Laboratory Phonology
Journal of the Association for Laboratory Phonology

Tilsen, S and Cohn, A C 2016 Shared Representations Underlie Metaphonological Judgments and Speech Motor Control. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1): 14, pp. 1–33, DOI: http://dx.doi.org/10.5334/labphon.52

JOURNAL ARTICLE

Shared Representations Underlie Metaphonological Judgments and Speech Motor Control

Sam Tilsen and Abigail C. Cohn
Department of Linguistics, Cornell University, US
Corresponding author: Sam Tilsen (tilsen@cornell.edu)

Researchers often use metalinguistic judgments to investigate phonological representations. The representations are assumed to govern speech motor control and thereby shape articulatory and acoustic characteristics of speech. Yet little is known about the relationship between metalinguistic judgments, phonological representations, and motor control. This paper reports on an experiment that directly investigates the relation between metalinguistic judgments and articulatory control, hypothesizing that the two share a common representation. This hypothesis predicts that differences in judgments should be correlated with differences in the acoustic characteristics of responses. An experiment was conducted in which syllable count judgments and productions of words with tense vowel/diphthong nuclei and liquid codas were obtained from native speakers of English. A subset of these words have previously been shown to exhibit variation in syllable count judgments. Acoustic analyses of productions showed that rime durations and formant trajectories differed between words associated with monosyllabic vs. disyllabic syllable count judgments. These results support the hypothesis that a common representation is utilized by the processes responsible for metaphonological judgments of syllable count and speech motor control.

Keywords: syllables; motor control; speech planning; metalinguistic judgments; liquids

1 Introduction

Metalinguistic judgments, or more specifically "metaphonological" judgments, are a commonly used tool for exploring the nature of phonological representations. It is typically assumed that such judgments utilize the same representations or are derived from the same mechanisms as those that are involved in speech production. Despite the pervasiveness of this assumption, connections between metalinguistic judgments and speech behavior have rarely been explicitly demonstrated. Here we present evidence that syllable count judgments (a metalinguistic task) utilize a representation that also governs articulatory control. We are concerned specifically with a subset of English words having a tense vowel or diphthong nucleus and a liquid coda (e.g., *peel*, *file*, *pier*, *fire*), which have been shown to exhibit variation in syllable count judgments. An experiment was conducted in which syllable count judgments and productions of words with liquid codas were obtained from native speakers of English. Acoustic analyses of responses showed that rime durations and formant trajectories differed between words associated with monosyllabic vs. disyllabic syllable count judgments. These results support the hypothesis that a common representation is utilized by the processes responsible for metaphonological judgments of syllable count and speech motor control.

1.1 Metalinguistic judgments and representations

Metalinguistic judgment tasks, such as lexical decision, wordlikeness judgments, explicit syllabification, stress placement, syllable counting, etc., have been used extensively to adduce characteristics of phonological representations. In this paper the focus is on phonological structure, and as such the experimental tasks of interest might be more specifically termed "metaphonological". What makes an experimental task "meta-" in our view is that it requires the use of explicit memory for linguistic representations, i.e., participants consciously use their linguistic knowledge to perform some task, which often involves an overt decision or judgment regarding some stimulus. Explicit memory tasks can be contrasted with implicit or procedural memory tasks, in which participants need not be consciously aware of the representations used in the task. Direct elicitation of speech is an example of a task using implicit memory, as is a recognition memory test in which participants decide whether they have heard a stimulus previously. Note that in some cases it may be difficult to draw a clear distinction between the use of implicit and explicit memory in a given task.

An obvious challenge in interpreting the results of meta-tasks is to avoid presuppositions about the status of the representations being investigated. It is always a troubling possibility that the task imposes upon participants representational distinctions with little relevance to the production or perception of speech (see Côté & Kharlamov [2011] and Derwing & Eddington [2014] for similar arguments). For example, syllabification tasks commonly presuppose that all consonants are uniquely associated with a single syllable. In the case of a VCV syllabification task, participants are forced to choose between the alternatives [V][CV] or [VC][V]. Treiman and Danis (1988) conducted an oral VCV syllabification task in which speakers produced disyllabic words with the order of syllables reversed, along with a written task in which speakers circled one of two choices. In both versions of the task, they observed substantial variation across speakers/words, with effects of orthography, stress pattern, and phonetic properties of the segments. Further research on VCV syllabification has replicated these findings and implicated word-level phonotactics as well (Eddington et al., 2013a, 2013b; Elzinga & Eddington, 2014). Yet it is not entirely clear what these findings tell us about representations vis-à-vis the production or perception of speech. Some theories of representation allow for ambisyllabic structures (e.g., Kahn, 1976), where an intervocalic consonant can be simultaneously associated with a preceding and following vowel. In that case, the forced choice between [V][CV] or [VC][V] in a syllabification task can be viewed as imposing a structure that differs from the one involved in normal speech.

Another example of the ambiguity involved in interpreting meta-tasks can be found in Frisch and Zawaydeh (2001), where wordlikeness judgments were obtained from Arabic speakers for novel Arabic verbs. Some of the novel words violated a hypothesized constraint against the co-occurrence of consonants with the same place feature in a root. The stimuli with violations were judged as less word-like than those which did not violate the constraint, and the judgments were not entirely predictable from the statistics of consonant co-occurrence in the lexicon. The authors therefore argued that the findings provide evidence that an abstract constraint is a psychologically real component of linguistic competence. Yet this conclusion begs the question of precisely what role is played by the constraint in normal speech behaviors. If wordlikeness judgments are viewed as a function of analogical mechanisms and lexical statistics, what evidence is there that these same mechanisms are relevant to the production or perception of speech in typical contexts?

Ultimately if firm conclusions are to be drawn from meta-tasks, an independent test is required to demonstrate that a metalinguistic judgment is correlated with a more implicit behavior. A good example of this is a wordlikeness study conducted by Frisch, Large, and Pisoni (2000). In the third experiment of that study, a wordlikeness judgment and a recognition memory task were conducted sequentially with English non-words. The number of syllables and sub-constituent probabilities of the non-word stimuli predicted their wordlikeness ratings, and crucially, the stimuli rated as more wordlike were more accurately recognized as previously heard in the recognition task. This finding indicates that the representational characteristics influencing wordlikeness judgments (i.e., number of syllables and constituent probability) are not solely used for the formation of wordlikeness intuitions, but also play an important role in the storage and retrieval of words from memory. One potential objection to this interpretation is that the meta-task of the wordlikeness judgment preceded the recognition task and so may have affected the outcome of the recognition task; ideally an independent test of the implicit behavior should precede the meta-task to avoid that potential interference.

The fundamental source of the difficulty in interpreting metalinguistic tasks is the assumption that the representations and processes governing the meta-task are also involved in the routine behaviors of normal speech production and/or perception. It is this assumption that the current study aims to test directly, focusing on metaphonological intuitions and articulatory planning/motor control. As a starting point for discussion, **Figure 1** contrasts a hypothesized model of the relation between metaphonological intuition-formation and articulatory planning/control with a null model in which the two are unrelated. Assume for expository purposes that speakers have two co-existing representations, X and Y, which may be differentially weighted. The model in **Figure 1A** depicts the situation in which the same representation underlies the processes of intuition-formation and articulatory control. This predicts that phonetic characteristics of production may correlate with metaphonological judgments. To be more specific, the intuition-formation process in **Figure 1A** might be interpreted as the result of a subvocal rehearsal mechanism. Because the subvocal rehearsal utilizes the same representation as an overt production, a correlation between intuitions and articulation is predicted by the model.

While the hypothesis of shared representation is the implicit assumption of research involving meta-tasks, it has rarely been tested directly. The corresponding null hypothesis

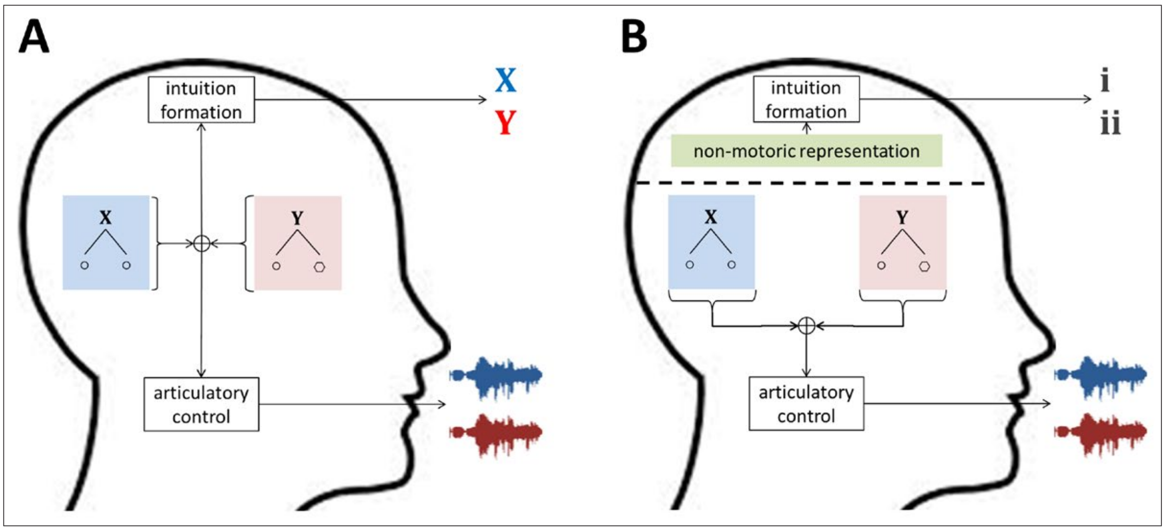

Figure 1: Alternative hypotheses regarding the relation between metaphonological intuitions and articulatory processes. **(A)** intuitions and control processes utilize a shared representation, and thus a correlation is predicted. **(B)** intuition-formation and articulatory control processes are not related, so no correlation between judgments and speech is predicted.

is shown in **Figure 1B**: representations X and Y influence articulation, but an alternative, non-motoric representation determines metaphonological intuitions (this could be, for example, a representation from sensory domains, or a multimodal representation constructed on-the-fly with analogical mechanisms). The null hypothesis predicts no correlation between intuitions and production, because the two are based on distinct representations.

The specific focus of the current experiment is on representations of subsyllabic structure and syllable count judgments. The aim was to test the null hypothesis that syllable count intuition-formation and articulatory control use independent representations (**Figure 1B**) against the hypothesis that syllable count intuitions and articulation share a common representation (**Figure 1A**). To conduct this test, correlations between syllable count intuitions and phonetic aspects of productions were examined, which required word stimuli for which syllable count intuitions can vary. The special class of rimes described below provides exactly the right sort of stimuli for this test.

1.2 Structural variation in liquid rimes

For most monomorphemic words in English, native speakers have robust, consistent intuitions regarding the number of syllables that comprise the word. These intuitions are robust even when syllabification (i.e., the mapping of segments to syllables) is unclear: words like *water* and *apple* are judged as disyllabic regardless of ambiguity in syllabification. The robustness of such intuitions seems to validate the notion that speakers use only syllable-level representations 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 across speakers in 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 studies have demonstrated interspeaker variation in σ-count judgments in variable-count words. 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 1½ 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 web-based 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 report 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

	pyre	***hire***	***fire***	***liar***	**TOTAL (σσ)**
wordcalc.com	–	σ	σ	σ	0
howmanysyllables.com	σ	σ	σ	σσ	1
poetrysoup.com	σ	σ	σσ	σσ	2
syllablecount.com	σ	σ	σσ	σσ	2
TOTAL # of σσ	0	0	2	3	

Table 1: Syllable counts reported by online syllable counters for selected words with diphthong-/r/ rimes. σ: 1 syllable, σσ: 2 syllables.

sequences are mapped to syllables, and hence those decisions may represent speaker intuitions at least indirectly.[1]

Cohn (2003) and Lavoie and Cohn (1999) pointed out a relation between the phenomenon of variable σ-count intuitions and mora-level representations. Specifically, they observed that variable-count words can be analyzed as having a trimoraic ("superheavy") syllable structure, and suggested that the origin of the disyllabic or greater-than-monosyllabic intuitions may lie in the subsyllabic structural organization of variable-count words. Whether the notions of "bimoraic" and "trimoraic" representation are the correct analyses of the relevant subsyllabic structural differences is somewhat tangential to the current focus, and the approach taken here does not presuppose any specific version of moraic theory, nor require a commitment to moraic theory in general. Rather, the crucial point is to differentiate between syllable-level organization and subsyllabic organization, allowing for subsyllabic organization to vary with the articulatory composition of a syllable. Hence "moraic structure" is used here in a generic sense, i.e., as structure that organizes segments or gestures within a syllable, and "bimoraic" organization is assumed to differ from "trimoraic" organization. Although a more concrete understanding of the nature of the structural difference is ultimately desirable, the first question that must be resolved is whether structural differences hypothesized to underlie the meta-task of σ-counting are indeed manifested in articulation.

The notion that subsyllabic/moraic structure has consequences for articulation has been supported by a variety of studies (cf. Cohn, 2003, for a review). For example, Broselow, Chen, and 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. 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 subsyllabic/moraic structure, it remains unknown whether structurally-conditioned variation occurs between speakers of the same language or between words with identical segmental content. Lavoie and Cohn (1999) observed some suggestive evidence that the structural configuration associated with $>1\sigma$

[1] Information regarding the dialects of speakers who designed the syllable counters is not available.

judgments has phonetic consequences for articulation. They examined the durations of variable-count and non-variable-count words with liquid codas for two speakers and 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. Although this finding is suggestive, in order to establish the relevance of the σ-counting meta-task to normal production, a correlation between judgments and acoustic measures of productions must be demonstrated, ideally within a large pool of speakers in which σ-count judgments vary.

