	ɾ/_#		ED, F1-3 HP-off		ER, F1-3 LP-off