Syllable count judgments and durations of liquid rimes in English

Sam Tilsen, Abby Cohn, and Eric Ricciardi

1 Introduction

For most monomorphemic words in English, native speakers have robust, consistent intuitions regarding the number of syllables that comprise the word. This observation seems to validate the notion that speakers use syllable-level representations alone in judging the syllable count of a given word. However, there is a small class of words, consisting of a diphthong or high/mid tense vowel nucleus and liquid coda (e.g. *pile*, *pail*, *pool, fire, fail, fool*) for which speakers do not exhibit consistent syllable-count judgments (henceforth "σ-count judgments"). The same variation is not observed with low or lax vowel nuclei, nor with non-liquid sonorant codas. This raises the question of why σ-count judgments are variable only for words with the aforementioned class of rimes. We will subsequently refer to the relevant class of words as "variable-count words", because of the inconsistency in speakers' σ-count judgments. While some speakers judge the variable words as comprised of one syllable, others judge them as comprised of two, and still others as more than one, but not quite two syllables.

Previous research (Cohn, 2003; Lavoie & Cohn, 1999) suggests a relation between the phenomenon of variable σ-count intuitions and mora-level representation; specifically, variable-count words can be analyzed as having a trimoraic ("superheavy") syllable structure. Here we consider two hypotheses regarding this relation. One is an intuition-based hypothesis: variability in σ-count judgments arises from cross-speaker differences in reliance on moraic structure for forming σ-count intuitions. In this hypothesis, all speakers have a trimoraic representation of variable-count words, but some speakers give more importance than others to moraic structure when they form a σ-count intuition. An alternative hypothesis is that variation in σ-count intuitions is attributable to variation in the moraic representations themselves: speakers who judge variable-count words as monosyllables have a bimoraic representation, speakers who judge them as more than one syllable have a trimoraic representation. These two hypotheses posit different origins of cross-speaker variation in σ-count judgments: the intuition-based hypothesis attributes the cause of variation to the *process* of forming σ-count intuitions, whereas the structurally-based hypothesis attributes the cause of variation to differences in the structural representation itself.

Although we have couched the above hypotheses in terms of "moraic" structure, we do so in a generic sense, not implying a commitment to a specific moraic theory. The crucial point is to differentiate between syllable level organization and subsyllabic organization. The hypotheses do not require a commitment to the view that moras are the correct analysis of sub-syllabic structure, nor to any particular instantiation of moraic theory. Rather, we refer to "moraic structure" as structure that organizes segments or gestures within a syllable.

In order to better understand variation in σ-count judgments, we posed the question of whether such variation correlates with how speakers produce variable-count words. Under the intuition-based hypothesis, variation in production is not predicted to correlate with variation in σ-count judgments, because the role of moraic structure in σ-count intuition formation process is independent from its role in production. Under the structurally-based hypothesis, variation in production is predicted to correlate with variation in σ-count judgments. Specifically, speakers who judge variable-count words as monosyllabic (i.e. =1σ) will produce them differently from those who judge them as more than one syllable or as two syllables (i.e. >1σ). We tested these hypotheses with sequential and parallel σ-count judgment and word production tasks, and found support for the structurally-based hypothesis: rime durations of variable-count words produced by speakers with >1σ judgments are longer than those produced by speakers with =1σ judgments.

1.1 Background

Lavoie & Cohn (1999) used a questionnaire to elicit σ-count judgments of variable-count words and monosyllabic/disyllabic controls from six speakers of northern American English. Participants were allowed to characterize each word as monosyllabic, disyllabic, or one-and-a-half syllables. Three of the participants consistently judged the variable-count words as monosyllabic, the other three consistently judged them as 1.5 or 2 syllables. All words with low or lax vowels and liquid rimes, and all words with nasal or stop codas were judged consistently as monosyllabic. Furthermore, words with strong orthographic cues to disyllabicity, such as a vowel-consonant-vowel sequence (e.g. “flower”), were consistently judged as disyllabic as opposed to words without such cues (e.g. “flour”), which were associated with variable judgments. Thus, the study established that there exists within-dialect variation in σ-count judgments of variable-count words.

Another form of evidence for variation in σ-count judgments of variable-count words can be observed by comparing online syllable-counting algorithms. Table 1 shows σ-counts reported by several websites for selected words with diphthong-/r/ rimes in English. The table shows that the syllable counters produce differing results. Syllable counters rely in part on orthographically-based algorithms to determine σ counts, so these results do not directly represent speaker intuitions. However, the algorithms themselves are designed by English speakers who must make decisions regarding how orthographic sequences are mapped to syllables, and hence those decisions represent speaker intuitions at least indirectly.

	pyre	*hire*	*fire*	*liar*	TOTAL (σ=2)
wordcalc.com	-	1	1	1	0
howmanysyllables.com	1	1	1	2	1
poetrysoup.com	1	1	2	2	2
syllablecount.com	1	1	2	2	2
TOTAL (σ=2)	0	0	2	3	

Table 1. Syllable counts reported by online syllable counters for selected words with diphthong-/r/ rimes.

The superheavy moraic structure hypothesized to account for >1σ judgments may also have phonetic consequences in production. Lavoie & Cohn (1999) examined the durations of variable-count and non-variable-count words with liquid codas. Words were produced by two female speakers in a frame sentence. They found that the presence of a coda /l/ contributed substantially more duration to the rime in diphthong-/l/ sequences than in low vowel-/l/ sequences, and argued that this can be accounted for if coda /l/ is moraic in the former but not the latter.

The notion that consonant moraicity has consequences for segmental duration has been supported by a variety of studies (cf. Cohn 2003, for a review). For example, Broselow, Chen, & Huffman (1997), comparing Malayalam and Hindi, found that vowel durations are shortened in the presence of codas which share a mora with the vowel, but not by moraic codas. Duanmu (1994) found that syllables with moraic codas in Mandarin Chinese are longer than matched syllables with non-moraic codas in Shanghai Chinese. Ham (2001) found that mora-sharing geminates exhibit a geminate-to-singleton duration ratio of 1.5 in Madurese and Bernese, while non-sharing geminates exhibit an approximately 2.0 geminate-to-singleton ratio in Levantine Arabic and Hungarian. All of these studies indicate that consonants associated with an independent mora contribute more duration to a syllable than consonants which share a mora.

Although the aforementioned studies demonstrate the existence of language-specific variation in production that is conditioned by moraic structure, it remains unknown whether structurally-conditioned variation occurs between speakers of the same language or between words with identical segmental content. The aim of the current experiment is to address this by examining the σ-count intuitions of a large sample of English speakers and assessing the correlation between σ-count intuitions and rime durations.

1.2 Hypotheses

The current experiment involved sequential and parallel σ-counting and production tasks. Participants made σ-count judgments of and produced variable-count words, along with invariant-count monosyllabic and disyllabic controls. The sequential task involved production of all word forms in the stimulus set, followed by the elicitation of σ-count judgments for all stimuli. The parallel task involved the elicitation of a σ-count judgment for a given word, followed by a production of that same word. All participants performed the sequential task first, then the parallel task. Hence in the sequential production, participants were unaware that σ-counts would be elicited subsequently, whereas in the parallel task, participants were aware that σ-counts were under investigation, and their productions were made with recent attention to their σ-count judgments.

The hypotheses below are based on the postulation that σ-count judgments reflect moraic (or sub-syllabic) structure and the assumption that σ-counting involves a subvocal rehearsal of a word, thereby engaging the phonological and motoric representations. It is worth emphasizing that these hypotheses entail no theoretical commitment to the concept of the mora—the same predictions follow if "moraic structure" is substituted with a more generic conception of "sub-syllabic structure," i.e. structure that mediates between syllable- and segment- or gesture-level representation, or structure that organizes segments or gestures into syllable units.

> Hyp. 1: *Variation in production is structurally conditioned*. This hypothesis predicts that rime durations in variable-count forms will vary as a function of σ-count judgments. Specifically, on the basis of previous findings regarding segment moraicity and duration, rime durations associated with >1 σ-count judgments will be longer than rime durations associated with =1 σ-count judgments.

The structurally conditioned variation hypothesis can be assessed on different levels of analysis, depending on the nature of observed variation. If σ-count judgments for a given speaker are mostly consistent across tasks and words with particular rime structure, then the relation between judgments and rime durations can be assessed in a by-speaker analysis. If σ-count judgments vary within rime structures (i.e. by word) for a given speaker but are consistent across tasks, then the relation can be assessed in a by-speaker, by-word analysis. If judgments vary across speakers, words, and tasks, then the relation can be assessed in a by-speaker, by-word, by-task analysis. The corresponding null hypothesis in all cases is that moraic structure does not influence production, which predicts that rime durations will not vary as a function of σ-count judgments.

> Hyp. 2: *Speaker attention to structure heightens structural conditioning*. The effects of structural representation on production may be stronger when speakers have recently attended to their structural representation. This predicts that the effect of σ-count judgment on rime duration will be greater in the parallel task than in the sequential task. The basis for this prediction is that in the parallel task participants produce a word having recently attended to its σ-count, whereas this is not the case in the sequential task.