1.3 Hypotheses

The current experiment conducted sequential and parallel production and σ-counting tasks with a large sample of native speakers of English. Stimuli were variable-count words along with unambiguously monosyllabic and disyllabic controls. The sequential task involved production of all words in the stimulus set, followed by the elicitation of σ-count judgments. The parallel task involved the elicitation of a σ-count judgment for a given word, followed immediately by a production of that same word. All participants performed the sequential task first, then the parallel task. Hence in the productions of the sequential task, 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. In both tasks, speakers were explicitly instructed to subvocally rehearse the stimuli to provide a basis for their σ-count judgments. The primary hypothesis of the study is as follows:

> Hyp. 1: *σ-count judgments and articulatory control are derived from a shared representation*. This hypothesis predicts that phonetic characteristics of variable-count forms, i.e., rime durations and formant trajectories, will vary as a function of σ-count judgments. Specifically, on the basis of previous findings regarding subsyllabic structure and duration, rime durations associated with >1σ-count judgments will be longer than rime durations associated with =1σ judgments. Formant trajectories associated with >1σ judgments will reflect delayed timing of the liquid gesture relative to the vocalic gesture, along with relatively less coarticulation between the liquid and vocalic gestures.

The corresponding null hypothesis is that distinct representations and/or processes are responsible for σ-count intuitions and articulation, which predicts that rime durations and formant trajectories will not vary as a function of σ-count judgments. The "shared representations" hypothesis is an instantiation of the more general hypothesis that metalinguistic intuitions and articulatory control processes share a common representation. The predictions regarding formant trajectories are detailed in Section 4.1, where a more specific interpretation of the relation between syllabic structure and articulatory control is presented.

It is important to clarify that the shared representations hypothesis does not presuppose that the only relevant aspect of representation is rime class, i.e., an abstract phonological category. Rather, given the wealth of evidence that phonological representations are sensitive to the lexical structure and statistics of usage, it seems plausible that both word-specific information and abstract phonological categories shape the relevant aspects of the representation. Therefore in general, the 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 a particular rime structure, then the relation between judgments and rime durations can be assessed in a by-speaker analysis. If σ-count judgments vary by word for a given speaker and rime 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-items analysis.

A secondary hypothesis regarding the effects of attention to structure was also tested. In the sequential task, speakers produce words without any attention being drawn to syllable count—indeed, without any knowledge that they will subsequently make σ-count judgments. In the parallel task, attention to σ-count is explicitly juxtaposed with the production of words. Hence if the processes of σ-count intuition formation and articulatory control have the potential to share a common representation, the heightened attention to structure in the parallel task might be expected to increase the likelihood that these processes will indeed share that representation, thus strengthening the correlation between them. This leads to the following hypothesis:

> Hyp. 2: *Attention to structure strengthens the relation between judgments and articulation.* This hypothesis predicts that the effects of σ-count judgments on rime durations and formant trajectories will be greater in the parallel task than in the sequential task.

The structural attention hypothesis assumes that the meta-task of σ-count judgment and articulatory control processes have at least the potential to utilize a shared representation, and holds further that temporal juxtaposition of the two should increase the concordance between them. It should be noted the absence of this effect might be interpreted in two ways. One possibility is that the processes of σ-count intuition formation and production are so tightly integrated that their correlation will reach a ceiling in the sequential task and thereby exhibit no augmentation in the parallel task. The other possibility is that conscious awareness of structure does not modulate the extent to which a shared representation influences production. The following section presents the methods employed to test the above hypotheses.

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. Six of the participants were excluded for reasons discussed in Section 2.3. Participant ages were in the range of 18–29 (median 20 years old). 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. All procedures were performed in compliance with institutional guidelines and approved by the Cornell Institutional Review Board. During the experiment, participants were seated in a sound-attenuating booth in front of a computer monitor and wore a head-mounted microphone.

The experimental session was organized into three phases, as schematized in **Figure 2**. 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 to hesitate before or after the word, not to emphasize the word that goes in the blank, and to "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 with a

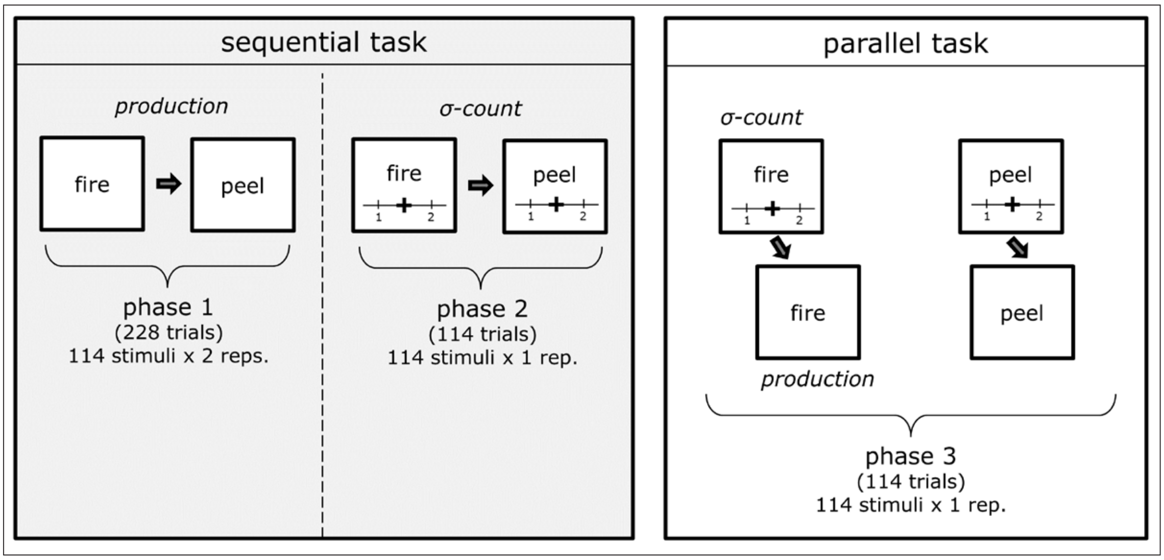

Figure 2: Sessions were organized into three phases: production, σ-count judgment, and production with σ-count judgment. The first two phases comprise the sequential task; the third phase is the parallel task.

dummy item how to produce the phrase without any major phrase-internal intonational breaks. The entire stimulus set was produced twice in the first phase (114 stimuli × 2 reps = 228 trials). In all phases, stimuli were presented in a pseudo-randomized 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.

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 syllables, σ-count judgments were elicited on a continuous scale with a mouse-guided pointer. The scale ranged from 0.5 to 2.5 and the values 1 and 2 were labeled with tick marks (see **Figure 2**). 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 up to 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 say 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 (the same instructions were given in the third phase, described below). 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 recently produced two judgments (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

Two sets of stimuli, targets and non-targets (fillers), were created for the experiment. Target stimuli included 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. On the basis of their phonological rimes, 13 of the 50 target stimuli are expected to be variable-count forms (**Table 2**, bolded words), while the remainder are unequivocally monosyllabic. No morphologically complex stimuli were included, and all of the variable-count stimuli were required to be 4 graphemes long. 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/.

Note that the lax vowel /ɪ/ does not occur in open monosyllables and that the tense/lax contrast in high front vowels is merged before /r/. There is some ambiguity as to 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 as this is consistent with the intuitions of the first author. Note also that the words *pall*, *ball*, and *fall* may have the vowel /ɔ/ or more commonly /a/, representing a merger that is increasingly characteristic of younger speakers, even in areas traditionally described as maintaining the /ɔ/-/a/ contrast.

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 varied, resulting in heterographic representations of homophonic

	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	par bar far
ai	pie buy vie	bide tide	pine fine vine	**pile** **bile** **vile** **file**	**pyre** **fire** **tire**

Table 2: Target stimuli include 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). Variable-count stimuli are in bold.

vocalic nuclei (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% 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 reduced 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 in the CELEX log-frequency 25–75% range.

2.3 Data processing and analysis

Despite the availability of a continuous response dimension in the σ-counting task, experiment-wide participant responses were highly multimodal. Modes near 1 and 2 are expected across the experiment because the stimuli included unambiguously monosyllabic and disyllabic words. The majority of participants exhibited 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.

Figure 3 illustrates the results of this procedure for three representative patterns of within-participant response distribution. **Figure 3**(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 just a few of the participants, in which intermediate values are more widely distributed rather than associated with a single intermediate mode. All subsequent analyses treat σ-count judgments as representing either 1 syllable ($=1\sigma$, i.e., belonging to the monosyllabic bin) or more than 1 syllable ($>1\sigma$, i.e., belonging to the intermediate or disyllabic bins).

Data from 6 of the 34 participants were excluded from subsequent analyses because these participants produced a high proportion (>10%) of nonstandard σ-count judgments for unequivocally monosyllabic and disyllabic words. The 10% criterion results in the same set of excluded participants regardless of whether all unequivocal stimuli or only non-target ones are considered. **Table 3** lists the excluded participants with >10%

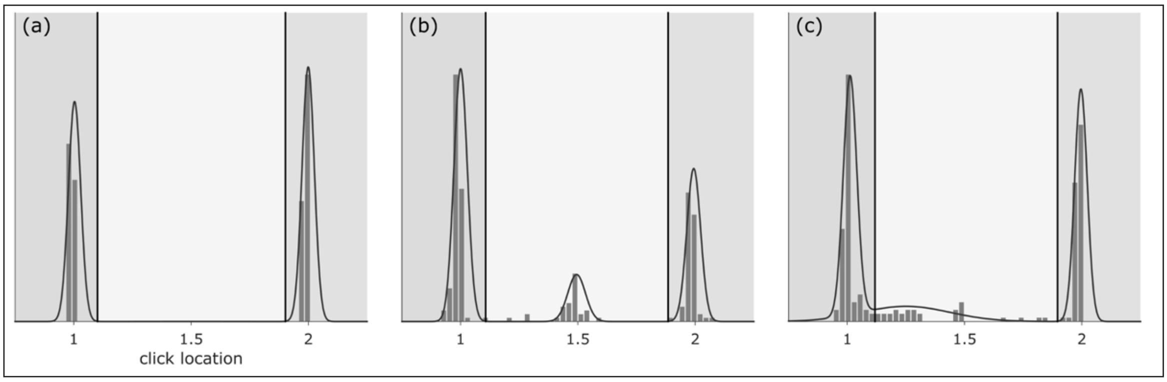

Figure 3: Three representative patterns of σ-count judgment distributions, bimodal/trimodal fits to density functions, and categorical partitioning. Vertical lines delineate monosyllabic, intermediate, and disyllabic bins. **(a)** bimodal response pattern; **(b)** trimodal response pattern; **(c)** bimodal response pattern with relatively widely distributed intermediate responses.

	all stimuli			non-target only			target only		
	non-standard	grapheme effects (p-value)		non-standard	grapheme effects (p-value)		non-standard	grapheme effects (p-value)	
participant	rate	over	under	rate	over	under	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 proportions of nonstandard judgments, along with *p*-values from a χ^2 test on the effect of grapheme length on the likelihood of over- and under-counting nonstandard judgments (values only shown when nonstandard judgments occur).

nonstandard judgment rates for all unequivocally monosyllabic or disyllabic stimuli, along with rates for target and non-target subsets separately. Of the 28 participants with nonstandard judgments rates lower than 10%, 19 produced 5 or fewer nonstandard judgments ($<2.5\%$) across the experiment.

The likely source of the nonstandard responses is over-reliance on grapheme length for syllable count judgments. Reliance on grapheme length is a viable 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 nonstandard judgments. An over-counting nonstandard judgment 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 nonstandard judgments 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 judgments and words with 3 graphemes were associated with under-counting judgments. The excluded participants not affected by grapheme length likely failed to attend closely to the experimental task.

Durations of acoustic intervals in productions of target words were identified automatically as follows. For each response, a vocalic energy amplitude envelope (Tilsen & Johnson, 2008) and sibilant energy envelope were obtained by lowpass-filtering (5 Hz cutoff) a bandpass-filtered version of the acoustic signal, with passbands of [80, 600 Hz] and [4000, 10000 Hz], respectively. Response onset/offset, along with the approximate midpoints of the alveolar fricatives and vowel [ei] in the carrier phrase (I [s][ei] — [s]ometimes) were located by identifying amplitude extrema in these signals. For each subject, trials with an outlying value (i.e., > ±2.0 s.d.) of any interval defined by these landmarks were identified and inspected by hand in Praat. The majority of the outliers were associated with trials in which speakers errorfully hesitated prior to the response item and restarted the response. When the restarted portion of the response included the [s]ay __ [s] interval (from *I say __ sometimes*) with no hesitation, the landmarks were adjusted by hand; otherwise the trial was marked as an error and excluded from subsequent analyses.