The hypothesis that attention to structure heightens structural conditioning is based on the notion that speakers have a mechanism for organizing production according to moraic structure, and that this mechanism can be weighted differently in different circumstances. The corresponding null hypothesis is that attention to structure does not influence structural conditioning, which predicts no difference in effects between the sequential and parallel tasks.

2 Method

2.1 Participants and Task

Thirty-four native speakers of English with no speech or hearing problems participated in the experiment; 18 were male, 16 female. Participants' ages were in the range of 18-29 y.o. (median 20 y.o.). Seventeen of the participants had resided in the Eastern U.S. the majority of their life, 12 in the Midwest or Western U.S., and 5 outside the U.S. During the experiment, participants were seated in a sound-proof booth in front of a computer monitor and wore a head-mounted microphone.

The experimental session was organized into three phases as schematized in Fig. 1. The first two phases constitute the sequential production and σ-counting task, the third phase the parallel production and counting task. Before each phase, participants read instructions on the computer monitor. In the first phase, participants were instructed to produce each word that appears on the screen in the phrase *I say* ___ *sometimes*. They were further instructed to "say the entire phrase in one piece", "not hesitate before or after the word", "not emphasize the word that goes in the blank," and "speak clearly but not slowly". If not familiar with a word, they were instructed to guess how to say it. Productions were monitored by an experimenter from outside the booth, and if the experimenter judged that the participant was producing a major intonational break within the phrase, or overly emphasizing the target word, the experimenter demonstrated how to produce the phrase without any major phrase-internal intonational breaks. The entire stimulus set was produced twice in this phase (114 stimuli x 2 reps = 228 trials). In all phases, stimuli were presented in a pseudorandomized order that was constrained such that target words never occurred on consecutive trials. Note that during the first phase, participants were unaware that syllable count judgments would be elicited subsequently.

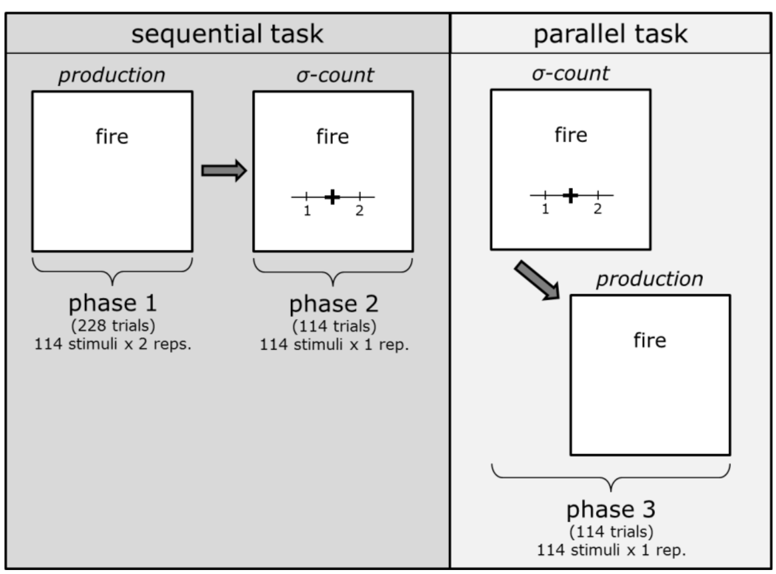

Fig. 1. Sessions were organized into three phases: production, σ-count judgment, and production with σ-count judgment. The first two phases comprise the sequential task, the third phase is the parallel task.

In the second phase of the session, participants produced a σ-count judgment for each stimulus. Because previous studies suggested that variable intuitions may be associated with the impression that a word contains more than 1 syllable but not quite 2 canonical syllables, count judgments were elicited on a continuous scale with a mouse-guided pointer. The scale ranged from 0.5 to 2.5 and the integers 1 and 2 were labeled with tick marks (see Fig. 1). At the start of each trial the pointer appeared at a value of 1.5. The pointer was constrained to move only horizontally along the scale. Participants were given 5 seconds to click the mouse to indicate their judgment, otherwise no response was recorded. Prior to beginning this phase, participants were given instructions that read as follows: "In this part of the experiment, you will decide whether there are one or two syllables in a word. Note that in some cases there is no right answer: people disagree on how many syllables are in some words. In addition, sometimes people feel that the number of syllables in a word is between whole numbers." Furthermore, participants were instructed that when a word appeared on the screen, they should silently read the word before responding. Each word was displayed on the screen for 1.5 seconds before disappearing, at which point the scale appeared. Participants were also explicitly instructed not to rely on how words are spelled, and told that they should rely on what they hear when they imagine saying the word. A σ-count judgment was elicited once for each stimulus in this phase.

In the third phase of the session, participants performed the production and σ-count judgment in tandem, once for each stimulus. On each trial, they first made a σ-count judgment for a word, and then produced that same word in the carrier phrase. Note that the productions in the third phase were made after participants were aware that σ-counts were being investigated, and that for each stimulus word they had produced two judgments recently (one in the second phase, one just prior to the production in the third phase). After completing all three phases, participants filled out a survey on their language background, geographic residence history, linguistic educational background, and familiarity with low-frequency target items in the experiment.

2.2 Stimuli

A list of target stimuli was constructed for elicitation of σ-count judgments and word production. Target stimuli involved all phonotactically licit combinations of the vowels {ɪ, i, a, ai} and codas {Ø, d, n, l, r} (Ø = no coda, i.e. an open syllable), as shown in Table 2 below. Note that the lax vowel /ɪ/ does not occur in open syllables and that the tense/lax contrast in high front vowels is merged before /r/. In order to facilitate automated acoustic analyses, all words were required to have a singleton labial onset consonant (i.e. /p/, /b/, /f/, /v/), or in the absence of viable candidates meeting this criterion, a singleton alveolar stop onset, either /t/ or /d/. No morphologically complex stimuli were included. On the basis of their phonological rimes, 13 of the 50 target stimuli are expected to be variable-count forms (Table 2, shaded cells). However, there is some ambiguity in whether high vowel-/r/ rimes are expected to be variable-count items, because of the tense/lax merger in high vowels before an /r/ coda; Cohn & Lavoie (1999) analyzed these forms as bimoraic and hence unequivocally monosyllabic, but here we include them as potentially variable-count. All of the variable-count stimuli were required to be 4 graphemes long.

	coda				
nucleus	Ø	d	n	l	r
ɪ		bid vid	pin bin fin	pill bill fill	beer fear pier
i	bee fee pea	bead feed	bean teen	peel feel veal	
a	pa bah fa	pod bod	bon Von	pall ball fall doll[1]	par bar far
ai	pie buy vie	bide tide	pine fine vine	pile bile vile file	pyre fire tire

Table 2. Target stimuli involve all phonotactically licit combinations of the syllable nuclei /ɪ/, /i/, /a/, and /ai/ with the codas /Ø/, /d/, /n/, /l/, and /r/ (Ø = no coda, i.e. an open syllable).

Factors such as orthographic composition of the rime, grapheme count, and word frequency would ideally be controlled across target stimuli. However, the English lexicon does not allow for perfect control over all of these factors. Hence for some rimes, orthographic composition was varied (e.g. *feel* vs. *veal*). An attempt was made to control for word frequency by preferring words with CELEX log-frequencies in the 25-75% percentile range (see Appendix: Table A.1 for target word log-frequencies). However, not all design cells could be sufficiently populated when holding to this criterion, and hence a few less frequent words, words with unknown frequency, and proper names were included (e.g. *pyre*, *vid*, *Von*). The influence of orthography could be mitigated to some degree if stimulus targets were cued with images rather than orthographically; however, due to variation in familiarity and grammatical category of target words, the use of image-based cues was deemed impractical.

In order to mitigate the potential for experiment-wide statistical properties of stimuli to create a response bias, non-target items (n=64, cf. Appendix: Table A.1) were selected so as to balance the stimuli in two ways. First, the total number of unequivocally monosyllabic and disyllabic stimuli was equal; hence judgments for the variable-count words cannot be attributed to an experiment-wide imbalance in stimuli. Second, the correlation between graphemic length and syllable count across all stimuli was minimized—hence there were approximately equal numbers of unequivocally monosyllabic and disyllabic words for a given graphemic length (all words ranged from 3-5 graphemes), discouraging participants from relying on graphemic length as a response strategy. All non-target items were required to be in the CELEX log-frequency 25-75% percentile range.