Rime durations in target responses were measured automatically with the following procedure. Vocalic and sibilant energy envelopes with greater temporal accuracy were calculated with lowpass filters having a higher cutoff frequency (25 Hz, 4th order Butterworth). The approximate midpoint of the rime vowel was identified as a peak in the vocalic energy envelope, and the onset of the rime vowel was located at the time of maximum velocity of vocalic energy preceding this peak. Inspection of these landmarks revealed a good fit to the onset of higher-harmonic energy visible in spectrograms. Offsets of open syllable and vowel-{l, r, n} rimes were located at the crossing point between normalized vocalic and sibilant energy envelopes that occurs prior to the sibilant energy peak associated with the post-target [s]. For vowel-[d] rimes the vowel offset was located at the vocalic energy velocity extremum associated with the decrease in energy in the V[d] transition, and the following [s] onset was located at the time of maximum sibilant energy velocity in the post-target [s]. Outlying rime durations were identified on a by-subject and by-rime basis, and inspected in Praat. Most of these were cases in which the vocalic onset had been errorfully located, and these were hand-corrected.

The relativized rime durations reported in Sections 3.4–3.6 are ratios obtained by dividing the duration of a rime for a given speaker/task by the mean duration of the corresponding open syllable rime for that same speaker in the corresponding task. In other words, the mean duration of /ai/ in open syllables for speaker X in the parallel task is the denominator of the ratio for tokens of /ail/ and /air/ from speaker X in the parallel task; likewise the mean /a/ duration in open syllables for speaker Y in the sequential task is the denominator for /al/ and /ar/ for speaker Y in the sequential task. 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.

Formant trajectories for open syllable and liquid rimes were calculated as follows. Raw F1, F2, and F3 trajectories were estimated using a robust LPC algorithm (Yao et al., 2010). These trajectories were time-normalized within subjects/vowels, and outlying points were excluded. Each trajectory was subsequently fit with a cubic smoothing spline. **Figure 9** in Section 3.7 represents the distributions of peaks in F2 trajectories for diphthong-lateral rimes. The trajectories, along with the measures of peak timing, are expressed in relativized time, analogous to the relativized duration described above. In other words, the relativized time is the raw time for a given token divided by the mean duration of the /ai/ formant trajectory for a given speaker. Normalized F2 rises are defined as the raw F2 rise from vowel onset to F2 peak, divided by the mean F2 rise in open syllable /ai/ for a given speaker. Note that inflection points in formant trajectories were not consistently present across speakers for vowels/formants other than F2 of [ail]/[air], and so the trajectory analyses conducted in Section 3.7 are confined to the diphthong-liquid rimes.

Details of statistical analyses are as follows: for the repeated measures ANOVAs of vowel duration in Section 3.3, the factors were Vowel (/i/, /ɪ/, /ai/, /a/), Coda (/r/, /l/), Word (nested within Vowel and Coda), and repeated measures over Subject, specified as a random factor. A separate ANOVA with Vowel level /ɪ/ excluded was used to estimate the Vowel × Coda interaction. For stepwise linear mixed effects regressions of rime durations in Section 3.4, the starting model included fixed effects of Rime, Task, syllable count Judgment, all interactions of these effects, and a random effect of Subject. The 3-way Rime × Task × Judgment interaction was removed first, then the Task × Judgment interaction, and lastly the main effect of Task.

3 Results

Analyses of correlations between σ-count judgments and phonetic properties of responses support the shared representations hypothesis. As predicted, responses associated with >1σ judgments had longer rime durations than those associated with =1σ judgments. In diphthong-liquid rimes, formant trajectories had later and higher F2 peaks that reflect later timing of the liquid coda gesture. Substantial interspeaker variation was observed in σ-count judgments of diphthong-liquid rimes (/ail/, /air/), and to a lesser extent in high-front/tense vowel-liquid rimes (/il/, /ir/). Within-speaker, word-specific variation in σ-count judgments was observed, as well as variation in σ-count judgments between the sequential and parallel tasks. Because effects were not consistently stronger in the parallel task compared to the sequential task, the structural attention hypothesis was not supported. To put the main results in context, we begin by reporting on the forms of variation observed in σ-count judgments.

3.1 Across-participant variation in σ-count judgments

Analysis of σ-count judgments revealed substantial interspeaker variation, particularly for the diphthong-liquid rimes. **Figure 4** 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. Specifically, 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. A similar pattern was observed for /air/ rimes. The /i/-liquid rimes exhibited less variation: all but a handful of the participants judged most of the /il/ and /ir/ rimes as =1σ.

Further analysis revealed that there was a moderate degree of consistency within participants with regard to σ-count judgments for a given nucleus. By-speaker proportions of >1σ judgments were correlated across rimes sharing a given nucleus (/ail/~/air/: $r = 0.69$, $p < 0.001$, $df = 26$; /il/~/ir/: $r = 0.68$, $p < 0.001$, $df = 26$). In other words, the by-speaker proportion of >1σ judgments for a given liquid coda accounts for about half of the variance for the other liquid coda with the same nucleus (i.e., $R^2 \approx 0.50$). In contrast, none of the correlations between rimes with different nuclei were significant ($p < 0.25$ for /ail/~/il/, /ail/~/ir/, /air/~/il/, /air/~/ir/). This disparity indicates that vowel nucleus was a more important factor than liquid identity (/r/ vs. /l/) in influencing σ-count judgments.

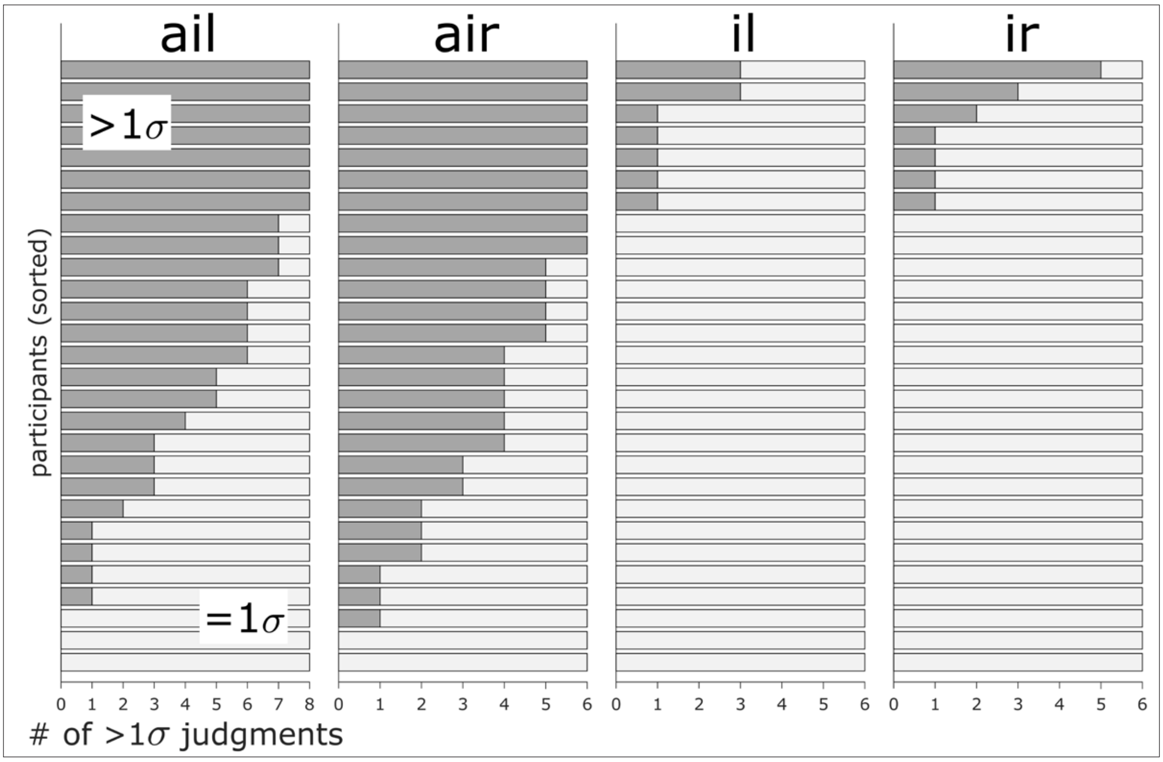

Figure 4: Counts of >1σ judgments for variable-count rimes by participant, sorted for each rime by within-participant proportion. The prevalence of intermediate counts for diphthong-liquid rimes indicates word- and/or task-specific variation.

3.2 Word- and task-specific variation in σ-count judgments

Analyses of word-specific variation in σ-count judgments showed a tendency for less frequent words to be more likely to receive >1σ judgments. **Figure 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 logistic regression of σ-count judgments with log-frequency as a predictor showed that word frequency was a significant factor in σ-count judgments ($t = 3.88$, $df = 670$, $p < 0.001$). A negative correlation was observed, i.e., lower frequency words were associated with a higher proportion of >1σ judgments.

An alternative source of the word-frequency effect could be graphemic composition, although we can only assess this alternative in the monophthong stimuli where both same-grapheme (*feel*, *peel*, *beer*) and mixed-grapheme (*veal*, *fear*, *pier*) nuclei occurred. A logistic regression of σ-count judgments with log-frequency and nucleus grapheme class (same- vs. mixed-) as predictors showed that indeed there was a significant effect of grapheme class on σ-count judgments in the monophthongs ($t = 2.15$, $df = 333$, $p = 0.03$). However, because this analysis is based on just three words in each grapheme class, and because there is a correlation between grapheme class and word frequency, caution should be warranted in inferring more general effects of graphemic composition on σ-count judgments. The same might be said for word frequency, where the analysis is based on 13 lexical items.

A majority of σ-count judgments were consistent within speaker across tasks (81%), but a substantial percentage of σ-count judgments changed for a given participant between the

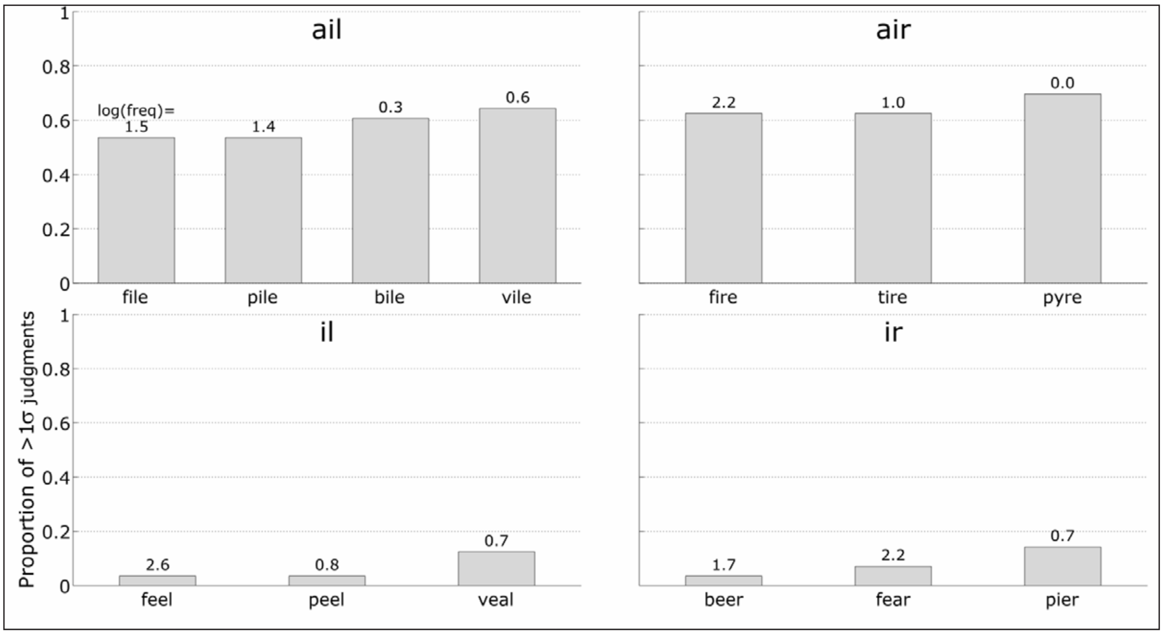

Figure 5: Experiment-wide proportions of >1σ judgments by word. CELEX log-frequency (per 1 million words) is shown above each bar.

rime	% of changed judgments	changed judgments	judgments per task	=1σ → >1σ changes	>1σ → =1σ changes
/ail/	23%	26	112	16	10
/air/	30%	25	84	9	16
/il/	13%	11	84	7	4
/ir/	10%	8	84	4	4
TOTAL	19%	70	364	36	34

Table 4: Percentage of judgment changes between tasks for each variable-count rime, and counts of changes in each direction from the sequential to parallel task.

sequential and parallel tasks (19%). **Table 4** shows the percentages 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.

The table shows that 23% and 30% of /ail/ and /air/ judgments changed between tasks, while changes were less frequent for /il/ and /ir/ rimes. Although judgment changes were not infrequent, no general trends are evident in the directions of the changes. To wit, in the diphthongal rimes there were 25 changes from $=1\sigma$ to $>1\sigma$ judgments and 26 changes in the other direction, and a similar balance in monophthongal rimes. Thus this result demonstrates the presence of variability in judgments, but does not support the hypothesis that attention to structure is a direct source of the variability.