2.3 Data Processing

Despite the availability of a continuous response dimension in the σ-counting task, experiment-wide participant responses were highly multimodal, with the majority of participants exhibiting

[1] The word *doll* may have the vowel /ɔ/ or more commonly /a/, representing a merger that is increasingly characteristic of younger speakers, even in areas described as maintaining a /ɔ/-/a/ contrast.

either bimodal distributions with modes near 1 and 2, or trimodal distributions with modes near 1, 1.5, and 2. However, there were several participants who used the continuum less discretely. Hence, in order to analyze syllable count as a multinomial variable, a participant-dependent procedure for mapping from gradient syllable count judgments to categories was employed. For each participant, an empirical Gaussian kernel density function of responses was calculated (bandwidth 0.025, 100 support points from 0.75 to 2.25) and this density function was fit with bimodal and trimodal Gaussian mixtures. The modes of the fitted Gaussians were constrained in the ranges (0.75, 1.25), (1.25, 1.75), and (1.75, 2.25), respectively. Bin edges for categorizing the gradient click values were set to 4 standard deviations above the estimated monosyllabic mode and 4 standard deviations below the estimated disyllabic mode. Parameters of the trimodal model were used if it provided a significantly better fit than the bimodal model, otherwise the parameters of the bimodal model were used.

Fig. 2 illustrates the results of this procedure for three representative patterns of within-participant response distribution. Fig. 2 (a) shows a common pattern, in which responses are distributed bimodally; (b) shows another common pattern, in which a mode near 1.5 is clearly present; (c) shows a pattern exhibited by several of the participants, in which intermediate values are more uniformly distributed rather than associated with a single intermediate mode. All subsequent analyses treat count judgments as representing either 1 syllable (=1σ, i.e. belonging to the monosyllabic bin) or more than 1 syllable (>1σ, i.e. belonging to the intermediate or disyllabic bin).

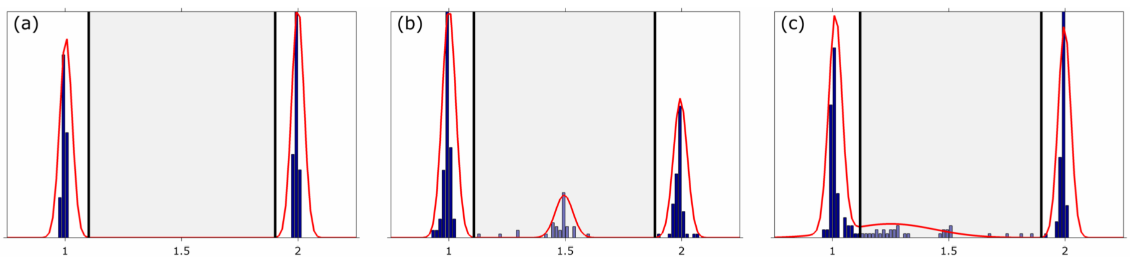

Fig. 2. Three representative patterns of σ-count judgment distributions, bimodal/trimodal fits to density functions, and categorical partitioning. Black vertical lines delineate monosyllabic, variable-count, and disyllabic bins. (a) bimodal response pattern; (b) trimodal response pattern; (c) bimodal response pattern with a relatively uniform intermediate response distribution.

2.4 Participant Exclusions

Data from 6 of the 34 participants were excluded from subsequent analyses because these participants produced a high proportion of non-standard σ-count judgments for unequivocally monosyllabic and disyllabic words. These participants were likely either not attending closely to the experimental task or were overly relying on grapheme length in their σ-count judgments. Table 3 lists participants with >10% mismatch rates (proportion of σ-count judgments deviating from the expected ones) for all items, along with mismatch rates for target and non-target items separately. 28 participants with mismatch rates lower than 10% were not excluded, and of these, 19 produced 5 or fewer mismatched judgments (< 2.5%) across the experiment.

The likely source of the mismatched responses is over-reliance on grapheme length for syllable count judgments. This is viable as a strategy for estimating syllable counts because words with more syllables generally have more graphemes. Table 3 shows the p-values of χ^2 tests for the effects of grapheme length on over- and under-counting mismatches (an over-counting mismatch occurs when an unequivocal monosyllable is judged as more than one syllable, and an under-counting mismatch occurs when an unequivocal disyllable is judged as one syllable). Significant values indicate that the over- or under-counting mismatches were biased by the number of graphemes in a stimulus for several of the excluded participants. Inspection of the distributions revealed that words with 5 graphemes were associated with over-counting mismatches and words with 3 graphemes were associated with under-counting mismatches.

	all items			**non-target only**			**target only**		
		grapheme effects (p-value)			grapheme effects (p-value)			grapheme effects (p-value)	
participant	mismatch rate	over	under	mismatch rate	over	under	mismatch rate	over	under
JA01	0.34	0.01		0.29	< 0.01		0.42	< 0.01	
IF01	0.28	< 0.01	0.05	0.43	0.01	0.00	0.03	0.26	
NS01	0.19	< 0.01		0.25	< 0.01		0.09	0.01	
AR01	0.18	< 0.01	0.00	0.27	< 0.01	0.00	0.03	0.26	
LX02	0.15	0.24	0.31	0.23	0.54	0.00	0.03	0.26	
LO01	0.12	0.04	0.72	0.12	0.04	0.04	0.14	< 0.01	

Table 3. Participants excluded because of relatively high rates of mismatch between canonical syllable count judgments and observed syllable count judgments, along with p-values from a χ^2 test on the effect of grapheme length on the likelihood of over- and under-counting mismatches (values only shown when mismatches occur).

3 Results

Substantial interspeaker variation was observed in σ-count judgments of diphthong-liquid rimes (/ail/, /air/), and to a lesser extent interspeaker variation was observed in high-front/tense vowel-liquid rimes (/il/, /ir/). Word-specific variation was also observed for both diphthong and high-front/tense vowel rimes in σ-count judgments. Variation in σ-count judgments was observed between the sequential and parallel tasks, but these task effects show no systematic pattern. For all variable-count targets, rime durations were significantly longer in words judged as >1σ than in words judged as =1σ, in both sequential and parallel tasks. This finding supports the hypothesis that variation in production is structurally conditioned. The hypothesis that attention to structure heightens the structural conditioning effects was not supported, because effects were not consistently stronger in the parallel task compared to the sequential task.

3.1 Across-Participant Variation in σ-Count Judgments

Analysis of the counts of >1σ judgments for each participant revealed substantial interspeaker variation, particularly for the diphthong-liquid rimes.

Fig. 3 illustrates the experiment-wide proportions of σ-count judgments for each participant and rime. Notably, the diphthong-liquid rimes exhibited more variation than the /i/-liquid rimes. 10/28 participants judged all or most (all but one) of the /ail/ rimes as >1σ, while 7/28 participants judged all or most of these rimes as =1σ. The remaining 11 participants exhibited intermediate proportions of >1σ judgments, reflecting within-participant, word- and/or task-specific variation. A similar pattern was observed for /air/ rimes. The /i/-liquid rimes exhibited less variation: all but 2 participants judged most of the /il/ rimes as =1σ and all but 3 judged /ir/ rimes as =1σ. Only one speaker was mostly consistent in judging /ir/ rimes as >1σ.

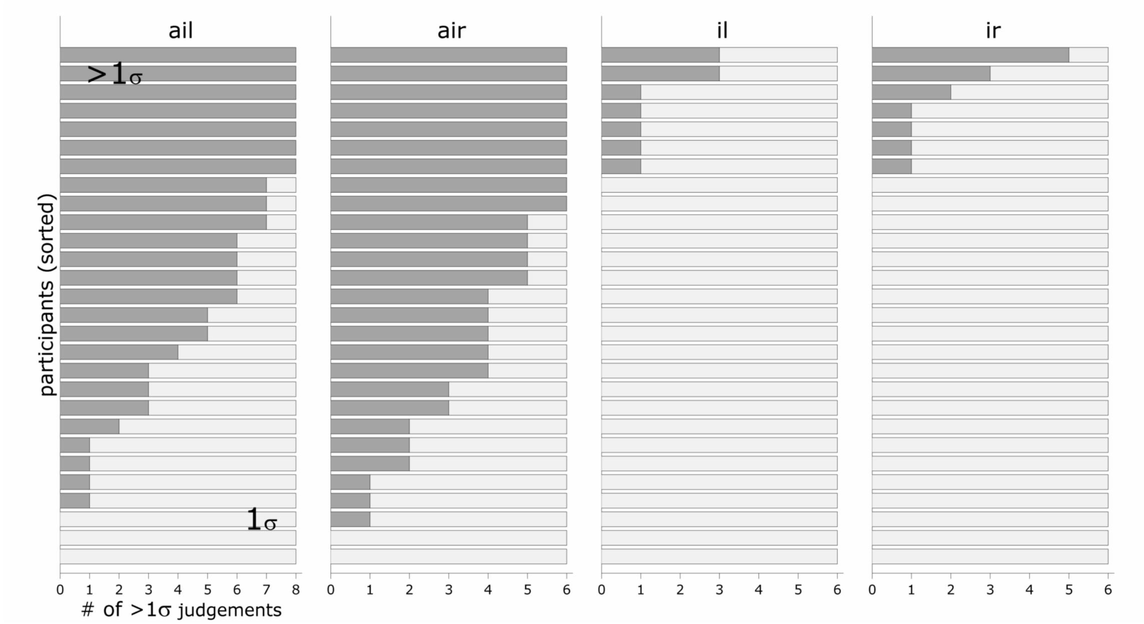

Fig. 3. Counts of >1σ judgments for variable-count rimes by participant, sorted by within-participant proportion. The prevalence of intermediate counts for diphthong-liquid rimes indicates word- and/or task-specific variation.

3.2 Word-Specific Variation in σ-Count Judgments

Analysis of word-specific variation in σ-count judgments shows that rime composition was not the only factor influencing σ-count: word-specific variation was observed as well. Fig. 5 shows experiment-wide proportions of >1σ judgments by word. For diphthong rimes, less frequent words such as *bile*, *vile*, and *pyre* were associated with a greater number of >1σ judgments than their more frequent counterparts *file*, *pile*, *fire*, and *tire*. For monophthong rimes, less frequent *veal* and *pier* were associated with a greater number of >1σ judgments than more frequent counterparts *feel*, *beer*, and *fear*. A stepwise linear regression of the experiment-wide proportions of >1σ judgments by word, with rime and log-frequency as predictors, showed that word frequency was a significant factor in σ-count judgments ($F = 7.9$, $p = 0.02$). The relation observed reflects a negative correlation, i.e. lower frequency words were associated with a higher proportion of >1σ judgments.

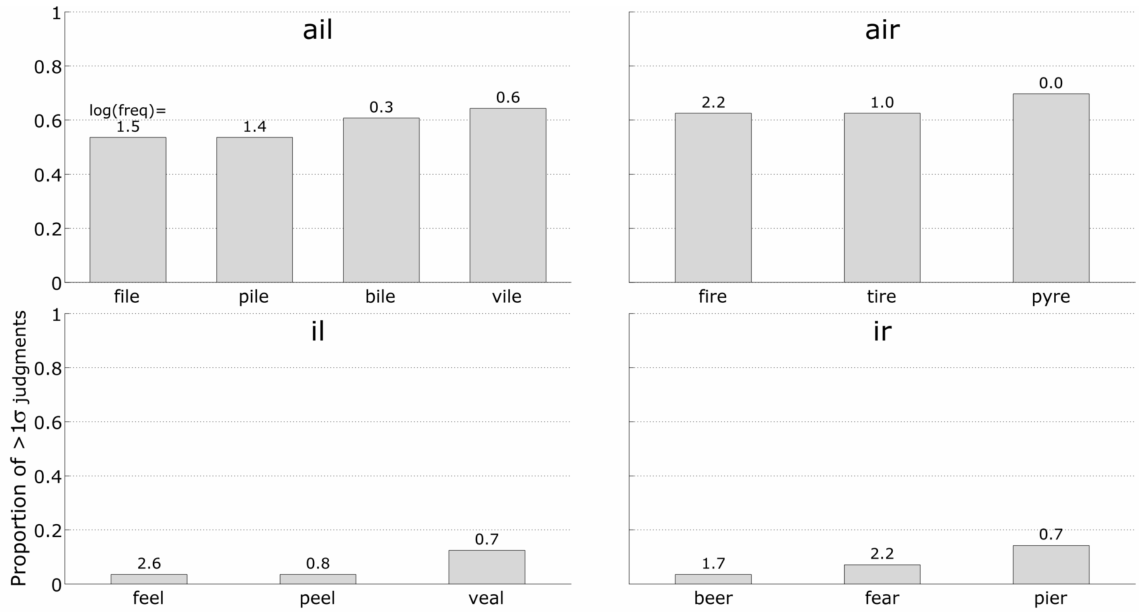

Fig. 5. Experiment-wide proportions of >1σ judgments by word. CELEX Log-frequency (per 1,000,000 words) is shown above each bar.

3.3 Task-Specific Variation in σ-Count Judgments

The majority of σ-count judgments were consistent across tasks, but a substantial proportion of σ-count judgments changed for a given participant between the sequential and parallel tasks. Table 4 shows the percentage of judgment changes over the experiment, i.e. the percentage of times that the judgment of a word changed between the sequential and parallel tasks. 23% and 30% of all /ail/ and /air/ judgments changed between tasks. Changes were less frequent for /il/ and /ir/ rimes.

However, no general trends are evident in changes of σ-count judgments between tasks. Fig. 6 shows the counts of participants who changed from >1σ to =1σ judgments and vice versa from the sequential to the parallel tasks. Three of the four /ail/ rime stimuli exhibited a predominance of changes to a >1σ judgment, but the pattern was reversed for the word *bile*, and in all cases some participants changed in the other directions. In contrast, the predominant change in the /air/ rimes was to a =1σ judgment, although some participants changed to >1σ judgments as well.

rime	% of changed judgments	# of changed judgments	# judgments per task
/ail/	23%	26	112
/air/	30%	25	84
/il/	13%	11	84
/ir/	10%	8	84

Table 4. Percentage of judgment changes between tasks for each rime.

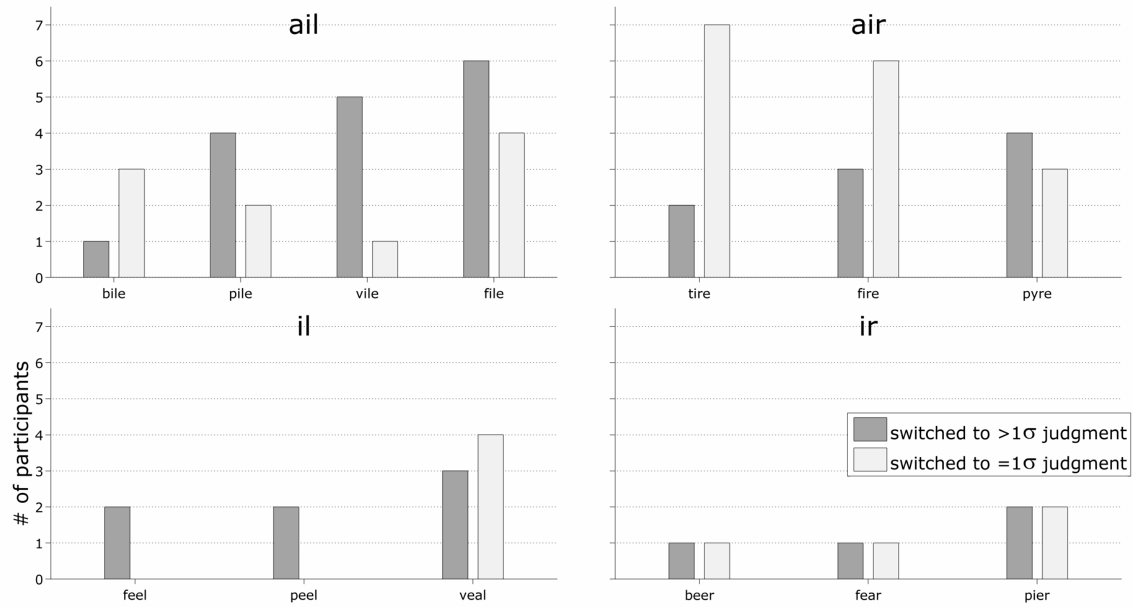

Fig. 6. Number of participants who changed judgments between tasks by word. Dark/light bars show the number of participants who changed to >1σ / =1σ in the parallel task.

3.4 Nucleus and Coda Effects on Rime Durations

Rime durations in target word productions were strongly influenced by rime composition, i.e. the nucleus and coda, and also by word identity. Fig. 7 and Fig. 8 show means and ranges of rime durations compared within nucleus and coda categories, respectively. Main effects of VOWEL, CODA, WORD (nested within VOWEL and CODA), and a VOWEL-CODA interaction were all significant in an ANOVA of rime duration (VOWEL: $F(3,4009) = 603.1$, $p< 0.001$; CODA: $F(4,4009) = 200.0$, $p< 0.001$; WORD: $F(42, 3967) = 15.5$, $p < 0.01$; VOWEL-CODA: $F(8,3359) = 43.3$, $p < 0.001$; note that the vowel-coda interaction effect was calculated in a separate ANOVA with lax vowel /I/ rimes excluded).