3.3 Segmental 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. **Figure 6** and **Figure 7** 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 a repeated measures ANOVA of rime duration (Vowel: $F(3,4009) = 603.1$, $p < 0.001$; Coda: $F(4,4009) = 200.0$, $p < 0.001$; Word: $F(42, 3967) = 15.5$,

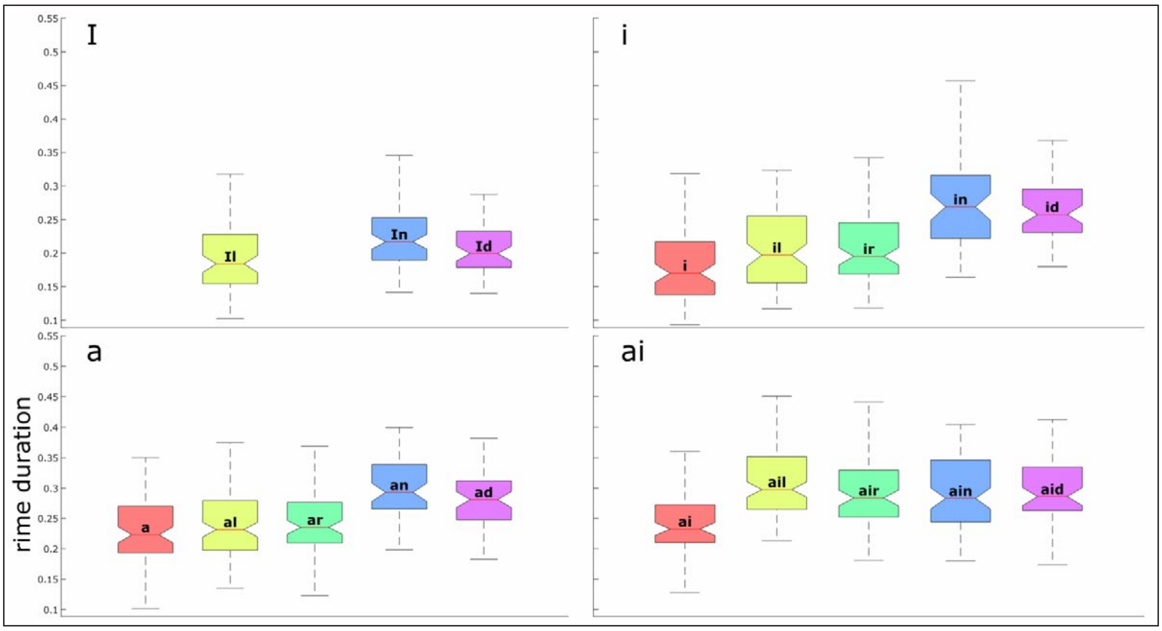

Figure 6: Rime durations from all tokens compared across codas within nuclei. Error bars show the range of data in the 5–95 percentile; boxes show range of data in the 25–75 percentile, and notches show ±2.0 standard error.

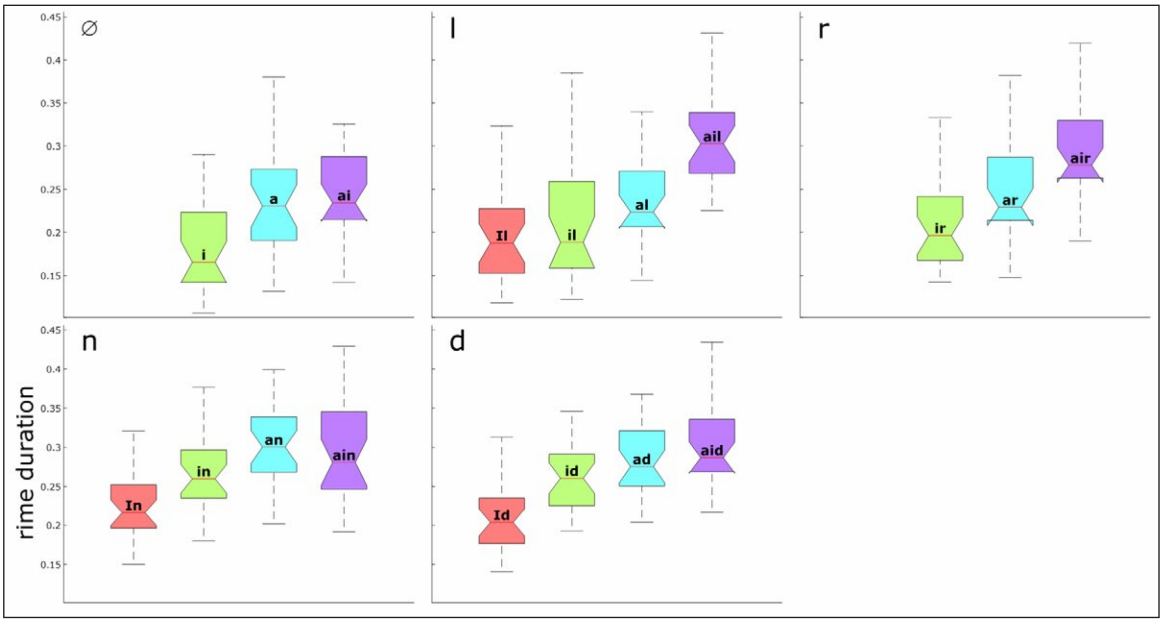

Figure 7: Rime durations compared across nuclei within codas. Error bars show the range of data in the 5–95 percentile, boxes show range of data in the 25–75 percentile, and notches show ±2.0 standard error.

$p < 0.01$; Vowel-Coda: $F(8,3359) = 43.3$, $p < 0.001$; note that the vowel-coda interaction effect was calculated in a separate repeated measures ANOVA with lax vowel /ɪ/ rimes excluded).

An important observation is that the liquid codas along with /n/ and /d/ contribute a substantial amount of duration to /ai/-nucleus rimes, resulting in rime durations that tend to be longer than those in the open syllable /ai/ (cf. **Figure 6**). In contrast, 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 **Figure 7**, 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.

3.4 Relation between rime duration and σ-count judgments

Analyses of rime durations supported the shared representations hypothesis: words associated with >1σ judgments were produced with greater rime durations than words associated with =1σ judgments. Two measures of rime duration were considered: (1) raw rime duration and (2) relativized rime duration (cf. Section 2.3). For both of these measures, linear mixed-effects regressions were performed to determine whether the effects of σ-count judgments on rime durations interacted with task and/or rime category.

Task effects and task-judgment interactions were not significant, whereas rime-judgment interactions were: likelihood ratio tests showed no significant improvement when an interaction between σ-count judgment and task was included (raw dur: $\chi^2 = 3.40$, $df = 1, p = 0.07$; rel. dur: $\chi^2 = 3.33$, $df = 1, p = 0.07$). Indeed, there was no improvement even when main effects of task were included (raw dur: $\chi^2 = 1.73$, $df = 1$, $p = 0.19$; rel. dur: $\chi^2 = 1.29$, $df = 1$, $p = 0.26$). In contrast, rime-judgment interactions significantly improved model fits for both measures (raw dur: $\chi^2 = 11.7$, $df = 3$, $p = 0.008$; rel. dur: $\chi^2 = 9.30$, $df = 3$, $p = 0.026$). The coefficient estimates for the main effect of judgment were 58 ms for raw rime duration ($t = 4.57$, $df = 708, p < 0.001$, 95% ci = [33, 83 ms]) and 0.29 for the relativized rime duration ($t = 4.46$, $df = 708$, 95% ci = [0.16, 0.42]). To assess the significance of effects of σ-count judgment within each of the four rime categories, two-sample *t*-tests were conducted for each rime category. **Figure 8** shows the *p*-values of these tests, along with boxplots of durations from each sample. A Bonferroni correction of $N = 4$ was used to adjust the significance threshold to $p = 0.0125$; equality of variance was not assumed.

For three of the four rimes there was a significant effect of σ-count judgment on relativized rime duration, and for /ir/ the effect was marginal ($p = 0.07$). Only in /il/ rimes was the effect significant for the raw duration measure, although marginal effects were observed for /air/ and /ir/ rimes. The ratio measures are preferred here because they better reflect the contribution of a liquid coda to rime duration. Although the task-judgment interaction did not exceed the significance threshold, for descriptive purposes **Table 5** presents the means, sample sizes, and standardized effect sizes within each rime and task for the relativized duration measure.

Additional analyses were conducted to assess the possibility that effects were driven by variation in speech rate or prosodic boundary structure. For example, it could be the case that speakers with >1σ judgments of a word use slower speech rates or higher-level phrase boundaries in producing the word, thus biasing rime durations in the observed directions. If so, effects of these confounds would be detectable in measures of the duration of the entire carrier phrase and/or the response-proximal portion of the phrase from the pre-target vowel to the post-target fricative (i.e., I s[ei _ s]ometimes). It is necessary to remove rime duration effects from this test, so rime durations were subtracted from the

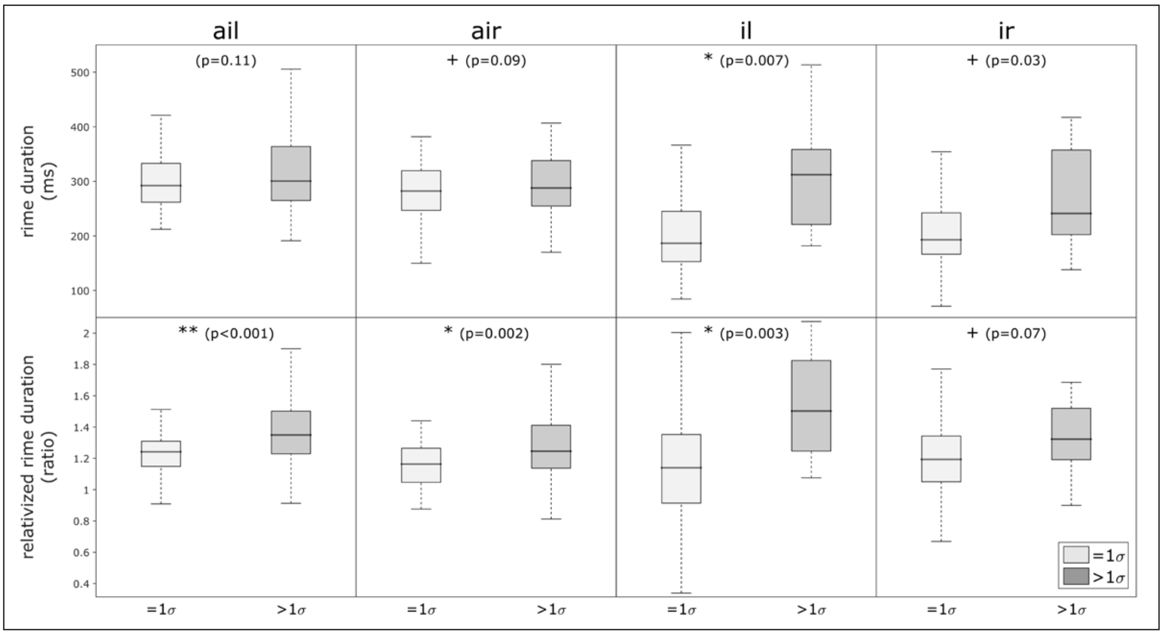

Figure 8: Effects of σ-count judgment on rime duration. (Top) Raw rime durations. (Bottom) Rime durations as a ratio of within-subject average open syllable duration. *P*-values are shown from two-sample *t*-tests: (**) $p < 0.001$, (*) $p < 0.0125$, (+) $p < 0.10$.

	sequential				**parallel**			
		mean				**mean**		
rime	***N* (=1σ, >1σ)**	**=1σ**	**>1σ**	**Cohen's *d***	***N* (=1σ, >1σ)**	**=1σ**	**>1σ**	**Cohen's *d***
/ail/	(50, 62)	1.225	1.344	0.84	(44, 68)	1.244	1.388	0.62
/air/	(26, 58)	1.176	1.250	0.48	(33, 51)	1.169	1.294	0.54
/il/	(80, 4)	1.186	1.438	0.40	(77, 7)	1.147	1.608	1.31
/ir/	(77, 7)	1.180	1.334	0.34	(77, 7)	1.207	1.338	0.76

Table 5: Sample sizes, means, and standardized effect sizes (Cohen's *d*) for the differences in relativized rime durations between syllable-count judgments for each rime and task.

phrase duration measures. Log-likelihood ratio tests of linear mixed effects models (with judgment-rime interactions included) showed no significant effect of σ-count judgment on the duration of the carrier phrase or response-proximal portion of the phrase (entire carrier phrase: $\chi^2 = 5.08$, $df = 4$, $p = 0.28$; response-proximal portion: $\chi^2 = 4.23$, $df = 4$, $p = 0.38$). Thus the observed effects on rime duration are unlikely to be an indirect consequence of differences in speech rate or prosodic boundary strength.

Although the patterns in rime durations strongly support the shared representations hypothesis, the structural attention hypothesis was not supported: task-count interaction effects were not significant, as shown above. This can also be seen by considering the Cohen's *d* effect sizes (cf. **Table 5**), which represent the standardized difference in sample means. The differences in effect sizes between tasks for the diphthong rimes are quite small. They are somewhat larger for the monophthongs, but these estimates are based on relatively few samples of >1σ judgments (7 or less), and so firm conclusions should not be drawn from them. Thus the results do not provide evidence for the hypothesis that attention to structure heightens the effects of shared representations.

By-word comparisons of the sizes of σ-count judgment effects on rime duration show that effects were not driven by a small subset of target words. **Table 6** shows standardized

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

Table 6: Effects of judgment on production by word. Difference between sample means ($\Delta = \mu_{[>1\sigma]} - \mu_{[=1\sigma]}$) and Cohen's *d* measure of effect size are shown for the raw duration and duration ratio measure. Words are sorted by effect size for raw duration.

(Cohen's *d*) and raw (Δ) effect sizes for raw and normalized rime durations in each word. The words are sorted by the effect size for the raw duration measure. Words with /i/ nuclei had the largest effect sizes, but the comparisons are based on fairly small sample sizes in the $>1\sigma$ group. Words with diphthong nuclei had more balanced distributions of $>1\sigma$ and $=1\sigma$ responses; in these words the standardized effect sizes ranged from 0.5 to 0.9. Because the effect sizes are similar across these lexical items, the observed relation between σ-count judgments and rime durations cannot be attributed to a subset of anomalous stimuli within the larger class of variable-count words.

3.5 Relations between σ-count judgments and formant trajectories

Further support for the shared representations hypothesis was observed in formant trajectory differences between $=1\sigma$ and $>1\sigma$ responses. Analyses of F2 peak timing and F2 rise showed that F2 peaks occurred later and had a higher F2 rise in $>1\sigma$ compared to $=1\sigma$ responses. Note that the F2 peak is indicative of an increase and subsequent decrease in the degree of palatal constriction in an [a]-[i]-liquid sequence, which predicts a rise and fall in the second resonance of the vocal tract due to the formation and release of a constriction near an antinode of that resonance. Thus the timing of the peak can be taken as an indirect reflection of the timing of the onset of the dorsal articulation (in the case of postvocalic /l/) or tongue root articulation (in the case of /r/).

ANOVAs of F2 peak timing and rise were conducted separately for /ail/ and /air/ rimes with main effects of task and σ-count judgment, and their interaction. None of the interactions were significant and hence an ANOVA with only the main effects was conducted. Task effects remained non-significant (/ail/ peak timing: $F = 0.30(1{,}275)$, $p = 0.58$; /ail/ rise: $F = 0.15(1{,}275)$, $p = 0.70$; /air/ peak timing: $F = 0.01(1{,}215)$, $p = 0.92$; /air/ rise: $F = 0.004(1{,}215)$, $p = 0.95$). In contrast, main effects of σ-count judgment were highly significant (/ail/ peak timing: $F = 9.9(1{,}275)$, $p = 0.002$; /ail/ rise: $F = 9.9(1{,}275)$, $p = 0.002$; /air/ peak timing: $F = 6.5(1{,}215)$, $p = 0.01$; /air/ rise: $F = 8.2(1{,}215)$,

F2 peak timing	**Δ%**	***t* (df) =**	***p*-value =**
/ail/	6%	2.2 (267)	0.030
/air/	14%	2.7 (165)	0.008
F2 rise			
/ail/	6%	9.9 (257)	0.002
/air/	9%	1.1 (160)	0.005

Table 7: Effect sizes (Δ) as a percentage of open syllable F2 peak timing/rise and *p*-values from *t*-tests comparing normalized F2 peak timing and rise between σ-count groups for each diphthongal rime.

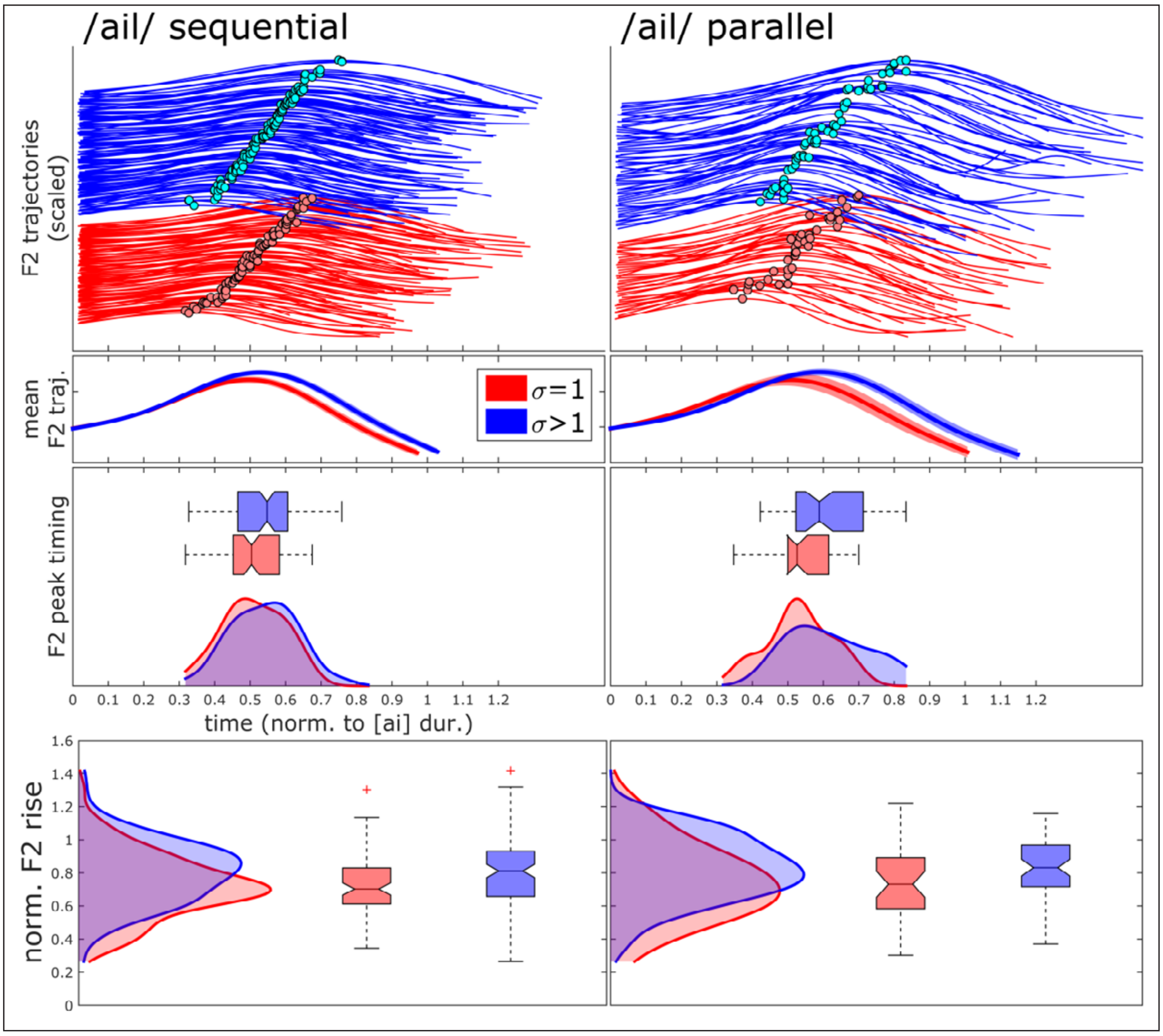

Figure 9: Comparisons of F2 peak timing and F2 rise in /ail/ responses by σ-count judgment. From top to bottom: F2 trajectories for responses associated with =1σ and >1σ judgments (sorted by F2 peak timing and scaled in amplitude for illustration); mean F2 trajectories ±2.0 s.e. normalized within response category; distributions of normalized F2 peak timing and boxplots; distributions of normalized F2 rise and boxplots.

p = 0.005). The results of two-sample *t*-tests (equal variance not assumed) for each rime and measure are summarized in **Table 7**, and trajectories of /ail/ are shown in **Figure 9** for reference. The figure shows that the F2 peak occurs later in responses associated with >1σ judgments than those associated with =1σ judgments. Although these effects were modest in size—6% and 14% of open syllable peak timing for /air/ and /ail/, respectively—they were statistically significant. The effects on F2 peak rise were also significant for both rimes, and amounted to 6% and 9% of open syllable F2 rise, respectively.

The observed effects on F2 peak timing and F2 rise were predicted by the shared representations hypothesis and thus constitute another form of support for it. The predictions follow from a specific interpretation of how shared subsyllabic structural representations influence articulatory control, which predicts a lesser degree of overlap between the vocalic and liquid gestures in structures that can be analyzed as trimoraic. The following section elaborates on this interpretation and provides further discussion of the experimental results.

4 Discussion

The main empirical result of the current study is that phonetic aspects of articulation are correlated with σ-count judgments. Specifically, rime durations and formant trajectories differed significantly between productions associated with $>1\sigma$ and $=1\sigma$ judgments. These findings suggest that when a speaker's representation of the subsyllabic structure of a word biases them toward a $>1\sigma$ judgment, the liquid gesture overlaps less with the preceding vocalic gesture than it does when their representation biases them toward a $=1\sigma$ judgment. This supports the primary hypothesis of the study, that the metaphonological process of judging σ-count and articulatory control processes share a common representation. The theoretical import of this is that at least one instantiation of a metalinguistic task—syllable counting—does indeed inform our understanding of more basic processes involved in speech. Yet a number of important questions remain: what is the connection between variation in subsyllabic structure and the observed phonetic effects? What are the implications of the observed variation in σ-count judgments for our models of intuition formation and production? More broadly, what are the implications of the results for interpretation of metalinguistic tasks? We address each of these questions in turn below.

4.1 Subsyllabic structure and gestural overlap

Why does the hypothesized variation in subsyllabic structure correlate with differences in gestural overlap between a liquid coda gesture and a preceding vocalic gesture? This finding concords with previous observations that moraic codas are associated with longer syllable durations than non-moraic codas (Broselow et al., 1997; Duanmu, 1994). However, a strictly categorical opposition between moraic and non-moraic codas seems to predict a stronger correlation than was in fact observed. Moreover, while the symbolic representation itself provides a starting point for understanding the processes that give rise to the correlation, ultimately a more mechanistic, explanatory basis for these predictions is desirable.

A plausible framework for understanding variation in overlap is the task-dynamic model of articulatory phonology (Browman & Goldstein, 1990; Saltzman & Munhall, 1989). In this model the lexical representation of a word is held to include a specification of the relative timing of articulatory gestures that comprise the word. These relative timing relations are the outcome of a system of gestural planning oscillators, which may be coupled to each other in one of two ways, in-phase or anti-phase (Goldstein et al., 2006; Nam & Saltzman, 2003). In a CV syllable, the planning oscillators associated with the onset consonantal gesture and the vocalic gesture are hypothesized to be in-phase coupled to one another, resulting in a high degree of synchrony in the relative timing of their associated gestures. In contrast, in a VC syllable, the planning oscillator associated with the coda consonantal gesture is anti-phase coupled to a preceding vocalic planning oscillator, resulting in asynchrony in relative timing. A key aspect of this model is that relative timing patterns emerge from local phasing interactions: patterns of timing result from a collection of pairwise interactions between coordinated gestures, rather than a hierarchical structure that organizes timing.

However, the articulatory phonology model of coda timing does not provide a basis for understanding the relation between σ-count judgment and gestural overlap observed in the current experiment, nor for durational differences between moraic and non-moraic codas mentioned above. The problem is that the articulatory phonology model provides just one option for control of the timing of a coda gesture relative to a preceding vocalic gesture: anti-phase coordination. The strength of the anti-phase coupling between these gestures might be varied to induce differences in gestural overlap, but this provides no explanation for why the variation is associated with σ-count intuitions, and also does not address the existence of other phonological patterns that appear to involve moraic structure. Ultimately what is needed is a model in which categorical patterns can be more clearly related to gradient variation in gestural overlap.

To address this, a recently developed theory of speech motor planning extends the task-dynamic model of articulatory phonology to account for cross-linguistic variation associated with moraic structure in rimes. This *selection-coordination* theory (Tilsen, 2013, 2014a, 2014b, 2016) holds that there are two prototypical regimes of articulatory control: competitive control and coordinative control. A key postulate of the theory is that phase-based control over timing (whether in-phase or anti-phase) is only available in the coordinative regime of control. In other words, in order for the timing of gestures to be coordinatively controlled through phasing mechanisms, the gestures must be selected together, i.e., co-selected, rather than competitively selected. **Figure 10** schematizes the difference between competitive and coordinative regimes with a simplified model of a diphthong-lateral sequence /ail/, where three articulatory gestures are shown: a pharyngeal constriction for the /a/ made with the tongue root (TR), a palatal constriction for the /i/ made with the tongue body (TB), and a velar constriction made with the tongue dorsum for the /l/ (TD).

In a purely competitive regime of control (**Figure 10A**), action plans are selected with mutual exclusion. As shown in the figure, once TR is selected, selection of TB and TD are delayed until sensory feedback results in suppression and deselection of TR; then TB is selected and subsequently deselected, allowing TD to be selected in turn. This competitive regime of control results in relatively little overlap between articulatory gestures.

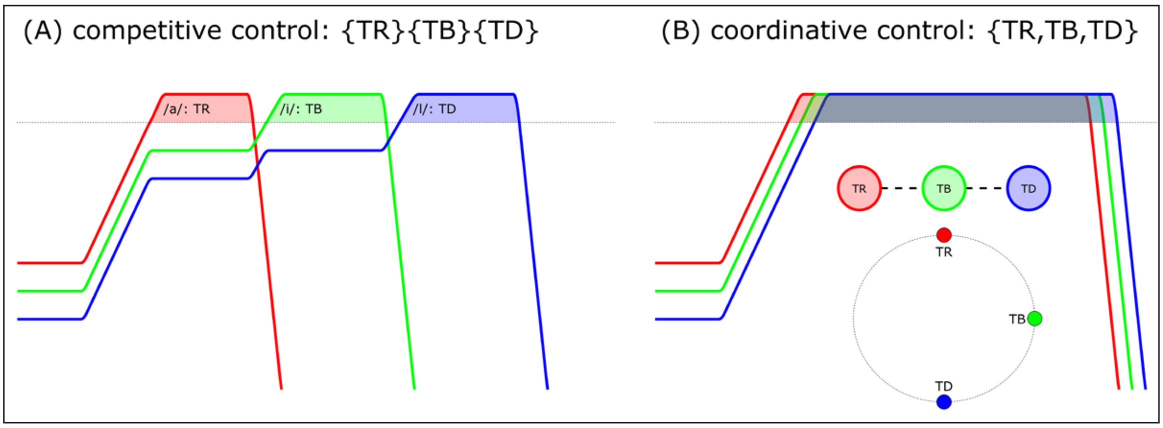
(A) competitive control: {TR}{TB}{TD}
/a/: TR
/i/: TB
/l/: TD
(B) coordinative control: {TR,TB,TD}
TR
TB
TD
TR
TB
TD

Figure 10: Illustrations of competitive and coordinative control over three gestures involved in /ail/: a pharyngeal constriction with the tongue root (TR), a palatal constriction with the tongue body (TB), and a velar constriction with the tongue dorsum (TD). **(A)** competitive control: gestural plans compete for selection. **(B)** coordinative control: phasing mechanisms determine the relative timing of movement initiation between a group of co-selected gestures. See text for details.