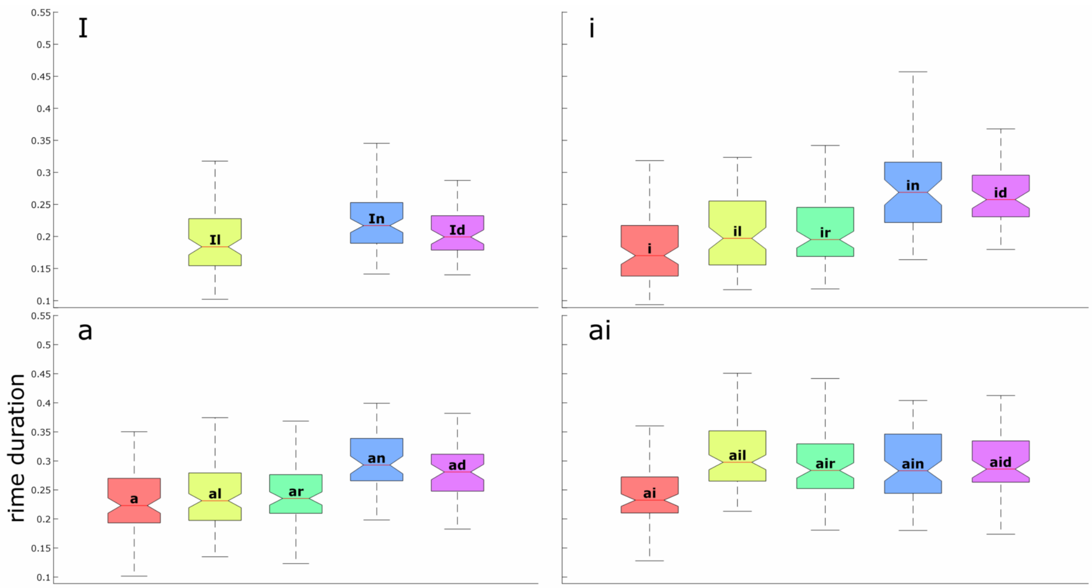

Fig. 7. Rime durations from all tokens compared across codas within nuclei. Error bars show the range of data in 5-95% percentiles, boxes show range of data in 25-75% percentiles, and notches show ±2.0 standard error.

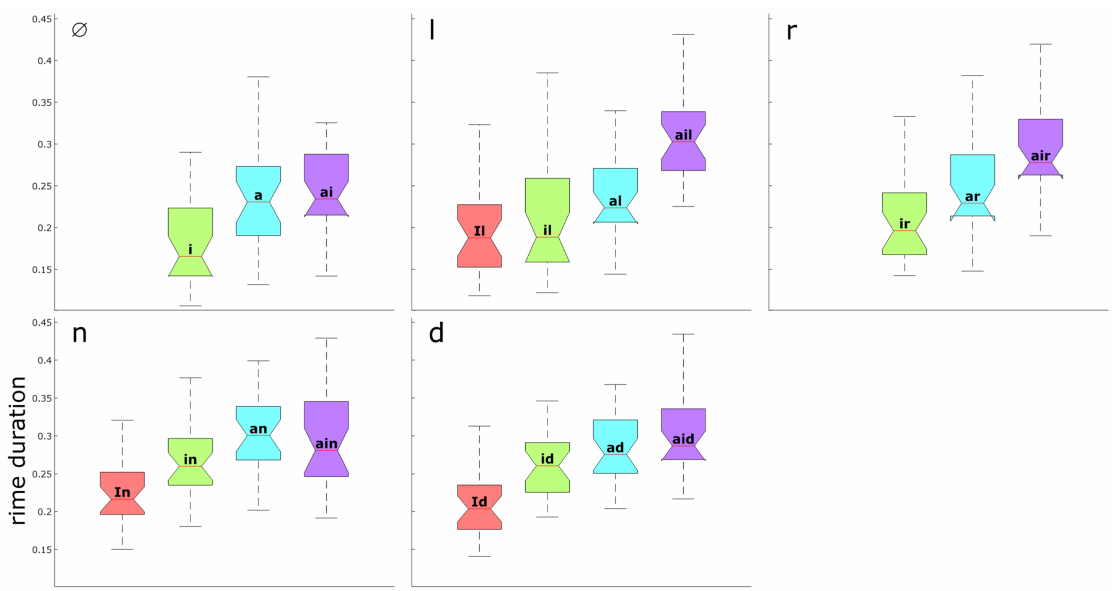

Fig. 8. Rime durations compared across nuclei within codas. Error bars show the range of data in 5-95% percentiles, boxes show range of data in 25-75% percentiles, and notches show ±2.0 standard error.

An important observation is that the liquid codas, /n/, and /d/ all contribute a substantial amount of duration to /ai/-nucleus rimes, resulting in rime durations that are significantly longer than those in the open syllable /ai/ (cf. Fig. 7); but in rimes with a low vowel /a/ nucleus, the liquids do not contribute a substantial amount of duration to the rime: only the /n/ and /d/ codas result in significantly greater rime duration compared to the open syllable /a/. A partly similar effect is observed with the high-front/tense vowel: /il/ and /ir/ rime durations are significantly greater than open /i/ rime durations; however, in these rimes, /n/ and /d/ codas contribute even more duration, resulting in rime durations that are significantly greater than /il/ and /ir/. These same observations can be seen from a different perspective in Fig. 8, which compares rime durations by vowel nucleus: /ail/ and /air/ rimes are significantly greater than /al/ and /ar/ rimes, respectively. These patterns indicate that across participants there is a general trend for liquid codas to contribute extra duration to the rime in variable-count words, i.e. in words with diphthong or high/front tense vowel nuclei.

3.5 Relation between Rime Duration and σ-Count Judgments

Analysis of rime durations shows that words associated with >1σ judgments were produced with greater rime durations than words associated with =1σ judgments. This finding supports the structurally conditioned variation hypothesis. Two measures of rime duration were analyzed: (1) absolute rime duration and (2) normalized rime duration, expressed as the ratio of absolute rime duration to the within-speaker average rime duration of a nucleus-matched open syllable. For example, the normalized measure expresses the rime duration of a given token of /ail/ as a ratio of its absolute duration to the mean duration of /ai/ produced by the same speaker. This normalization accommodates the fact that speakers differ in their baseline word durations and allows the contribution of a liquid coda to rime duration to be characterized in a more speaker-independent fashion. Two-sample one-sided t-tests were conducted for each speaker and rime category. Significant differences between rime durations are indicated with * ($p < 0.05$) and marginal differences with + ($p < 0.10$) in Fig. 9, along with boxplots of durations from each sample. Table 5 shows p-values, Cohen's d standardized effect size measure, and counts of =1σ and >1σ judgments in each sample.

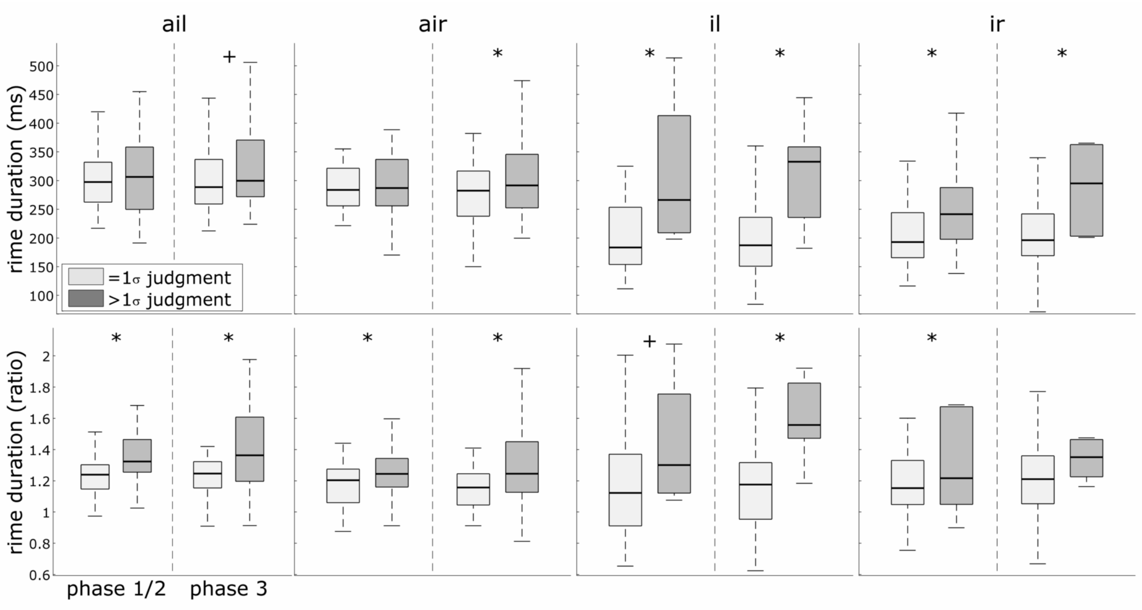

Fig. 9. Effects of σ-count judgment on rime duration in the sequential (phase 1/2) and parallel (phase 3) tasks. (Top) Raw rime durations. (Bottom) Rime durations as a ratio of within-subject average open syllable duration. (*) Two-sample, one-sided t-test $p < 0.05$, (+) Two-sample, one-sided t-test $p < 0.10$.