In a purely coordinative regime of control (**Figure 10B**), gestures are co-selected, i.e., selected contemporaneously without competition, and the relative timing of gestural execution is controlled through coordinative phasing mechanisms. This regime corresponds to the standard articulatory phonology model of a syllable, in which coda gestures and the second gestural components of a diphthong are anti-phase coordinated with a preceding gesture. The phasing relations are represented by a coupling graph in which TB is anti-phase coupled to TR, and TD is in turn anti-phase coupled to TB. The phases of the gestural plans can be envisioned to rotate counter-clockwise around a circle, and as each reaches the top of the circle its gesture is initiated. The relative timing of movements is thus derived from the relative phases of the oscillatory planning systems.

The key difference between competitive and coordinative regimes is that sensory feedback plays no role in coordinative control, but is essential for competitive control. Tilsen (2014b, 2016) hypothesizes that the transition from competitive to coordinative control arises from internalization of feedback, i.e., development of a predictive/anticipatory model of the sensory consequences of outgoing motor commands, which allows for diminished reliance on external sensory feedback (Desmurget & Grafton, 2000; Wolpert & Kawato, 1998). Various phonetic and phonological patterns observed in the course of development suggest that transitions from competitive to coordinative regimes of control are common (Tilsen, 2014b, 2016).

Importantly, the theory holds that competitive and coordinative control are prototypical modes of control, analogous to endpoints of a continuum, as illustrated in **Figure 11**. This continuum can also accommodate intermediate degrees of feedback internalization that result in intermediate degrees of gestural overlap. Panels A and B show endpoints of the continuum, where the TD gesture of /l/ is competitively selected relative to the TB gesture of /i/ (panel A) or coordinated with the TB gesture (panel B). (Note that the TR and TB gestures of the diphthong /ai/ are assumed to be coordinated.) Panel C shows the consequences of an intermediate degree of internalization. The internal feedback model (not shown) allows the TD gesture to be selected prior to deselection of the TB gesture.

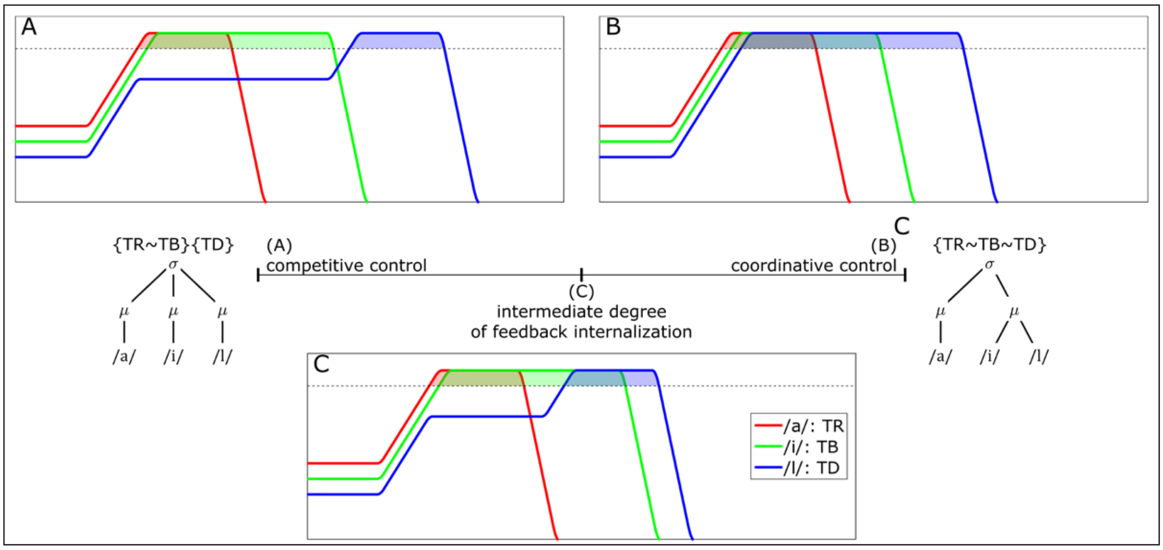

Figure 11: Illustration of a continuum between competitive and coordinative control over the tongue body (TB) gesture associated with /i/ and the tongue dorsum gesture associated with /l/ in the rime /ail/. Panels A and B show competitive and coordinative control regimes, respectively. Panel C shows a control regime with an intermediate degree of feedback internalization, which allows the TD gesture to be selected prior to suppression of the TB gesture.

This control regime is "intermediate" because the internal feedback model is not sufficiently anticipatory to allow for co-selection and coordination of the TB and TD gestures.

The distinction between moraic and non-moraic codas thus has a straightforward interpretation in selection-coordination theory: moraic codas are associated with coda gestures for which competitive control remains dominant, entailing that feedback is not sufficiently internalized to allow co-selection. Consequently, overlap between the coda gesture and preceding vocalic gesture is less extensive. Thus the model predicts that for subsyllabic structures that are "trimoraic" (i.e., the liquid gesture is competitively selected), there will be less overlap between the liquid gesture and preceding vocalic gesture. Conversely, for subsyllabic structures that are "bimoraic" (i.e., the liquid gesture is coordinated), there will be more overlap between the liquid gesture and the preceding vocalic gesture.

Figure 12 illustrates the model's predictions with hypothesized gestural scores for two versions of the rime /ail/, from productions associated with =1σ and >1σ, respectively. Each score represents the intervals of time in which articulatory gestures comprising the rime are active (Browman & Goldstein, 1990; Saltzman & Munhall, 1989). The figure also shows F2 trajectories associated with each of the gestural scores. One phonetic effect of the predicted difference in overlap is that rime durations will be longer for >1σ productions than =1σ productions. Furthermore, in diphthong-liquid rimes the sequence of articulatory gestures results in a rise and fall of F2, due to the formation and release of a palatal constriction. Thus another effect of a lesser degree of overlap is a delay in the location of the F2 peak relative to the start of the rime. This results in a third effect, which is an increase in the height of the F2 peak. The reason for the increase is that the delay of the liquid gesture allows more time for the palatal constriction gesture of [i] gesture to reach its target, and so F2 is predicted to rise higher in >1σ productions than it rises in =1σ productions. All three of these predictions were observed in association with >1σ judgments, and they can be readily understood as the consequence of variation in gestural phasing.

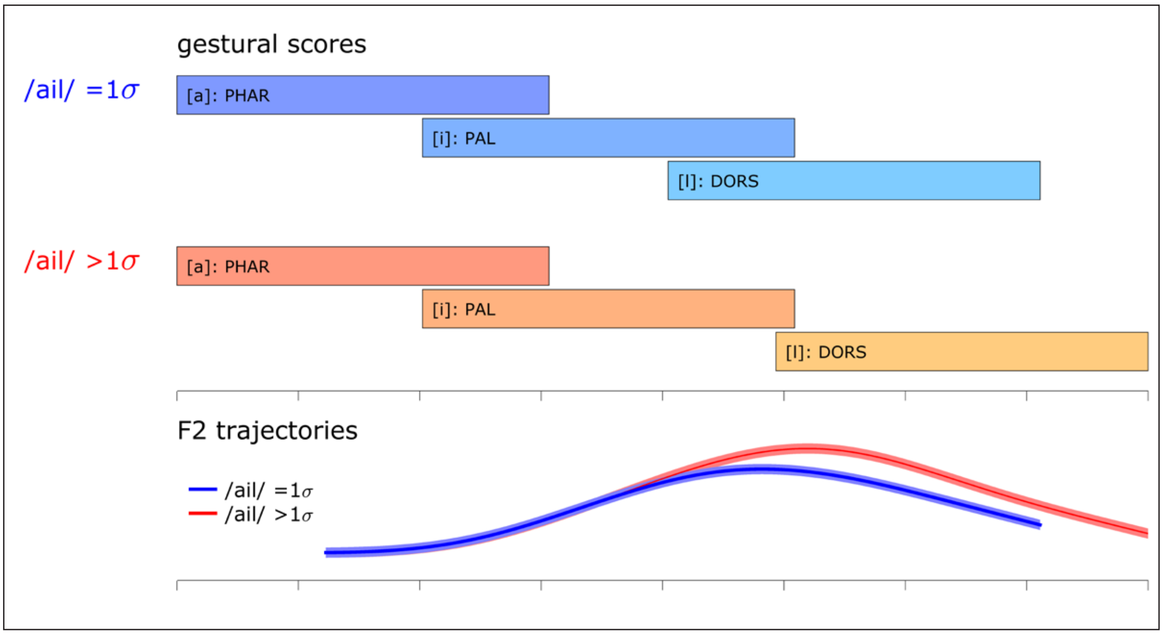

Figure 12: Gestural scores and formant trajectories predicted by the shared representations hypothesis. Shaded intervals represent the times during which pharyngeal, palatal, and dorsal constriction gestures are active. The palatal constriction results in a rise in F2; the dorsal constriction associated with /l/ results in a fall in F2.

Although the selection-coordination framework provides a detailed and general account of the mechanisms responsible for the observed variation in rime duration and formant trajectories, alternative interpretations might be considered. For example, target durations and constriction degrees/formant values may be represented in motor plans more directly. A full discussion of speech motor representations is beyond the scope of the current work, but is nonetheless an important endeavor for future research. Better knowledge of the nature of representations will undoubtedly help inform our understanding of sources of variation in judgments and production, which are considered in the following section.

4.2 Models of variation in σ-count judgments and production

The experimentally observed variation in σ-count judgments and judgment-production correlations informs our understanding of the processes involved in σ-count intuition formation and articulatory control. Previous investigations 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.

As a starting point for discussion of the mechanisms underlying the observed variation, consider the proposal of Cohn (2003) and Lavoie & Cohn (1999) that variable-count words may have a trimoraic structure which influences σ-count intuitions. In this view, the presence of a third mora in a syllable (or more neutrally, additional subsyllabic structure) biases speakers toward an intuition that the syllable is "larger" than a canonical syllable. This interpretation can be augmented by associating the trimoraic structure with competitive control over the liquid gesture and associating the bimoraic structure with coordinative control, as discussed in the previous section. This leads to a number of possibilities for models of variation in judgments and production.

For purposes of simplicity, we focus on three components of such models: a representation, an intuition formation process, and an articulatory control process, as illustrated in **Figure 13**. One possible constraint on the representation is that it is monovalent and categorical for a given speaker and word. In other words, the representation is either bimoraic or trimoraic for some lexical item for a particular speaker, and this representation is static over time for adult speakers. Alternatively, bimoraic and trimoraic representations may co-exist, and hence the representation is bivalent in the sense that any given judgment or production may derive from one representation or the other. Both processes—intuition formation and articulatory control—take a representation as input, where σ- and μ-level structure contribute to the output of the processes. The contributions of σ/μ-level structure are w_μ/w_σ and a_μ/a_σ for the intuition and articulation processes, respectively.

A model in which σ-count variation originates in a σ-intuition formation process is not consistent with the observed correlation between judgments and production. For example, consider the *monovalent representation, variably weighted intuitions* model in **Figure 14A**. In this model, 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 $=1\sigma$; speakers for whom w_σ and w_μ are more balanced may produce $>1\sigma$ 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 interword variation can be likewise accounted for. However, this model cannot account for judgment-production correlations, because the intuition formation process and its weighting

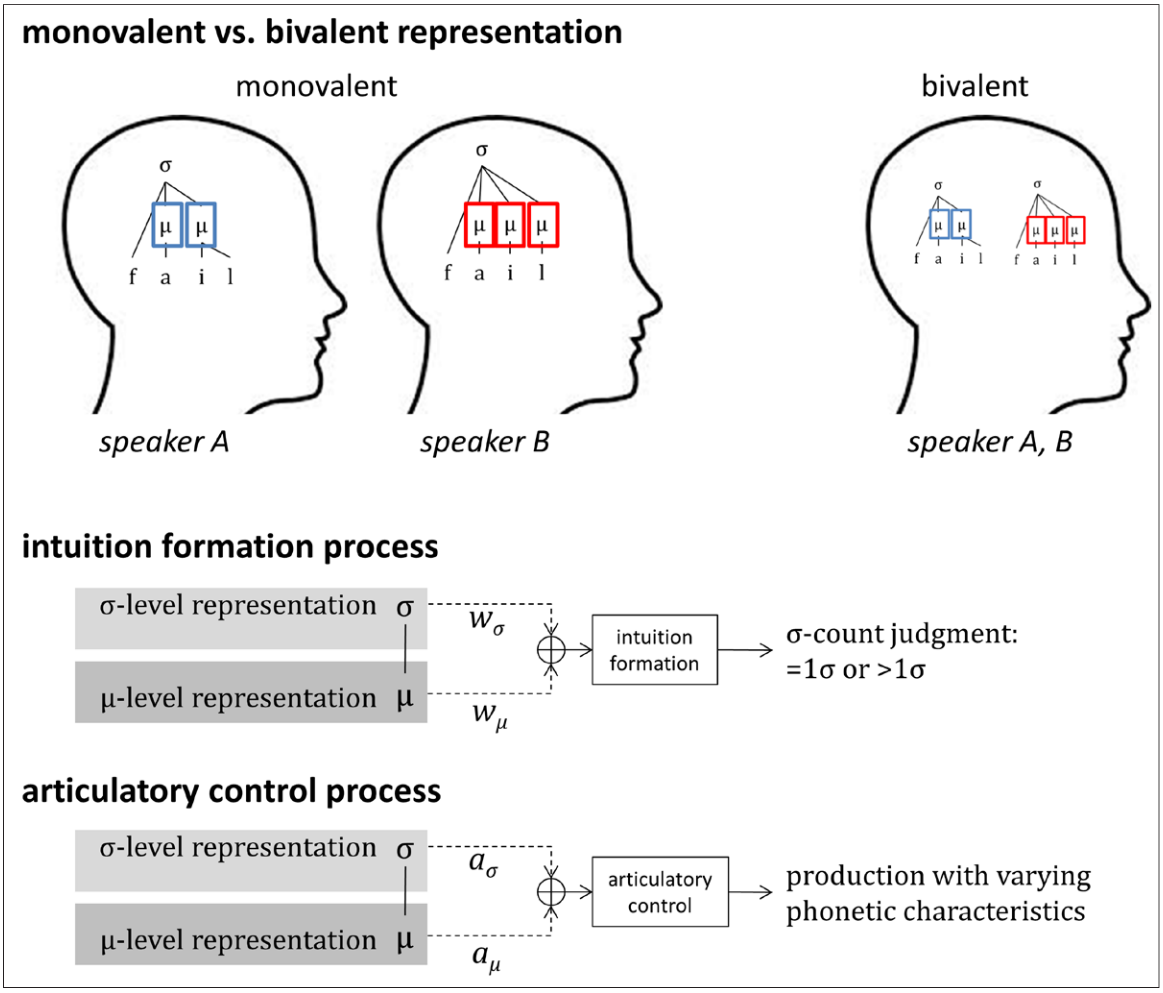

Figure 13: Along with a structural representation, an intuition formation process and articulatory control process are the three main components of schematic models of variation in σ-count judgment and production. The structural representation may be monovalent for a given speaker/word, or may be bivalent. Intuition formation and articulatory control processes involve weighting terms that modulate the contributions of different aspects of the representation to the response.

terms are independent from 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 **Figure 14B**. In this *bivalent representation* model, speakers may have either a bimoraic or trimoraic representation of variable-count words. Speakers with a bimoraic representation make =1σ judgments and those with trimoraic representations make >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.

Somewhat unanticipated was the relatively high degree of variation in σ-count judgments observed within speakers/words between tasks. 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. Changes from >1σ to =1σ and from =1σ to >1σ occurred with approximately the same frequency. Because there was no clear bias in these judgment reversals, they cannot be

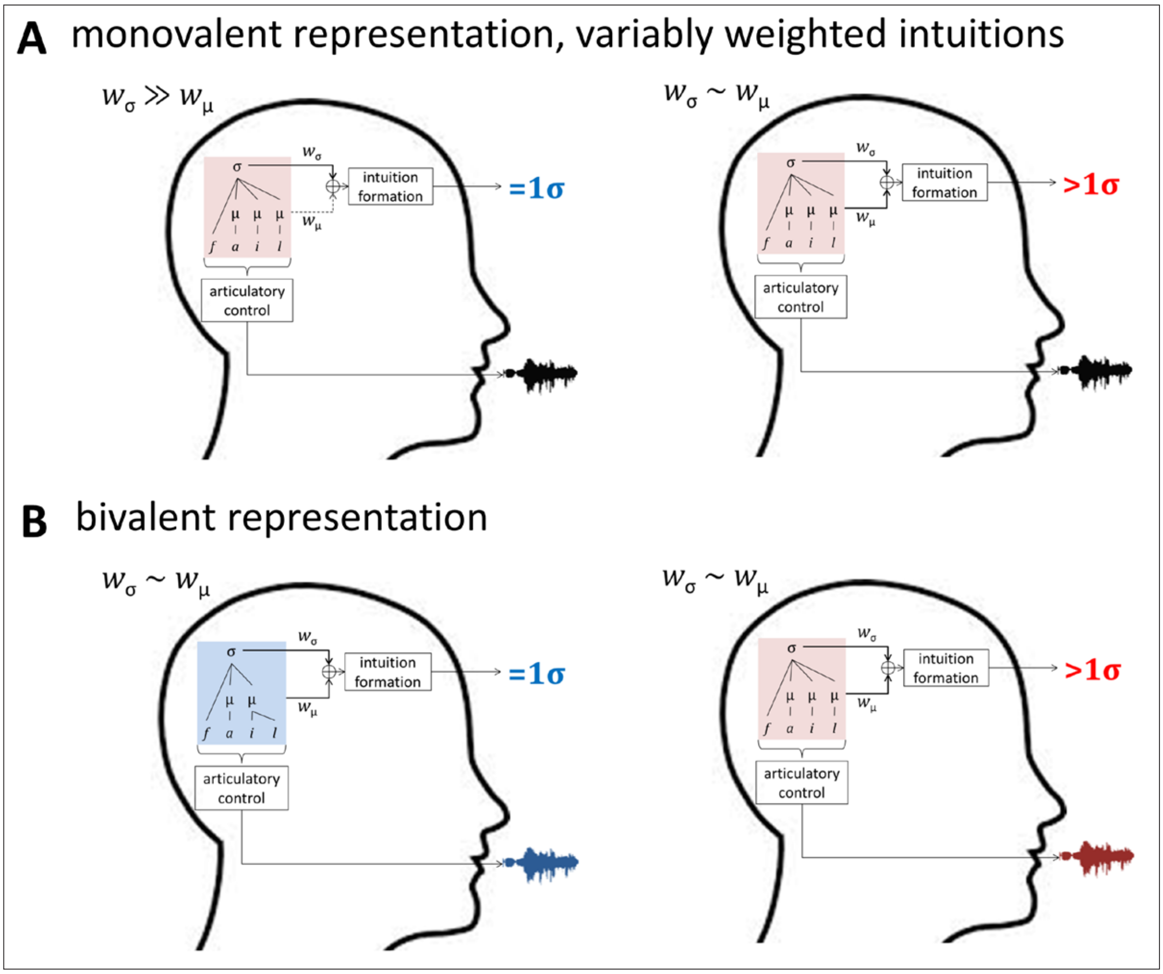

Figure 14: Schematic models of inter-speaker and inter-word variation in σ-count judgments. **(A)** *monovalent representation, variably weighted intuitions*: all variable-count words have a trimoraic representation, and variation in σ-count judgments arises from differences in relative weighting of syllabic and subsyllabic structure in the process of σ-count intuition formation; no correlation between production and σ-count judgments is predicted. **(B)** *bivalent representation*: variable-count words may have a bimoraic or trimoraic representation; correlations between judgments and production are predicted.

attributed straightforwardly to a task effect or task-order effect. For example, if contrary to fact >1σ judgments had 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. In actuality, such support for the structural attention hypothesis was not observed. One possible explanation for the absence of support is that the recency of attention to structure indeed does not have any effect on the influence that structure exerts on articulatory control. However, this conclusion may be somewhat premature given the complex interaction of factors that influence judgments and productions. Greater statistical power, which could be achieved with either a larger sample size or a reduction in variability by exerting more control over nuisance factors, may be necessary to conclusively test the structural attention hypothesis.

The observed token-level variation suggests that several factors may have interacted in a speaker-specific fashion. For instance, some speakers may have begun with a bias toward = 1σ judgments of /ir/ words but subsequently switched to >1σ judgments with recent experience of /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 that might favor relatively hypoarticulated variants. The interaction of effects of this sort has the potential to introduce additional variation in judgments, consistent with the relatively high degree of token-level variation that was observed. Regardless of its origin, the token-level variation highlights the complexity of the relation between production and metalinguistic judgments. Although further studies are necessary to test hypotheses regarding the origins of such variation, it is useful to contemplate several possibilities to provide starting points for future research.

One possible model of token-level variation would incorporate random or externally conditioned variation in the weighting terms in the monovalent representation model (**Figure 14A**). If the weighting terms of this model vary 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 likely to occur. However, as already mentioned, the monovalent representation model cannot account for judgment-production correlation. Incorporating token-specific weighting into the bivalent representations model (**Figure 14B**) 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, such as gestural overlap. The models in **Figure 15** assume that speakers potentially have both bimoraic and trimoraic representations, or that these representations correspond to endpoints of a continuous parameter dimension. One possibility is that intuition-formation and articulatory control processes are independent, each having their own weighting terms (**Figure 15A**). 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.

Another possibility is that intuition-formation and articulatory control involve a shared mechanism and shared weighting terms (**Figure 15B**). 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. Note that because our task instructions explicitly directed participants to conduct a silent rehearsal of the stimulus before producing a judgment, one might question whether the observed correlations are a product of the design. Our own impressions in producing σ-count judgments is that a subvocal rehearsal may in fact be required for this task: it seems that to produce a σ-count judgment, a speaker must to some extent engage the motor routines that they would use to produce the form. In a design without an explicit instruction of this nature, we would merely expect a greater proportion of participants to rely on orthographic properties of stimuli. Nonetheless, it is worth noting that even in the context of a subvocal rehearsal, correlation between articulatory control and intuition formation is not a logical necessity: intuition formation processes could determine syllable count judgments parallel to and independent of subvocal rehearsal.

One quite useful feature of the shared process model is that it allows for token-level variation with random or externally-driven variation in articulatory weighting parameters. These parameters 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.

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

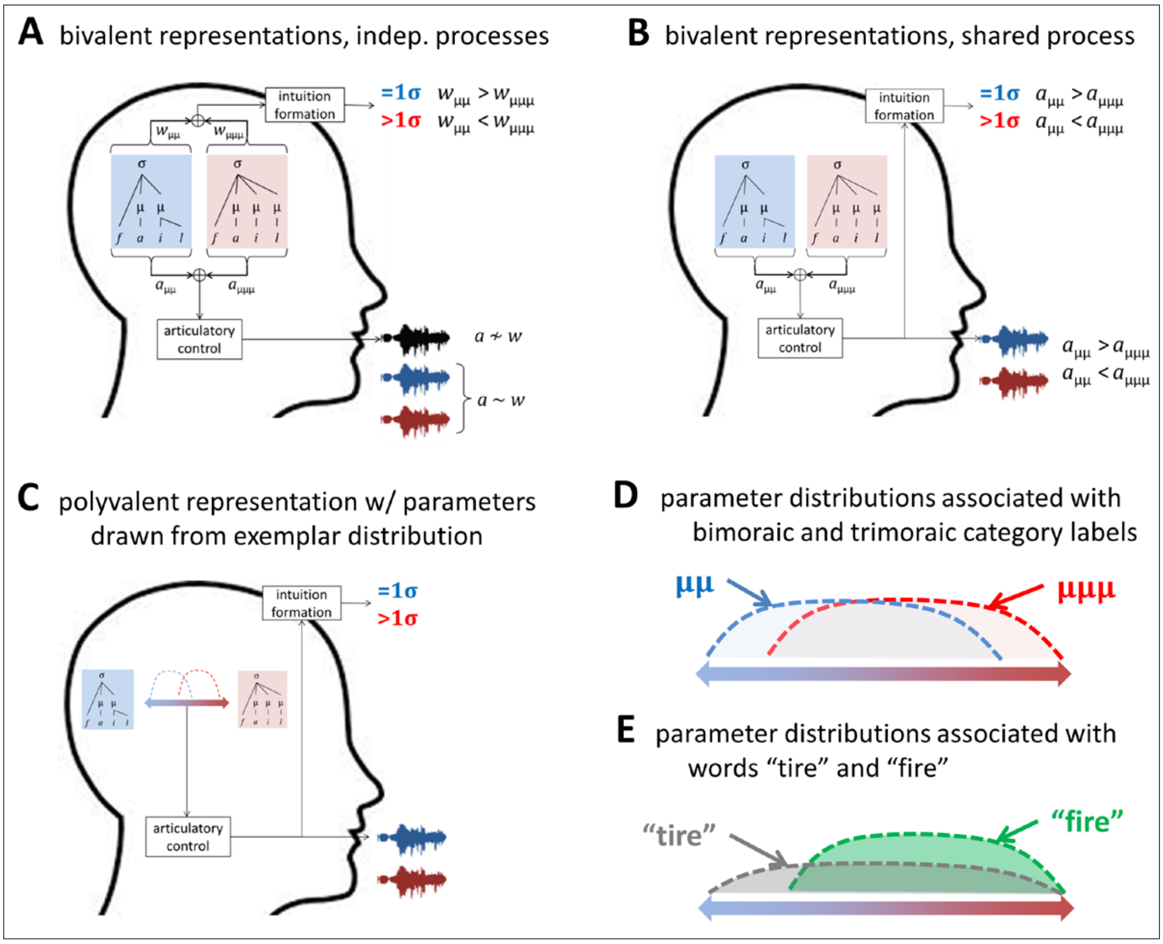
A bivalent representations, indep. processes
intuition formation
=1σ $w_{\mu\mu} > w_{\mu\mu\mu}$
>1σ $w_{\mu\mu} < w_{\mu\mu\mu}$
$w_{\mu\mu}$
$w_{\mu\mu\mu}$
σ μ μ f a i l
σ μ μ μ f a i l
$a_{\mu\mu}$
$a_{\mu\mu\mu}$
articulatory control
$a \nsim w$
$a \sim w$
B bivalent representations, shared process
intuition formation
=1σ $a_{\mu\mu} > a_{\mu\mu\mu}$
>1σ $a_{\mu\mu} < a_{\mu\mu\mu}$
σ μ μ f a i l
σ μ μ μ f a i l
$a_{\mu\mu}$
$a_{\mu\mu\mu}$
articulatory control
$a_{\mu\mu} > a_{\mu\mu\mu}$
$a_{\mu\mu} < a_{\mu\mu\mu}$
C polyvalent representation w/ parameters drawn from exemplar distribution
intuition formation
=1σ
>1σ
articulatory control
D parameter distributions associated with bimoraic and trimoraic category labels
μμ
μμμ
E parameter distributions associated with words "tire" and "fire"
"tire"
"fire"