The majority of rimes exhibited significant effects of σ-count judgment on rime duration in both the sequential and parallel tasks. For /ail/ and /air/ rimes, the ratio measures were significant in both tasks, but the raw duration measure was not significantly different for =1σ and >1σ judgments in the sequential task. In the parallel task, raw durations were significantly different between samples for /air/ and marginally significant for /ail/. Both raw and ratio measures were significant or marginal in both tasks for /il/ and /ir/ rimes. The lack of significance in raw durations for diphthong-liquid rimes may be attributable to interspeaker variation in baseline durations, which heightens the variability in rime duration and thereby decreases statistical power. The ratio measures provide a better reflection of the contribution of a liquid coda to rime duration, and thus the results support the hypothesis that structural variation across words and speakers conditions differences in production.

	sequential		parallel		sequential		parallel	
	t-test p-value	Cohen's d	t-test p-value	Cohen's d	t-test p-value	Cohen's d	t-test p-value	Cohen's d
	/ail/				**/air/**			
N(=1σ,>1σ)	(50,62)		(44,68)		(26,58)		(33,51)	
raw dur.	0.329	0.08	0.052	0.32	0.325	0.11	0.048	0.38
dur. ratio	0.000	0.80	0.001	0.65	0.048	0.40	0.010	0.54
	/il/				**/ir/**			
N(=1σ,>1σ)	(80,4)		(77,7)		(77,7)		(77,7)	
raw dur.	0.002	1.51	0.000	1.53	0.037	0.71	0.002	1.24
dur. ratio	0.075	0.75	0.000	1.57	0.034	0.73	0.112	0.52

Table 5. P-values from two-sample one-sided t-tests for raw and ratio rime durations by syllable-count judgment, along with the Cohen's d standardized effect size. Counts of =1σ and >1σ responses for each rime/task are shown in parentheses.

To rule out the possibility that effects were driven by variation in speech rate or prosodic boundary structure (e.g. speakers with >1σ judgments of a word might use slower speech rates or higher-level phrase boundaries in producing the word), analyses were conducted on the duration of the entire response phrase and the period of time from the midpoint of the pre-target carrier vowel to the midpoint of the post-target carrier fricative (i.e. I s[ei __ s]ometimes), with the duration of the rime subtracted from these measures. These durations did not differ significantly in either task for any rime, and hence the observed effects on rime duration are not likely to be an indirect consequence of differences in speech rate or prosodic boundary strength.

The hypothesis that structurally conditioned effects on production are heightened by attention to structure was not supported. This can be seen by considering the Cohen's d effect sizes (cf. Table 5), which represent the standardized difference in sample means. For the raw duration measure in /ail/ and /air/ rimes, the effect size increased in the parallel task compared to the sequential task. However, for the duration ratio measure, the differences in effect sizes are somewhat smaller and for /ail/ the effect size decreased in the parallel task. The absence of a task effect in the ratio measure suggests that the task effect in raw durations may be attributable to other factors, such as decreased response variability in the parallel task or other changes in participant behavior. Furthermore, although /il/ and /ir/ appear to show some large increases in effect size, these estimates are based on relatively few samples of >1σ judgments (merely 7 or 4), and so firm conclusions should not be drawn from them. Thus the results do not provide strong evidence for the hypothesis that attention to structure vis-à-vis a count judgment heightens structural conditioning.

3.6 Word-Specific Variation in the Relation between Rime Duration and σ-Count Judgments

In order to assess whether the relation between σ-count judgment and rime duration is driven by specific words, one-sided t-tests on raw and normalized rime durations associated with =1σ and >1σ judgments were conducted for each variable-count word. Table 6 shows p-values, Cohen's d, and the difference in means for both variables. The words are sorted by the effect size for the raw duration measure. Notably, the words with the strongest effects are the high-front tense vowel rimes, which also exhibited relatively few >1σ judgments. Words with diphthong nuclei had smaller effect sizes. In the duration ratio measure, the effects for diphthong nuclei were all significant or marginally significant, and the effect sizes are fairly comparable. Thus the observed relation between σ-count judgment and rime duration does not appear to be driven by specific word forms.

	N =1σ	N >1σ	raw duration p-value	raw duration Cohen's d	raw duration Δ (s)	duration ratio p-value	duration ratio Cohen's d	duration ratio Δ
beer	54	2	0.001	3.21	0.196	0.04	1.80	0.416
feel	54	2	0.001	2.73	0.182	<0.001	2.04	0.636
peel	54	2	0.01	1.71	0.117	0.05	1.22	0.344
veal	49	7	0.01	1.10	0.081	0.02	0.85	0.302
fear	52	4	0.03	1.03	0.060	0.08	0.75	0.163
pier	48	8	0.02	0.79	0.051	0.04	0.67	0.152
pile	26	30	0.08	0.40	0.023	0.001	0.88	0.163
pyre	17	39	0.12	0.35	0.022	0.05	0.49	0.088
fire	21	35	0.12	0.33	0.019	0.04	0.50	0.114
vile	20	36	0.21	0.23	0.014	0.03	0.55	0.097
tire	21	35	0.31	0.14	0.009	0.04	0.48	0.102
bile	22	34	0.36	0.10	0.006	0.06	0.55	0.105
file	26	30	0.47	0.02	0.001	0.01	0.69	0.137

Table 6. Effects of judgment on production by word. P-values of t-tests, Cohen's d measure of effect size, and difference between sample means (>1σ - =1σ) are shown for the raw duration and duration ratio measure. Words are sorted by effect size for raw duration.

4 Discussion

σ-count judgments showed substantial interspeaker, word-specific, and task-specific variation for diphthong-liquid rimes (/ail/, /air/); more limited variation was observed in high-front/tense vowel-liquid rimes (/il/, /ir/). Crucially, normalized rime durations were significantly longer in words judged as >1σ than in words judged as =1σ in both sequential and parallel tasks, although no clear difference in the strength of this effect between tasks was observed. Hence the results support the hypothesis that variation in production is structurally conditioned, but do not support the hypothesis that attention to structure heightens structural conditioning.

Previous investigations have observed interspeaker variation in σ-count judgments for variable-count words (Cohn, 2003; Lavoie & Cohn, 1999); the current study, using a larger sample of participants, replicates this interspeaker variation but also reveals a more complex picture which includes word- and task-specific variation. Although the current study examined only a subset of the variable-count rimes (those with the diphthong /ai/ and high-front/tense vowel /i/), these findings are expected to extend to other variable-count rimes.

As a starting point for discussion of the mechanisms underlying the observed variation, consider the hypothesis of Cohn (2003) and Lavoie & Cohn (1999) that variable-count words have a trimoraic structure which may influence σ-count intuitions. In this view, the presence of a third mora in a syllable (or more neutrally, additional sub-syllabic structure of some sort), biases speakers toward an intuition that the syllable is "larger" than a canonical syllable. Below we discuss several schematic models of how this effect may arise. Again, although these models refer specifically to *moraic* structure, the reader should note that nothing hinges on any particular theory of moraic structure. In all cases, "moraic structure" can be substituted with a more generic notion of sub-syllabic structure which organizes segments and/or articulatory gestures, and the bimoraic vs. trimoraic distinction can be understood as a difference in the complexity of sub-syllabic structure.

A model in which σ-count variation originates in a σ-intuition formation process is not consistent with the observed correlation between judgments and rime durations. For example, consider the *invariant representation, variably weighted intuitions* model in Fig. 10A. In this model, all speakers have trimoraic representations of variable-count words, but they differ with regard to whether moraic structure influences their σ-count intuitions. Specifically, an intuition formation process takes syllable-level structure and mora-level structure as input, with weights w_σ and w_μ, respectively. Speakers for whom w_σ is substantially greater than w_μ will always judge variable-count words as =1σ; speakers for whom w_σ and w_μ are more balanced will produce >1σ judgments. Thus interspeaker variation is accounted for by variation in the weighting of syllabic and moraic structure in σ-count intuition formation. If w_μ and w_σ are furthermore allowed to vary on a word-specific basis, then inter-word variation can be likewise accounted for. However, this model cannot account for judgment-production correlations, because the intuition formation process (and its associated weighting terms) is independent of the production process, which takes the trimoraic representation as input in all cases.

An alternative model which does accommodate judgment-production correlation is shown in Fig. 10B. In this *variable representation* model, speakers may have either a bimoraic or trimoraic representation of variable-count words. Those speakers with a bimoraic representation produce =1σ judgments and those with trimoraic representations produce >1σ judgments. Hence the representation itself (rather than the intuition formation process) is the origin of variation in σ-count judgments. Because articulatory control is also driven by the representation, judgments and production can be correlated. The model can furthermore account for word-specific variation if different words are allowed to have different representations. Weighting of syllable-level and mora-level structure is unnecessary to account for judgment variation and judgment-production correlation. However, if such weighting is incorporated into the model, the weighting might be expected to weaken the judgment-production correlation by introducing another source of variation in σ-count judgments.

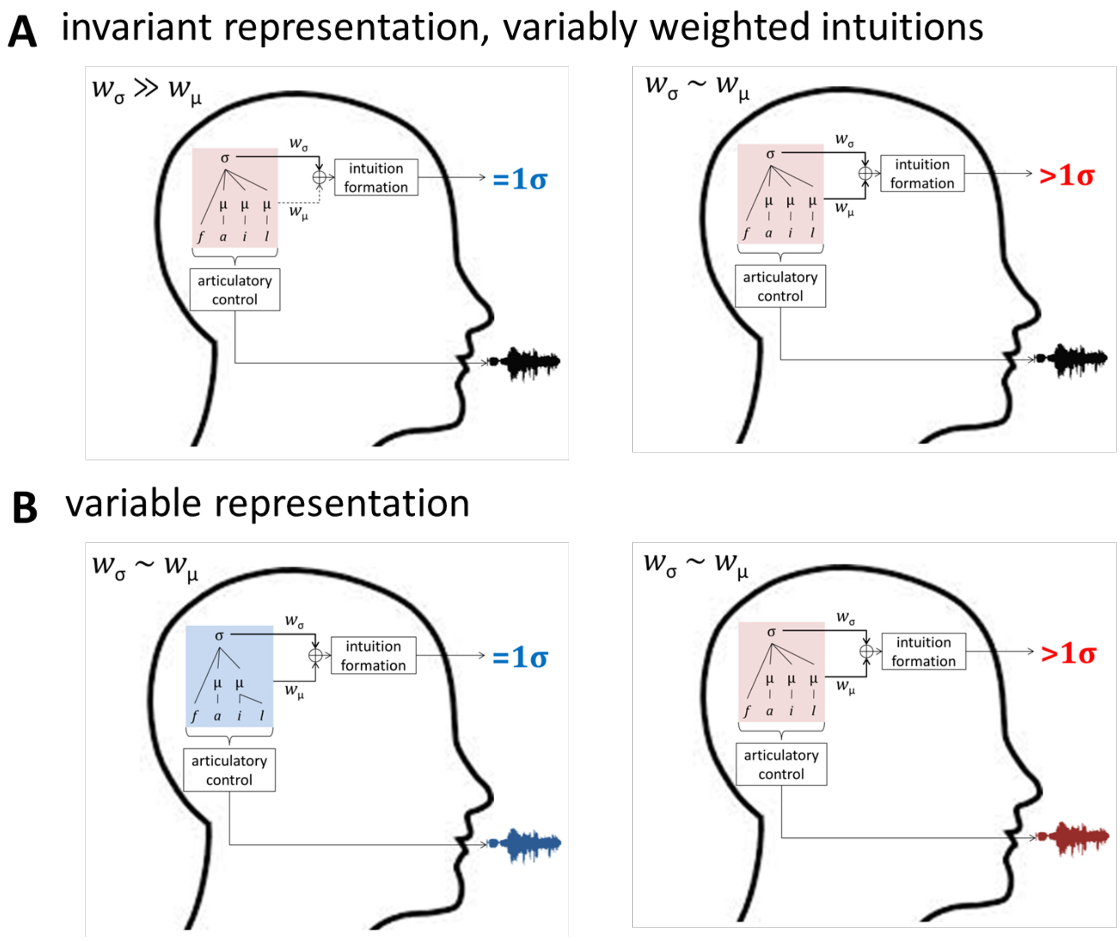

Fig. 10. Schematic models of inter-speaker and inter-word variation in σ-count judgments. (A) *Invariant representation, variably weighted intuitions*: all variable-count words have a trimoraic representation, and variation in σ-count judgments arises from differences in relative weighting of syllabic and moraic structure in the process of σ-count intuition formation; no correlation between production and σ-count judgments is predicted. (B) *Variable representation*: variable-count words may have a bimoraic or trimoraic representation; correlations between judgments and production are predicted.

Somewhat unanticipated was the relatively high degree of intra-speaker/word variation observed in the experiment. This "token-level" variation arises when a participant produces different judgments in the sequential and parallel tasks for a given word. Judgments were changed for 23% of /ail/ rimes and 30% of /air/ rimes. Both changes from >1σ to =1σ and from =1σ to >1σ were observed in both cases. Because there was no clear bias in the judgment reversals, they cannot be attributed straightforwardly to a task effect or task-order effect. For example, if >1σ judgments tended to increase in the parallel task, then this increase could be attributed to heightened awareness of structure in the parallel task, a stimulus repetition effect, or a cross-stimulus priming effect. The absence of a bias does not necessarily negate the possible role of such factors, but instead suggests that they were manifested in a speaker-specific fashion or interacted in a complex way. For example, some speakers may have begun with a bias toward =1σ judgments of /ail/ variable-count words but subsequently switched to >1σ judgments upon becoming aware of the ambiguity in /air/ rimes. At the same time, some of these speakers may have become biased toward =1σ in the parallel task due to a repetition effect. The interaction of these effects could result in the observed patterns.

One possible model of token-level variation would incorporate random or externally conditioned variation in the weighting terms in the invariant representation model (Fig. 10A). If the weighting terms of this model vary randomly from judgment to judgment, or are influenced by other factors (stimulus repetition, cross-stimulus priming, etc.), then token-level variation in σ-count judgments is possible. However, as already mentioned, the invariant representation model cannot account for judgment-production correlation. Incorporating variable weighting into the variable representations model (Fig. 10B) could account for token-level variation, but it predicts that only speakers with a trimoraic representation would exhibit such variation.

Token-level variation can be more readily modeled if speakers are allowed to have both representations or if variation in judgments and productions is associated with a continuous parameter dimension. The models in Fig. 12 assume that speakers potentially have both bimoraic and trimoraic representations. This allows for several interpretations of token-level variation. One possibility is that intuition-formation and articulatory processes are independent, each having their own weighting terms (Fig. 12A). Random or externally-driven variation in the intuition-weighting parameters ($w_{\mu\mu}$ and $w_{\mu\mu\mu}$) can account for token-level variation in judgments. Under this scenario, judgment-production correlations should only be observed when the articulatory weighting parameters ($a_{\mu\mu}$ and $a_{\mu\mu\mu}$) are correlated with the intuition weights.

A simpler possibility is that intuition-formation and articulatory control involve a shared mechanism and shared weighting terms (Fig. 12B). This could, for example, correspond to a model in which the process of forming a σ-count intuition involves a subvocal rehearsal of the word form. Our impression in producing our own σ-count judgments is that this is indeed the case: it seems that to produce a σ-count judgment, a speaker must engage the motor routines that they would use to produce the form. The shared process model thus allows for token-level variation with random or externally-driven variation in articulatory weighting parameters, which are assumed to determine the characteristics of both overt articulation and sub-vocal rehearsal. Judgment-production correlation is expected as long as the token-to-token variation in articulatory weighting parameters is not too extreme.

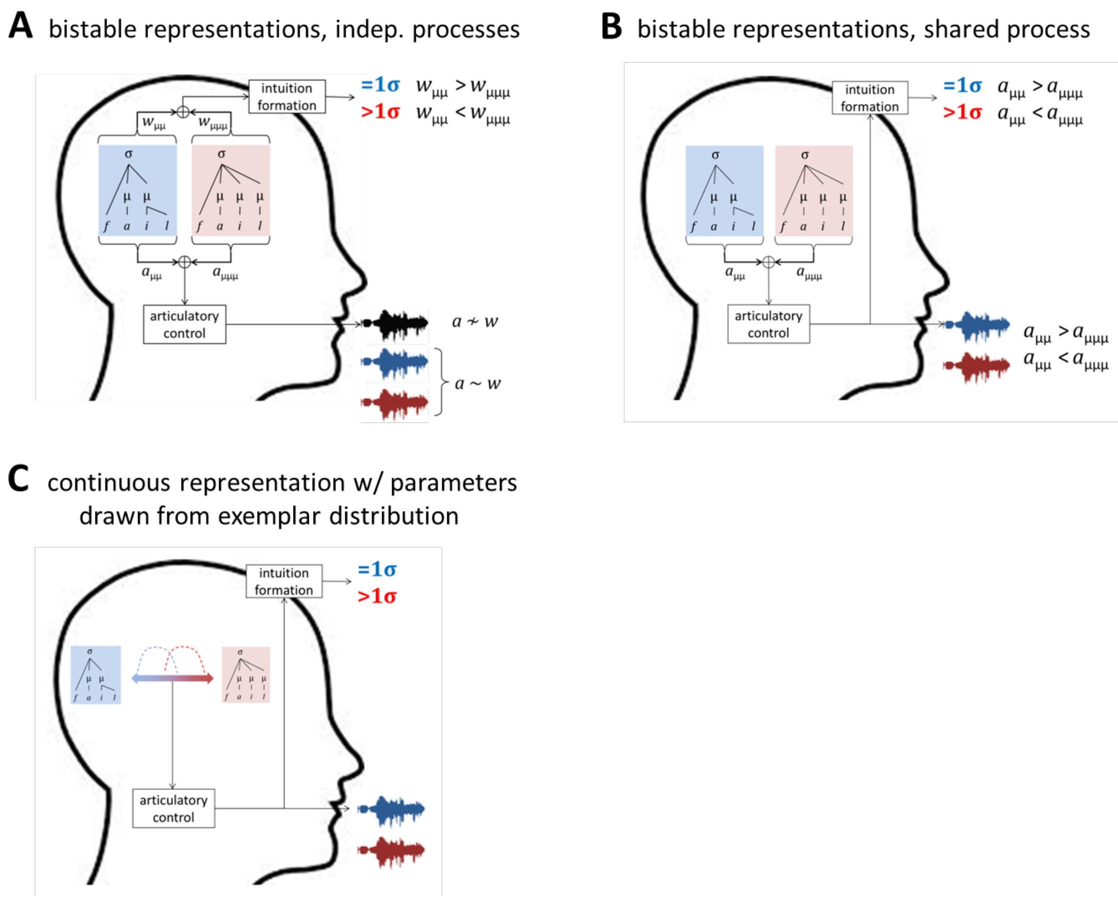

Fig. 12. Schematic models of intra-speaker/word variation. (A) speakers maintain two representations and judgments/production processes are independent with different weighting parameters; judgment-production correlation occurs when weighting parameters are correlated. (B) speakers maintain two representations but judgment and production rely on a shared process. (C) reconceptualization of (B) with continuous variation between representations.

A number of factors beyond structural representation plausibly play a role in influencing production and σ-count judgments of variable-count words. For one, the observation of a negative correlation between word frequency and proportion of >1σ judgments suggests that structural representations are influenced by language experience in a word-specific fashion. Consider also that a handful of speakers were excluded from the analysis because their σ-counts judgments were found to be unduly influenced by orthography. Furthermore, we have mentioned that stimulus repetition effects, cross-stimulus structural priming, and task-related factors may interact in a complex way to influence judgments and productions.

An exemplar-based perception and production model (Johnson, 1997; Pierrehumbert, 2001, 2002) with a continuous parameter dimension provides a useful framework for accommodating the above factors. In this model (Fig. 12C), bimoraic and trimoraic representations are viewed as category labels that are associated with endpoints of a continuous parameter dimension. A relevant parameter dimension in this case might be rime duration, which previous studies and our results indicate is an articulatory correlate of coda moraicity. When speakers engage the motor system—either for intuition formation or articulation—they randomly select a parameter value from a distribution of values that reflects their experience. This distribution can be modulated by contextual factors, such as recent memories of other tokens, and hence allows for cross-stimulus priming. Because exemplars of a given word include associations with orthography, word-orthographic variation can influence parameter selection. The parameter distribution is also influenced by the frequency of a word because more frequent words will have more exemplars in memory. While an exemplar-based model with shared articulatory and intuition formation processes provides the mechanisms to accommodate a wide array of factors that influence judgments and production, the question of which parameter(s) are most directly relevant remains unresolved.

A specific proposal regarding the articulatory manifestation of moraic structure developed in Tilsen (under review) holds that moraic codas are associated with a competitive regime of gestural selection, while non-moraic codas are associated with a co-selective, coordinative regime. In the competitive regime ({V}{C}), the vowel gesture and coda gesture in a rime compete for selection: the selection of the coda gesture is delayed until feedback is received regarding achievement of the preceding vocalic target. In the coordinative regime, the vowel and coda gestures are selected together ({VC}), and the relative timing of their initiation is controlled through phasing mechanisms. Between these two prototypical regimes exists a continuum corresponding to varying degrees of feedback internalization. This model predicts that rime durations associated with a competitive {V}{C} control regime will be longer than rime durations associated with a coordinative {VC} regime, because the vocalic and consonantal gestures in the coordinative regime will overlap to a greater extent. Hence the model predicts that a rime in which a liquid coda is moraic will be longer in duration than a rime in which a liquid coda shares a mora with the preceding vowel. The continuous parameter associated with the observed variation in the current experiment is thus hypothesized to be the degree of internalization of control over the liquid coda gesture: the extent to which control is internalized determines the articulatory phasing of the vocalic and coda gestures, their degree of overlap, and thus the duration of the rime.

The current experiment found substantial variation in σ-count judgments across speakers and words, along with a correlation between judgments and rime durations. These observations support the hypothesis that sub-syllabic structure influences production, and provides constraints on models of σ-count judgments. Specifically, the judgment-production correlation requires variation in sub-syllabic representation—not merely variation in the intuition formation process—and suggests that the process of forming an intuition involves engaging a motor representation in some fashion. The observation of token-level variation suggests that speakers may maintain both representations, and perhaps that these are related by a continuous parameter dimension. A better understanding of the factors influencing σ-count intuitions and production of variable-count word forms should be sought through future studies, because understanding these factors will shed light on the nature of phonological representations. The current findings ultimately highlight the importance of studies that address the relation between cognitive processes and representations in behavioral tasks which involve explicit judgments of phonological structure.

References

Broselow, E., Chen, S.-I., & Huffman, M. (1997). Syllable weight: convergence of phonology and phonetics. *Phonology*, *14*(1), 47–82.

Cohn, A. (2003). Phonological Structure and Phonetic Duration: The Role of the Mora. *Working Papers of the Cornell Phonetics Laboratory*, *15*, 69–100.

Duanmu, S. (1994). Syllabic weight and syllabic duration: A correlation between phonology and phonetics. *Phonology*, *11*(1), 1–24.

Ham, W. H. (2001). *Phonetic and phonological aspects of geminate timing*. Routledge.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. *Talker Variability in Speech Processing*, 145–165.

Lavoie, L., & Cohn, A. (1999). Sesquisyllables of English: the structure of vowel-liquid syllables. In *Proceedings of the XIVth International Congress of Phonetic Sciences* (pp. 109–112). Retrieved from http://conf.ling.cornell.edu/pdfs/Lavoie_Cohn_1999.pdf

Pierrehumbert, J. (2001). lenition and contrast. *Frequency and the Emergence of Linguistic Structure*, *45*, 137.

Pierrehumbert, J. (2002). Word-specific phonetics. *Laboratory Phonology*, *7*, 101–139.

Tilsen, S. (under review). The developmental origin of hierarchical phonological structure.

Department of Linguistics
Cornell University
Ithaca, NY 14850
tilsen@cornell.edu
acc4@cornell.edu

Appendix

Table A.1 Stimuli

word	# of syllables	CELEX log(frequency)	graphemes	word	# of syllables	CELEX log(frequency)	graphemes	word	# of syllables	CELEX log(frequency)	graphemes
Target words				*Non-target words*							
bee	1	0.8	3	disc	1	0.9	4	agent	2	1.6	5
fee	1	1.1	3	pass	1	1.3	4	today	2	1.0	5
pea	1	0.3	3	dock	1	1.0	4	given	2	1.5	5
bead	1	0.3	4	guest	1	1.4	5	sunny	2	1.0	5
feed	1	1.7	4	watch	1	1.6	5	piggy	2	1.3	5
bean	1	0.6	4	scent	1	1.1	5	widow	2	1.1	5
teen	1		4	queen	1	1.7	5	study	2	2.0	5
peel	?	0.8	4	quick	1	1.8	5	bacon	2	1.2	5
feel	?	2.6	4	cause	1	1.9	5	exist	2	1.2	5
veal	?	0.7	4	thank	1	1.4	5	stony	2	0.8	5
beer	?	1.7	4	youth	1	1.8	5	honey	2	1.3	5
fear	?	2.2	4	sight	1	2.0	5	giant	2	1.5	5
pier	?	0.7	4	pants	1	1.2	5	gonna	2	0.9	5
bid	1	1.0	3	faint	1	1.4	5	tummy	2	1.0	5
vid	1		3	bunch	1	1.2	5				
pin	1	1.1	3	dough	1	1.0	5				
bin	1	0.7	3	bound	1	1.7	5				
fin	1	0.6	3	sweet	1	1.6	5				
pill	1	1.1	4	waste	1	1.6	5				
bill	1	1.7	4	beast	1	1.2	5				
fill	1	1.6	4	shape	1	1.8	5				
pa	1	1.1	2	steak	1	0.9	5				
bah	1		3	ego	2	1.0	3				
fa	1	0.6	2	icy	2	1.0	3				
pod	1	0.3	3	via	2	1.3	3				
bod	1		3	any	2	1.7	3				
bon	1		3	copy	2	1.5	4				
Von	1		3	oven	2	1.3	4				
pall	1	0.3	4	busy	2	1.7	4				
ball	1	2.0	4	tiny	2	1.9	4				
fall	1	2.0	4	poet	2	1.2	4				
doll	1	1.2	4	mama	2	1.0	4				
par	1	1.0	3	easy	2	1.0	4				
bar	1	1.8	3	bias	2	1.0	4				
far	1	2.7	3	exam	2	0.9	4				
pie	1	1.1	3	cafe	2	1.3	4				
buy	1	2.1	3	unto	2	0.9	4				
vie	1		3	taxi	2	1.4	4				
bide	1		4	exit	2	1.0	4				
tide	1	1.4	4	unit	2	1.8	4				
pine	1	1.1	4	menu	2	0.8	4				
fine	1	2.2	4	acid	2	1.2	4				
vine	1	0.5	4	navy	2	1.2	4				
pile	?	1.4	4	tidy	2	0.9	4				
bile	?	0.3	4	echo	2	0.8	4				
vile	?	0.6	4	ocean	2	1.4	5				
file	?	1.5	4	event	2	1.8	5				
pyre	?		4	ahead	2	2.0	5				

fire	?	2.2	4	muddy	2	1.0	5
tire	?	1.0	4	china	2	1.1	5