Figure 15: 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)** continuous variation between representations with exemplar distributions. **(D)** exemplar distributions associated with bimoraic and trimoraic category labels. **(E)** exemplar distributions associated with the words *tire* and *fire*.

negative correlation between word frequency and proportion of $>1\sigma$ 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-count judgments were found to be unduly influenced by orthography. This suggests that despite the explicit instruction to base their judgments on a silent rehearsal of the word, some participants consciously or unconsciously incorporated the number of graphemes into their decision. Moreover, as discussed previously, 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) provides a useful framework for accommodating the observed variation. In this model (**Figure 15C**), bimoraic and trimoraic representations can be viewed as category labels that are associated with distributions of gradient parameter values derived from integrating individual memories. In this case the representation can be considered "polyvalent," i.e., taking on many values. The relevant parameter dimension might be the degree of overlap between the liquid gesture and preceding vocalic gesture. When speakers engage the motor system—either for intuition formation or articulation—a parameter value is selected from a speaker's distribution of previously experienced values. The distribution is built from the parameter values of past memories, which are weighted by a variety of linguistic, paralinguistic, and contextual factors, including memories of recently heard and spoken tokens.

The weighting of memories in exemplar models allows for a variety of effects to occur. For example, if the speaker is judging or producing the word *fire*, then parameters associated with memories of the word *fire* may be weighted more highly than parameters associated with the word *tire*, as shown in **Figure 15E**. Thus if the identity of a word, i.e., as *fire* or *tire*, exerts any bias on parameter values, this will be reflected in the parameter distribution and will in turn exert a bias on the selected parameter. Similarly, if the speaker is judging a less frequent or unfamiliar word such as *veal*, then the distribution of parameter values may be biased toward values associated with a more frequent/familiar word such as *steal*. Moreover, if the speaker incorporates orthographic representations in the weighting function, memories associated with orthographically similar words such as *steal*, *deal*, and *real* potentially may exert a stronger bias on the parameter distribution than memories associated with *feel*, *kneel*, and *heel*. The context-specific weighting of exemplars can account for a variety of effects that are relevant to understanding the current results, including effects of word frequency, rime similarity, orthographic similarity, and cross-stimulus priming.

The questions of which parameter(s) are selected from exemplar memories, exactly how parameter selection works, and what factors can influence the weighting of exemplar memories, are currently unresolved. Nonetheless, the exemplar-based model has two important advantages over a model in which purely abstract and categorical representations determine articulatory control parameters. First, it incorporates an associative memory network that can accommodate the wide array of factors that influence judgments and production. Second, it incorporates a mechanism for relating gradient, detailed memories and discrete, categorical representations. Specifically, trimoraic and bimoraic representations (or any alternative conceptualization of subsyllabic structural categories) can be interpreted in an exemplar model as distinct category labels that are associated with memories of parameters defined in a continuous dimension, e.g., segment duration or gestural overlap. A model of representation without an associative memory network or interactions between abstract and detailed memories is simply not powerful enough to account for the empirical observations.

4.3 Implications for metalinguistic judgments

The complexity of interactions in the current experiment highlights the fact that interpreting metaphonological judgments with regard to cognitive representations requires some caution. As discussed in Section 1.1, without independent confirmation that the meta-task judgments correlate with implicit memory-driven speech behaviors, one cannot be sure that the representations used explicitly in a meta-task indeed play a role in normal speech processes. Even if independent confirmation is obtained, detailed inferences about the nature of representations are still unclear. The results of the current study show that meta-task intuitions *can* utilize the same representations as articulatory control processes: syllable count judgments are derived at least in part from the same representations that speakers use for controlling the timing of articulatory gestures.

Furthermore, the patterns of variability observed in σ-count judgments offer some hints at the nature of this representation. The observation that word frequency influenced syllable count intuitions suggests that the representation must to some extent emerge from statistical generalizations made over previous experience, as in exemplar models. Along these same lines, the identification of several speakers who overly relied on orthography indicates that motoric phonological representations can interact with visual/orthographic ones. The observation that σ-count judgments in variable-count words were more consistent within speakers for a given vowel nucleus than a given coda suggests that the vocalic gesture(s) in a rime are more influential than the identity of the liquid. The observation that judgments were not uncommonly switched between tasks suggests that the representational distinction between judgments is mutable—a variety of factors can bias the representation in one direction or the other.

Finally, the finding that syllable count judgments and articulatory control share a common representation begs the question of whether other sorts of metaphonological tasks exhibit the same connection with typical production or perception processes. To determine whether a given meta-task, such as a wordlikeness judgment or explicit syllabification, shares some aspect of representation that plays a role in normal production or perception, independent tests are necessary to correlate the meta-task behavior with an implicit behavior. However, as we suspect that the basis for the connection between σ-count judgments and production is their shared reliance on a subvocal rehearsal, this connection might be extended to make guesses regarding other meta-tasks. For instance, if wordlikeness ratings are derived from subvocal rehearsal, then presumably they too can inform our understanding of the motor representation for normal speech. Yet if the attempted subvocal rehearsal for a wordlikeness rating requires online construction of representations not typically employed by the production system, the problems with interpreting meta-task behaviors remain.

5 Conclusion

The current experiment found that σ-count judgments correlate with phonetic aspects of production. Specifically, $>1\sigma$ judgments were associated with less gestural overlap between the coda liquid and preceding vocalic gesture compared to $=1\sigma$ judgments. This finding supports the hypothesis that σ-count judgments and articulatory control utilize the same representations. While this validates the use of meta-linguistic σ-count judgments to probe the cognitive representation of phonological structure, some caution is warranted, as evidenced by the complexity of the variability in such judgments.

In addition to a correlation between meta-linguistic judgments and production, substantial variation in σ-count judgments was observed. This variation occurred not only between speakers and words, but also at the level of individual tokens, i.e., between tasks. This suggests that speakers maintain both representations, and that a continuous parameter dimension such as gestural overlap may relate the two. The observations that word frequency, nucleus category, and orthographic composition influence judgments highlights the complexity of the system responsible for meta-linguistic judgments and begs for a model that allows for polymodal representations and relations between categorical and gradient memories, as in exemplar theory.

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 in metalinguistic judgments, representations, and speech production.

Supplementary Files

The supplementary file for this article can be found as follows:

- **Supplementary File: Appendix.** http://dx.doi.org/10.5334/labphon.52.s1

Acknowledgements

We thank Eric Ricciardi for his help in testing and running this study.

Competing Interests

The authors declare that they have no competing interests.

References

Broselow, E., Chen, S.-I., & Huffman, M. 1997. Syllable weight: convergence of phonology and phonetics. *Phonology*, *14*(1), 47–82. DOI: http://dx.doi.org/10.1017/S095267579700331X

Browman, C., & Goldstein, L. 1990. Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, *18*(3), 299–320.

Cohn, A. 2003. Phonological structure and phonetic duration: The role of the mora. *Working Papers of the Cornell Phonetics Laboratory*, *15*, 69–100.

Côté, M.-H., & Kharlamov, V. 2011. The impact of experimental tasks on syllabification judgments: A case study of Russian. In Cairns, C. E. & Raimy, E. (eds.), *Handbook of the Syllable*. Leiden: Brill, pp. 273–294.

Derwing, B. L., & Eddington, D. 2014. The experimental investigation of syllable structure. *The Mental Lexicon*, *9*(2), 170–195. DOI: http://dx.doi.org/10.1075/ml.9.2.02der

Desmurget, M., & Grafton, S. 2000. Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences*, *4*(11), 423–431. DOI: http://dx.doi.org/10.1016/S1364-6613(00)01537-0

Duanmu, S. 1994. Syllabic weight and syllabic duration: A correlation between phonology and phonetics. *Phonology*, *11*(1), 1–24. DOI: http://dx.doi.org/10.1017/S0952675700001822

Eddington, D., Treiman, R., & Elzinga, D. 2013a. Syllabification of American English: Evidence from a large-scale experiment. Part I. *Journal of Quantitative Linguistics*, *20*(1), 45–67. DOI: http://dx.doi.org/10.1080/09296174.2012.754601

Eddington, D., Treiman, R., & Elzinga, D. 2013b. Syllabification of American English: Evidence from a large-scale experiment. Part II. *Journal of Quantitative Linguistics*, *20*(2), 75–93. DOI: http://dx.doi.org/10.1080/09296174.2013.773136

Elzinga, D., & Eddington, D. 2014. An experimental approach to ambisyllabicity in English. *Topics in Linguistics*, *14*(1), 34–47. DOI: http://dx.doi.org/10.2478/topling-2014-0010

Frisch, S. A., Large, N. R., & Pisoni, D. B. 2000. Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language*, *42*(4), 481–496. DOI: http://dx.doi.org/10.1006/jmla.1999.2692

Frisch, S. A., & Zawaydeh, B. A. 2001. The psychological reality of OCP-place in Arabic. *Language*, *77*(1), 91–106. DOI: http://dx.doi.org/10.1353/lan.2001.0014

Goldstein, L., Byrd, D., & Saltzman, E. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In *Action to language via the mirror neuron system*. Cambridge: Cambridge University Press, pp. 215–249. DOI: http://dx.doi.org/10.1017/CBO9780511541599.008

Ham, W. H. 2001. *Phonetic and phonological aspects of geminate timing*. New York: Routledge.

Johnson, K. 1997. Speech perception without speaker normalization: An exemplar model. In Johnson, K. & Mullenix, J. W. (eds.), *Talker Variability in Speech Processing*. San Diego: Academic Press, pp. 145–165.

Kahn, D. 1976. *Syllable-based generalizations in English phonology* (Vol. 156). Indiana University Linguistics Club Bloomington.

Lavoie, L., & Cohn, A. 1999. Sesquisyllables of English: the structure of vowel-liquid syllables. In *Proceedings of the XIVth International Congress of Phonetic Sciences*, 109–112.

Nam, H., & Saltzman, E. 2003. A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th International Conference on Phonetic Sciences*. Barcelona, Spain, pp. 2253–2256.

Pierrehumbert, J. 2001. Lenition and contrast. *Frequency and the Emergence of Linguistic Structure*, *45*, 137. DOI: http://dx.doi.org/10.1075/tsl.45.08pie

Pierrehumbert, J. 2002. Word-specific phonetics. *Laboratory Phonology*, *7*, 101–139. DOI: http://dx.doi.org/10.1515/9783110197105.101

Saltzman, E., & Munhall, K. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, *1*(4), 333–382. DOI: http://dx.doi.org/10.1207/s15326969eco0104_2

Tilsen, S. 2013. A dynamical model of hierarchical selection and coordination in speech planning. *PloS One*, *8*(4), e62800. DOI: http://dx.doi.org/10.1371/journal.pone.0062800

Tilsen, S. 2014a. Selection and coordination of articulatory gestures in temporally constrained production. *Journal of Phonetics*, *44*, 26–46. DOI: http://dx.doi.org/10.1016/j.wocn.2013.12.004

Tilsen, S. 2014b. Selection-coordination theory. *Cornell Working Papers in Phonetics and Phonology, 2014*, 24–72.

Tilsen, S. 2016. Selection and coordination: the articulatory basis for the emergence of phonological structure. *Journal of Phonetics*, *55*, 53–77. DOI: http://dx.doi.org/10.1016/j.wocn.2015.11.005

Tilsen, S., & Johnson, K. 2008. Low-frequency Fourier analysis of speech rhythm. *Journal of the Acoustical Society of America, 124*(2), 34–39. DOI: http://dx.doi.org/10.1121/1.2947626

Treiman, R., & Danis, C. 1988. Syllabification of intervocalic consonants. *Journal of Memory and Language*, *27*(1), 87–104. DOI: http://dx.doi.org/10.1016/0749-596X(88)90050-2

Wolpert, D. M., & Kawato, M. 1998. Multiple paired forward and inverse models for motor control. *Neural Networks*, *11*(7), 1317–1329. DOI: http://dx.doi.org/10.1016/S0893-6080(98)00066-5

Yao, Y., Tilsen, S., Sprouse, R., & Johnson, K. 2010. Automated measurement of vowel formants in the Buckeye Corpus. *Gengo Kenkyu* (Journal of the Linguistic Society of Japan), *138*, 99–113.

How to cite this article: Tilsen, S and Cohn, A C 2016 Shared Representations Underlie Metaphonological Judgments and Speech Motor Control. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1): 14, pp. 1–33, DOI: http://dx.doi.org/10.5334/labphon.52

Published: 25 October 2016

]u[*Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS