Original Article

Language *and* Speech

Effects of Surprisal and Entropy on Vowel Duration in Japanese

Language *and* Speech
1–35

DOI: 10.1177/0023830917737331
journals.sagepub.com/home/las
SAGE

Jason A. Shaw
Yale University

Shigeto Kawahara
Keio University

Abstract

Research on English and other languages has shown that syllables and words that contain more information tend to be produced with longer duration. This research is evolving into a general thesis that speakers articulate linguistic units with more information more robustly. While this hypothesis seems plausible from the perspective of communicative efficiency, previous support for it has come mainly from English and some other Indo-European languages. Moreover, most previous studies focus on global effects, such as the interaction of word duration and sentential/ semantic predictability. The current study is focused at the level of phonotactics, exploring the effects of local predictability on vowel duration in Japanese, using the Corpus of Spontaneous Japanese. To examine gradient consonant-vowel phonotactics within a consonant–vowel-mora, consonant-conditioned Surprisal and Shannon Entropy were calculated, and their effects on vowel duration were examined, together with other linguistic factors that are known from previous research to affect vowel duration. Results show significant effects of both Surprisal and Entropy, as well as notable interactions with vowel length and vowel quality. The effect of Entropy is stronger on peripheral vowels than on central vowels. Surprisal has a stronger positive effect on short vowels than on long vowels. We interpret the main patterns and the interactions by conceptualizing Surprisal as an index of motor fluency and Entropy as an index of competition in vowel selection.

Keywords

Information-Theoretic, Entropy, Surprisal, vowel duration, corpus study, mora-timing, motor fluency, speech planning, Japanese

Corresponding author:
Shigeto Kawahara, The Institute of Cultural and Linguistic Studies, Keio University, 2−15−45 Mita, Minato-ku Tokyo, Japan 108-8345, Japan.
Email: kawahara.research@gmail.com

1 Introduction

Recent research has shown that speech production behavior can vary according to the distribution of information in a message. In phonetics, for example, it has been demonstrated that the duration of syllables and words can be influenced by how much information they carry (e.g., Aylett & Turk, 2004, 2006; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Bell et al., 2003; Cohen Priva, 2012, 2015; Hume, 2016; Jurafsky, Bell, Gregory, & Raymond, 2001).[1] More specifically, for instance, Aylett and Turk (2004) show that in English, more predictable vowels, that is, those that carry less information, are shorter and more centralized (see also Jurafsky et al., 2001). Bell et al. (2009) likewise show that more predictable content words are shorter in duration in conversational English. Hall, Hume, Jaeger, and Wedel (2016) provide a recent, extended review of similar findings in which information content, or predictability more generally, seems to affect phonetic implementation patterns.

Shaw, Han, and Ma (2014) demonstrate that gradient predictability can play a role at the morphophonological level as well. In forming new words through the Modern Standard Chinese pattern of compounding and truncation, what survives in truncation tends to be those segments that carry more information about the underlying word, that is, those segments that best enable listeners to recover the original, untruncated words. Information in that study is measured in terms of a combination of paradigmatic (morphological family size) and syntagmatic (frequency ratio between compound frequency and character frequency) factors influencing predictability. To provide one more example, past phonological research has reported the generalization that, given a consonant cluster straddling a syllable boundary, an onset consonant never deletes; it is only the coda consonant that can delete (Wilson, 2001). McCarthy (2008) develops a theory of constraint interaction that accounts for this observation, by postulating that only coda consonants can be deleted because they are targeted by CodaCondition (Ito, 1989). In a corpus study, however, Raymond, Dautricourt, and Hume (2006) uncover exceptions to the generalization. They examined the Buckeye corpus of spontaneous interview speech and found that onset [t] and [d] can delete in frequent words such as *somebody, lady*, and *better*, especially when the following context makes those words predictable (i.e., when they have high backward transitional probabilities); for example, [d] in *ladies and gentlemen*. Raymond et al. (2006) thus show that even phonologically-privileged sounds like onset consonants can delete, when they are highly predictable from their context. See also Cohen Priva (2015) for a similar finding in which onset [t] in English can delete when it is not informative.

A cross-linguistic study by Piantadosi, Tily, and Gibson (2011) shows that average information content may affect lexical organization—they show that information content is a better predictor of word length (the number of segments in a word) than word frequency; more informative words tend to have more segments. Seyfarth (2014) shows that words that are usually predictable tend to reduce, and that they appear as reduced even in non-predictable contexts, suggesting that the reduced forms are stored in the lexicon. Jaeger (2010) argues that predictability may affect syntactic patterns in that speakers attempt to distribute information more or less consistently across the signal, making use of the optionality afforded by the grammar. Kuperman and Bresnan (2012) show that syntactic predictability effects at choice points influence phonetic duration as well. To summarize then, information content, operationalized as some measure of predictability, seems to play a non-trivial role at every level of our linguistic behavior, from phonetics to syntax.

This growing body of research is evolving into a general hypothesis that speakers may articulate a linguistic unit with more information more robustly. This general thesis seems plausible from the viewpoint of efficient communication (Hall et al., 2016; Hume, 2016): linguistic units that carry high information should not be misperceived by the listener; on the other hand, linguistic units that can be predicted from contextual information—those with inherently low information—can be recovered by listeners, even if the signals are degraded. Indeed, as reviewed above, recent research shows that this principle may be at work in governing several aspects of our linguistic behavior.

The current study extends this previous research in two directions: (a) target languages; and (b) the targeted level of linguistic representation. First, most previous studies in this research program target Indo-European languages, including English (e.g., Arnon & Cohen Priva, 2013; Aylett & Turk, 2004, 2006; Bell et al., 2003, 2009; Cohen Priva, 2015), Dutch (Kuperman, Pluymaekers, Ernestus, & Baayen, 2007; Pluymaekers, Ernestus, & Baayen, 2005; Son & Pols, 2003), French (Bürki, Ernestus, Fougeron, Gendrot, & Frauenfelder, 2008; Torreira & Ernestus, 2009), (Brazilian) Portuguese (Everett, Miller, Nelson, Soare, & Vinson, 2011), and Spanish (Cohen Priva, 2012); the only exceptions that we know of are the study of Egyptian Arabic by Cohen Priva (2012), and a study of second-mention reduction effects in Indian English and Korean (Baker & Bradlow, 2007). The principle that the signal is controlled to maximize communicative efficiency should apply to any language, and thus needs to be tested in languages beyond Indo-European languages.

The second gap in this research is that it seldom explores interactions at segmental levels; nor does it consider possible interactions of predictability effects with other linguistic phonetic factors. Most work examines syllable or word duration (e.g., Arnon & Cohen Priva, 2013; Aylett & Turk, 2004; Bell et al., 2003, 2009), segment duration (Bürki et al., 2008; Cohen Priva, 2015; Hanique et al., 2010; Kuperman et al., 2007; Torreira & Ernestus, 2009), or a couple affixes (Pluymaekers et al., 2005) without revealing the specific phonological locus of the effects, that is, the part(s) of words, affixes, syllables, etc. that are shortened as a function of contextual predictability or how predictability relates to other aspects of articulation known to influence movement duration, for example, the distance that articulators must travel to achieve their targets (e.g., Munhall, Ostry, & Parush, 1985; Ostry & Munhall, 1985). This is important because, owing to factors independent of predictability, probabilistic reduction is unlikely to be uniform across phonetic environments. Thus, the degree to which a word is subject to probabilistic reduction may rest on the specific phonetic targets and the transitions between them, factors absent in most analyses. Relatedly, Watson, Buxó-Lugo, and Simmons (2015) raise the possibility that aspects of word duration variation could follow from competition in the selection of segments for articulation. In the context of a comprehensive review of possible accounts of probabilistic reduction, Jaeger and Buz (2017) note both the plausibility of this claim and the current lack of evidence.

With these two issues in mind, this paper assessed whether Japanese vowel duration is influenced by predictability conditioned by the preceding consonant within a consonant–vowel (CV)-mora. Our analysis targeted phonotactics, an aspect of phonological grammar, asking whether the trade-off between predictability and reduction applies at this level of grammar. Specifically, we ask whether vowels are reduced in contexts where their identity is more or less predictable from the preceding consonant, and, likewise, whether vowels are produced with longer duration when vowel identity is not predictable from the preceding consonant. We computed two measures of predictability, Surprisal, which we relate to motor fluency, and Entropy, which we relate to selection competition, and tested their effect on vowel duration. Besides these two Information-Theoretic measures (Shannon, 1948), this study also examines various other factors—vowel quality, phonological length, preceding consonantal features, syllable structure, and others—that have been previously found to affect vowel duration in Japanese. This analysis allows us to assess whether local predictability influences vowel duration beyond these potentially confounding factors, and possibly reveal how Information-Theoretic factors and linguistic factors may interact.

Japanese provides an interesting new test case, since it differs rhythmically from Indo-European languages. Japanese uses duration to express short versus long phonological contrasts including both phonemically long vowels and geminate consonants (Han, 1962, 1994; Homma, 1981). Japanese is also thought to be "mora-timed" (e.g., Han 1962, 1994; Kawahara, 2017; Port, Dalby, & O'Dell 1987; cf. Beckman, 1982; see Warner & Arai, 2001 for a critical review), such that there is some force in the language encouraging mora-based isochrony. Together these points highlight that the rhythmic organization of Japanese is different in important ways from those of other languages on which measures of predictability have been shown to influence phonetic duration. There are also specific phonetic details of Japanese that make it a particularly intriguing test case for phonologically-localized predictability. Primary amongst these is that various consonantal factors are already known to affect vowel duration in Japanese (as we will confirm below). A coarse generalization is that the same vowel tends to be produced longer after a phonetically shorter consonant than after a longer consonant. This observation has been taken as evidence that Japanese speakers keep the duration of CV units more or less constant (e.g., Homma, 1981; Port, Al-Ani, & Maeda, 1980; Sagisaka & Tohkura, 1984). Port et al. (1987) found a strong linear correlation between word duration and the number of moras that the word contains (see also Han, 1994; Kawahara, 2017 for further support; and Arai, Warner, & Greenberg, 2001 for critique). What is most interesting about the above facts in connection with the current study is that Japanese is a language in which there is substantial variation in vowel duration within the CV unit that appear to be conditioned in some way by properties of the preceding consonantal environment. On the way to testing our main research question—whether probabilistic reduction operates within Japanese phonotactics—we provide a description of Japanese vowel duration across environments that replicates and extends several past results.

The remainder of this paper is structured as follows. In section 2, we describe the details of the corpus used for analysis (2.1) and how Entropy and Surprisal were calculated, along with the theoretical motivation for these particularly measures (2.2). Section 3 reports the results. We first report patterns of Entropy (3.1) and Surprisal (3.2) calculated across the corpus and then provide a description of how duration patterned across vowel quality (3.3) and across phonetic environments (3.4). These results add to the descriptive work on Japanese vowel duration but also serve to motivate control variables in our modeling. Section 3.5 presents the modeling results. In 3.5.1, we assess the effects of Surprisal and Entropy on vowel duration through model comparison and use the best-fitting model to report predictions for control variables of interest. Section 3.5.2 focuses on the level of individual vowels to assess significant interactions found in the full model. Section 4 provides discussion of the main results including possible explanations for the differential effects of Surprisal and Entropy across vowel quality and length. Section 5 briefly concludes.

2 Method

2.1 The speech corpus

The analysis is based on the Corpus of Spontaneous Japanese (CSJ) (Maekawa, Koiso, Furui, & Isahara, 2000), one of the largest annotated speech corpora of Japanese. The CSJ contains several speech styles, including, but not limited to, Academic Presentation Style and Spontaneous Presentation Style, the former of which is based on real academic speech, which is more formal. The latter is solicited speech recorded at the recording room in the National Institute for Japanese Language and Linguistics. The speakers were given a topic about their life, for example, "tell us the happiest moment of your life" as a prompt. The speech was monologue, but there were three to four listeners at the time of recording. The speakers' ages ranged mainly from 30s to 70s. The gender was more or less balanced, although there were slightly more male speakers.[2]

The current analysis sampled from the core portion of the corpus (known as the CSJ-RDB), which comes with rich annotation and phonetic information. Our sample consisted of 11,559 unique words produced by 70 speakers, and over 346,000 vowel tokens. The CSJ-RDB includes annotated segmental intervals, created by hand, rather than automatic forced alignment.[3]

2.2 Entropy and Surprisal

Following other recent work (e.g., Cohen Priva, 2012, 2015; Daland, Oh, & Kim, 2015; Hall, 2009; Hall et al., 2016; Hume, 2016; Hume & Mailhot, 2013; Kawahara, 2016b), the current study made crucial use of Shannon's (1948) Entropy, and the related measure Surprisal. As Shannon's formulation of these quantities is general, a crucial component of their application in any specific study is how the relevant probabilities are computed and how the quantities are interpreted with respect to the theoretical constructs under study. While our main research question is whether probabilistic reduction/enhancement can be observed within a CV mora in Japanese, we have selected measures of predictability that plausibly index relevant aspects of the speech production mechanism: selection competition, in the case of Entropy; and motor fluency, in the case of Surprisal. Preceding in this manner facilitates interpretation of these predictors and interactions with other linguistic phonetic factors. We first explain how the probabilities underlying Surprisal and Entropy were computed in this study and then elaborate on the theoretical interpretations, given the computations described below.

Both Entropy and Surprisal values were calculated based on the conditional probabilities of vowels given preceding consonants in the CSJ-RDB corpus. We will write $Pr\left(v \mid C\right)$ to indicate the conditional probability that a vowel, *v*, occurs, given that the preceding consonant is *C*. For example, $Pr\,(u \mid n)$ is calculated as the frequency of /nu/ divided by the frequency of /nV/ (the onset /n/ followed by any vowel).

There are different ways to compute frequency. The token frequency of a CV mora, such as /nu/, is the absolute number of times that the sequence occurs in the corpus. The type frequency is the number of times that the sequence occurs in the list of words that occurs in the corpus. In keeping with the theoretical interpretations of these measures described above and below, we calculated Entropy based on type frequencies and Surprisal based on token frequencies.[4] As an index of motor fluency, Surprisal should be based on the number of times that a CV sequence is produced, regardless of the lexical status of the speech. Entropy on the other hand is about the number of distinct competitors, which is appropriately grounded in type frequencies across the lexicon. For the purposes of computing type-based Entropy, we counted each distinct pronunciation of a word in our sample of the CSJ corpus as a unique type.

Surprisal is the negative (base 2) log probability. Log-transforming the probability renders the units in *bits*. For the case at hand:

(1) SURPRISAL (of *v* in context *C*) $= -\log_2 Pr_t(v \mid C)$

where the subscript (t) indicates the conditional probability is calculated from token frequencies. Surprisal provides a measure of how unexpected a particular vowel is given the preceding consonant, C. CV sequences with lower Surprisal are those that are produced more frequently and are, consequently, more practiced.

Whereas Surprisal is directly related to the probability of a particular vowel, the Shannon Entropy, sometimes represented by the symbol *H*, is a property of the whole probability distribution. In this case, we are interested in the Entropy of the vowel distribution, *V*, given the preceding consonantal context. Entropy is defined as the weighted average of the Surprisal in a given context. The un-transformed probability serves as the weight. For our case:

$$(2)\quad \text{Entropy (of V in context C)} \quad = \textstyle\sum_v Pr_T(v|C) \times \left(-\log_2 Pr_T(v|C)\right)$$

where the subscript (T) indicates that the conditional probability is calculated from type frequencies. As noted above, we considered each unique pronunciation in the CSJ corpus to be a separate type. Entropy was calculated over the five Japanese vowels, /a/, /e/, /i/, /o/, /u/, in each consonantal environment in the corpus yielding a measure of consonant-conditioned vowel Entropy, henceforth Entropy. This measure quantifies vowel uncertainty in a given consonantal context. The higher the Entropy, the greater the uncertainty about vowel identity in the specified consonantal context.

2.2.1 Entropy as selection competition. We chose Entropy as a measure of probabilistic reduction for this study because it both quantifies information content and can also be related to a specific mechanism involved in speech production, that is, selection competition. All models of speech production, despite many architectural differences, share the assumption of competition and selection between sub-lexical units. Minimally, some such mechanism is required to account for the type of categorical substitution error observed in tongue twisters (Frisch & Wright, 2002) as well as naturally occurring speech errors (Fromkin, 1971; Garnham, Shillcock, Brown, Mill, & Cutler, 1981; Shattuck-Hufnagel, 1983), but competition between sub-lexical units has also been used to account for other types of data including production latencies (reaction times) under different conditions (Dell, 1986; Gafos & Kirov, 2010; Levelt, Roelofs, & Meyer, 1999; Roon & Gafos, 2016) and the phonetic details of the resulting speech (Baese-Berk & Goldrick, 2009; Gafos & Kirov, 2010; Goldrick & Blumstein, 2006; Roon & Gafos, 2016).[5]

A key assumption in linking selection competition to phonetic variation is that planning and articulation overlap in time and interact such that residual activation of a competing (non-selected) phonological form influences articulation. Phonetic variation found in a variety of contexts has been attributed to such planning–articulation interaction. Results include the phonetics of tongue twisters (Goldrick & Blumstein, 2006) and other experimental tasks designed to elicit slips of the tongue (McMillan & Corley, 2010; McMillan, Corley, & Lickley, 2009), words (with and without minimal pair lexical competitors) read in isolation (Baese-Berk & Goldrick, 2009) or in sentences (Fox, Reilly, & Blumstein, 2015), lenition over time (Gafos & Kirov, 2010), perception–production links (Roon & Gafos, 2016), and spontaneous speech (Bell et al., 2009); see also Buz and Jaeger (2016) for discussion. In their model of lenition, Gafos & Kirov (2010) use Entropy as an index of selection competition at the phonological level. For the specific case of vowel planning and production, there is experimental evidence that a planned (but not produced) vowel influences the phonetics of the target vowel (Tilsen, 2009; Whalen, 1990). Given the body of evidence that phonological competition (through incomplete competitor inhibition) can influence vowel production· we interpret Entropy, as we have calculated it for the purpose of this study, both in terms of its Information-Theoretic basis, as an index of probabilistic reduction, but also as an index of selection competition. We thereby test the effects of predictability defined at the level of Japanese phonotactics (our broader research question) but also a specific mechanistic basis through which probabilistic reduction may be achieved.

Japanese has phonotactic restrictions that reduce the number of vowels that can follow certain consonants. These categorical restrictions are reflected as decreases in Entropy. For example, since front vowels are prohibited after palatalized consonants, it is easier to predict vowel quality, either

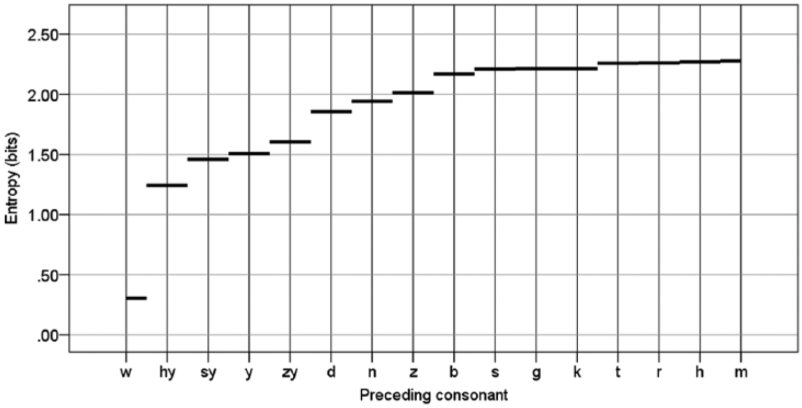

Figure 1. Contextual Entropy (Consonant-conditioned Vowel Entropy) ordered from low to high. [Xy] represents a palatalized version of X, the orthographic convention used in the CSJ. /hy/ is phonetically realized as [ç], /sy/ as [ɕ], and /zy/ as [ʑ]. See Vance (2008) for details.

/a/, /u/, or /o/, in these environments; that is, in these cases, Entropy is low. On the other hand, the distribution of the five vowels can be unpredictable given a preceding consonant, in which case Entropy is high. Besides categorical phonotactic restrictions, Entropy also reflects gradient differences in predictability. For example, all five vowels can follow both /m/ and /n/ in Japanese, but there are differences in the balance of conditional probabilities across the vowels. Within the portions of the CSJ corpus we analyzed, /i/ was more probable than /u/ following /n/—there were 74 unique words containing /nu/ and 314 containing /ni/—but /i/ and /u/ had more similar probabilities following /m/—there were 277 unique words containing /mu/ and 326 containing /mi/. These facts contribute to a difference in Entropy across /n_/ and /m_/ environments; /m/ is a higher entropy environment than /n/, although the differences in Entropy due to categorical phonotactic restrictions are much greater. As we will observe (see Figure 1), the degree of variability in Entropy across consonantal environments is sufficient to quantitatively access the effects of Entropy on vowel duration.

2.2.2 Surprisal as motor fluency. Alongside Entropy, we also assessed the role of Surprisal, another Information-Theoretic measure, on vowel duration. As described above, we calculated Surprisal across tokens so as to index motor fluency with particular CV sequences. Vowels with low Surprisal could be shorter due to motor fluency, that is, if more frequent CV colocations are produced more efficiently/consistently (e.g., Fujimura, 1986; see also Tilsen, 2014 for evidence that CV are selected together; and Tilsen, 2016 for claims that motor experience encourages co-selection of CV as a coordinated unit). We note that Surprisal could also index perceptional frequency; as such, vowels with low Surprisal could also be shorter due to audience design considerations (e.g., Arnold, 2008). Importantly, motor fluency and audience design are distinct from the competition effects that motivate the inclusion of Entropy in our analysis. The specific computation of Entropy and Surprisal for this study also ensures their statistical independence. While Entropy is calculated over the set of vowels that can occur in a given context (and based on types), Surprisal is computed separately for each vowel in each context (and based on tokens). For example, /nu/ and /ni/ have the same Entropy (recall that vowel Entropy is lower for /nV/ than for /mV/ owing to the fact that /ni/ occurs more often than /nu/) but they differ in Surprisal. The Surprisal for /nu/, which occurs 162 times in the corpus, is much higher than for /ni/, which occurs 8,060 times. Thus, while we may expect both /i/ and /u/ to be longer in /ni/ than /mi/ because /nV/ has lower Entropy than /mV/, we also predict that /i/ should be shorter than /u/ before /n/ due to the differences in Surprisal. Entropy and Surprisal can be independent in speech production, as they are in other domains of language cognition (see, e.g., Linzen & Jaeger, 2015 for an example from self-paced reading).

3 Results

3.1 Entropy by preceding consonant environment

Figure 1 shows how Entropy varies across consonantal environments. We have excluded consonants that are under-represented, showing only consonant environments with at least 1,000 occurrences in the corpus.[6] The vertical axis represents Entropy. Consonant environments, shown on the horizontal axis, are ordered from low to high entropy. The theoretical maximum of Entropy given five vowels is $2.32\,(-\log_2 p(0.2))$, which happens when all five vowels appear with the same probability (1/5 = 0.2). The solid black line indicates the Entropy of the vowel in each consonantal environment in Japanese. The consonantal environment that conditions the highest vowel entropy is /m/, which is close to the theoretical maximum. There are several other consonants, for example, /h/, /r/, /t/, /k/, /g/, /s/, with comparably high Entropy. At the left side of Figure 1, we find the consonant environments that condition low Entropy. The consonant environment with the lowest Entropy, /w/, is almost always followed by /a/, except in some loanwords, for example, [wisukii] ‘whisky’, [wizaado] ‘wizard’, [webu] ‘web’, or hyper-articulation of the particle /o/, produced as [wo]. Thus, /w/, is a near perfect predictor of following vowel quality. Since the vowel following /w/ is highly predictable, it carries little information content, and its Entropy is near zero. In between low entropy /w/ and the group of high entropy consonants there is a roughly linear increase across the various palatal consonants, /hy/, /sy/, /y/, /zy/, and then voiced coronals, /d/, /n/, /z/, and /b/.

Overall, Figure 1 indicates that there is substantial range in vowel Entropy as a function of the preceding consonant environment. This variation allows us to assess whether Entropy affects vowel duration.

3.2 Surprisal by preceding consonant environment

Figure 2 shows Surprisal (*y*-axis) in each consonantal environment (*x*-axis) for each vowel (separate panels). The consonants on the *x*-axis are ordered from low to high Entropy, the same order as in Figure 1. It is clear from this ordering that the change in Surprisal across consonantal contexts for any given vowel is distinct from the change in Entropy in that context—that is, as Entropy increases monotonically from left to right, Surprisal fluctuates up and down. Thus, a vowel in a low Entropy context, such as /w/, could have high Surprisal, such as /o/ in /wo/, or low Surprisal, such as /a/ in /wa/. The missing Surprisal values in the figure correspond to gaps in the corpus, most of which are due to well-known phonotactic restrictions or to the exclusion of devoiced vowels (see section 3.3 for a full description of exclusions).

3.3 Vowel duration for each vowel

Figure 3 shows the distribution of vowel duration for each of the five Japanese vowels. The white bars show phonemically short vowels; the gray bars show phonemically long vowels. A total of 274,602 vowel tokens were included in the analysis. Of these, 28,673 were phonemically long and 245,929 were phonemically short. We excluded vowels that followed low frequency consonants, consonants that occurred fewer than 1,000 times in the corpus (*n* = 6,902), and values that were

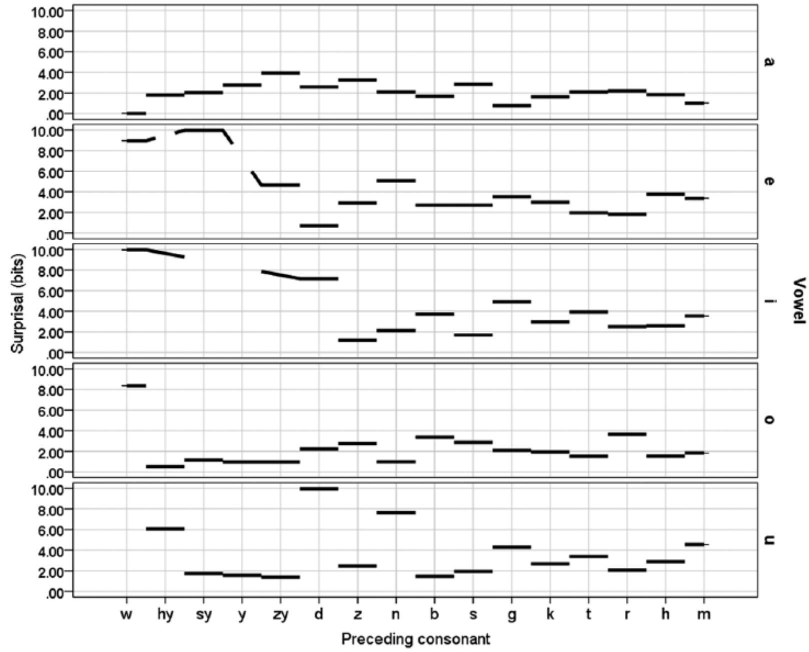

Figure 2. Surprisal by consonantal context (x-axis) and vowel (separate panels). The ordering of the consonants is from low to high Entropy (as per Figure 1).

extreme outliers ($n = 3{,}191$), defined as those with residuals from our baseline model (section 3.5) greater than three standard deviations from the mean (cf. Baayen & Milin, 2015). We also excluded devoiced vowels ($n = 22{,}450$), as their durations were not reliably measured in the CSJ measurement protocol and recent articulatory research has shown that they are sometimes deleted (Shaw & Kawahara, 2017). The shape of the distributions for each of the short vowels is similar—all have long right tails and steeper left tails that fall towards zero. The distributions of the phonemically long vowels are contained within the right tails of the phonemically short vowels. Thus, phonemic length in no way guarantees a millisecond difference in vowel duration.

Table 1 provides descriptive statistics for vowel duration by vowel quality and length. The standard deviation of vowel duration is rather high in the corpus. For phonemically short vowels, except for /a/, the standard deviation is greater than half the mean. The mean duration of the five phonemically short vowels follow the order of /a/ > /e/ > /o/ > /i/ > /u/, which is compatible with what is found in the previous studies on Japanese vowel duration (Arai, Warner, & Greenberg, 2001; Campbell, 1992, 1999; Han, 1962; Sagisaka, 1985; Sagisaka & Tohkura, 1984)—we take this replication as evidence that our data source, the CSJ-RDB, provides a valid sample of Japanese.

The pattern of duration by vowel quality for phonemically long vowels is slightly different from the pattern described above for short vowels. For phonemically long vowels, /o:/ and /a:/ are similar in duration (with /o:/ only 1 ms longer than /a:/), followed by /e/, then /i/ then /u/. The duration ratio between long and short vowels also varies across vowel qualities: /a/, /e/, /u/ are similar, around 1.6 (/a/ = 1.59; /e/ = 1.59; /u/ = 1.63) while the ratios for /o/ and /i/ are larger, around 1.90 (/o/ = 1.88; /i/ = 1.92). Thus, on average, duration is a stronger cue to phonemic length for /o/ and /i/ than for /a/, /e/, and /u/. We note that the range of duration ratios found in the CSJ is substantially

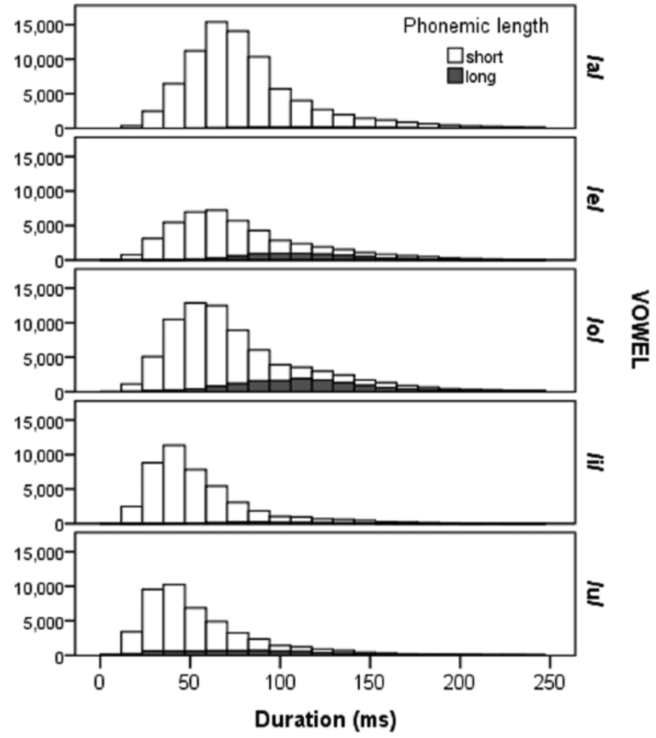

Figure 3. The distribution of vowel duration for each vowel.

Table 1. The number of valid token counts along with the mean and standard deviation (*SD*) in milliseconds of the five vowels in Japanese.

	Mean	*SD*	*n*
/a/	81	36	78,154
/a:/	129	47	1,595
/e/	74	41	39,130
/e:/	118	44	6,680
/o/	69	36	61,913
/o:/	130	78	14,315
/i/	55	31	35,742
/i:/	106	43	1,217
/u/	52	30	30,990
/u:/	85	46	4,866
Total			274,602

smaller than what are typically reported in experimental settings (e.g., Hirata, 2004). This may be due to the range of speech registers and speech rates that characterize spontaneous speech or possibly due to the age ranges of the speakers or other factors that are typically controlled in lab experiments. We also note that the counts of the phonemically long vowels are substantially fewer than the phonemically short vowels. Moreover, neither long nor short vowels are uniformly distributed across vowel quality. The most frequent phonemically short vowel in the corpus is /a/ followed by /o/; /e/, /i/, and /u/ have similar counts. The most frequent phonemically long vowel is /o:/ followed by /e:/ and /u:/, which have similar counts. There are a comparatively small number (~1,500 instances) of /a:/ and /i:/ in the corpus.

3.4 Vowel duration in different environments

Before evaluating whether our measures of vowel predictability condition vowel duration, we first report several other phonetic factors that influence following vowel duration and, therefore, need to be factored into the analysis. For example, Japanese vowels recorded in laboratory speech have been reported to be shorter following voiceless stops than following voiced stops. The explanation for the voicing effect given in past work is that since voiced stops are shorter, the following vowels are longer due to mora-timing, a tendency to maintain the temporal isochrony of mora-sized units (Port et al., 1980; Sagisaka & Tohkura 1984). An alternative is that inter- and intra-gestural timing between consonant and vowel gestures in a CV mora is constant (e.g., Smith, 1995), a consequence of which is that laryngeal abduction, characteristic of voiceless consonants, overlaps more of the vowel, leading to a shorter period of voicing, that is, shorter acoustic duration. Figure 4 shows how voicing of the preceding consonant influences acoustic vowel duration in the corpus under study here. Consistent with results from laboratory production experiments, phonemically short vowels tend to be shorter following voiceless consonants than when following voiced consonants. The effect of voicing on long vowel duration (right side of Figure 4) is absent or at least greatly attenuated relative to short vowels.

Another factor that influences vowel duration is the place of articulation of the preceding consonant. Figure 5 displays vowel duration before consonants of different places of articulation. Phonemically short vowels tend to increase in duration with the anteriority of the preceding consonant, that is, vowels preceded by more front consonants are longer in duration. This pattern, again, is in accordance with the finding previously reported in lab-read speech (Homma, 1981). For phonemically long vowels, the pattern is different: vowels are longest following labials and glottals, then velars, and finally, shortest in duration when following coronal consonants.

Another factor known to influence vowel duration is syllable structure. Japanese has closed syllables, where the coda consonants are limited to a so called “coda-nasal” or the first part of a geminate (Kawahara, 2016a; Vance, 2008). Figure 6 illustrates the durations of vowels in open and closed syllables. As shown in previous production studies, phonemically short Japanese vowels are longer in closed syllables than in open syllables (Campbell, 1999; Han, 1994; Idemaru & Guion, 2008; Kawahara, 2006; Port et al., 1987). Phonemically long vowels do not occur in closed syllables (Kubozono, 2003); hence, there is only one bar on the right side of Figure 6. We note that phonemically short vowels in closed syllables are on average intermediate in duration between short vowels in open syllables and long vowels in open syllables.

As illustrated in Figures 4–6, certain aspects of the local phonetic context appear to influence vowel duration in systematic ways. In order to evaluate the effect of predictability, as captured by measures of Surprisal and Entropy, we need to take these other factors into account. To that end, we proceed in the following section by first fitting a baseline model to the data and then evaluating whether Information-Theoretic factors increase the likelihood of the data. Before proceeding with the modeling, we close our description of vowel duration across contexts by summarizing vowel duration distributions in each of the consonantal contexts for which we have computed Entropy (Figure 1) and Surprisal (Figure 2) for short vowels in open syllables (CV), short vowels in closed syllables (CVC), and long vowels in open syllables (CV). Table 2 provides the mean, standard

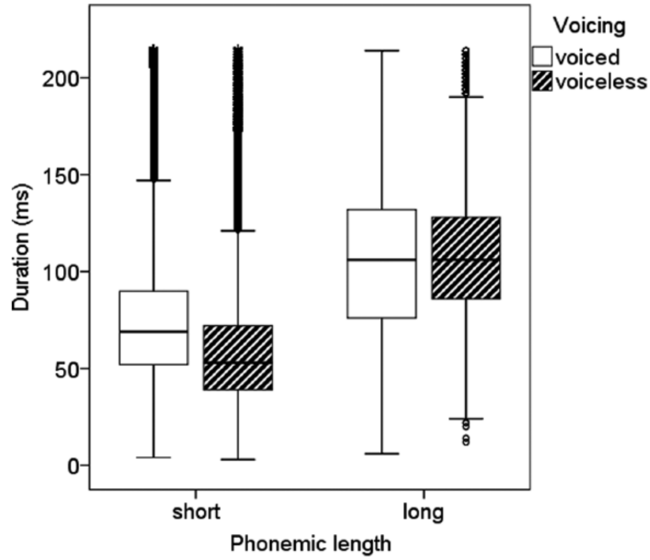

Figure 4. The average vowel duration after voiced (including both voiced obstruents and sonorants) and voiceless consonants. To simplify the visual display, outliers greater than three standard deviations from the mean have been suppressed.

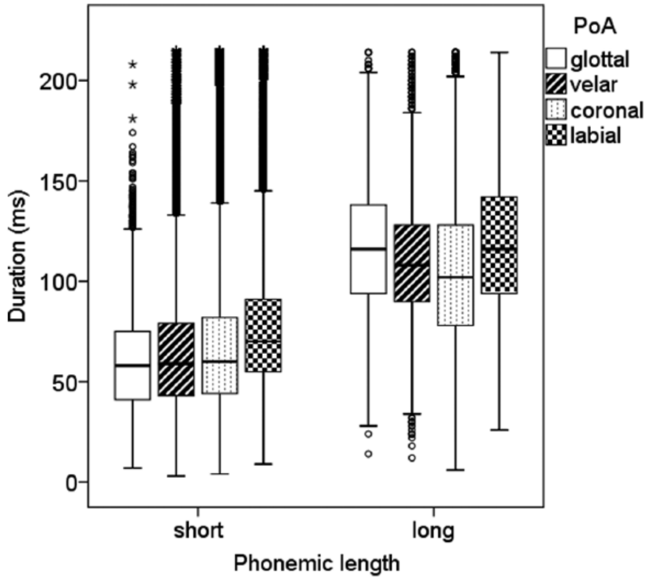

Figure 5. The average vowel duration after consonants with different primary place of articulation. To simplify the visual display, outliers greater than three standard deviations from the mean have been suppressed.

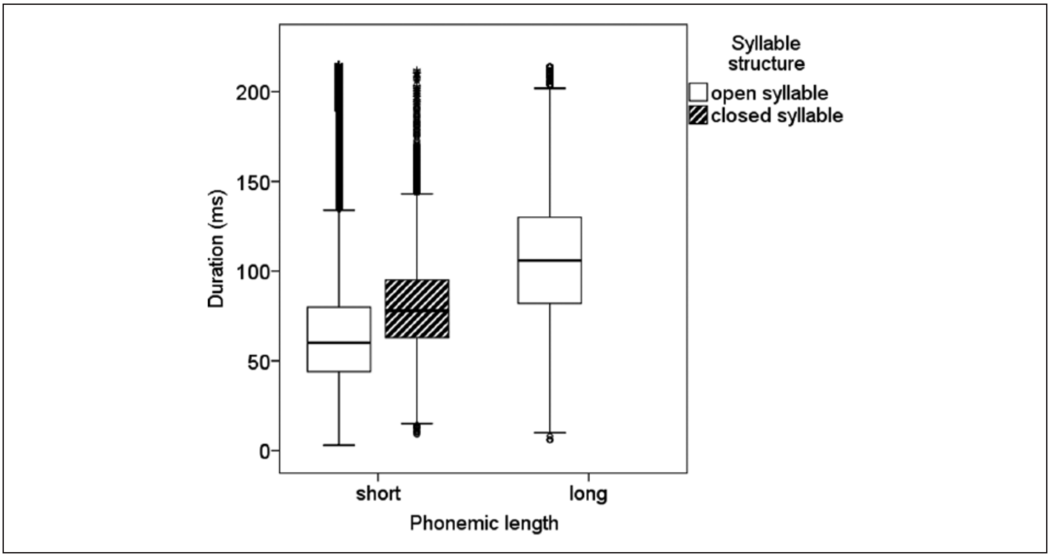

Figure 6. The effects of syllable structure on vowel duration. No long vowels appear in closed syllables. To simplify the visual display, outliers greater than three standard deviations from the mean have been suppressed.

deviation and number of observations for each cell. The consonantal contexts are ordered (top to bottom) by increasing Entropy.

3.5 Model comparison

3.5.1 The full model. We fit a series of three generalized linear mixed effects models to the vowel duration data. The first model is the baseline model, which involves phonological factors that condition vowel duration, including those presented above. The second model adds SURPRISAL as a factor and the interaction between SURPRISAL and LENGTH. The third model adds to the Surprisal model ENTROPY as a factor and interactions with VOWEL and LENGTH. A comparison between these models allows us to assess the effect of predictability in the presence of other factors that are known to influence vowel duration. We then explore the predictions of the best-fitting model to assess the statistical reliability of the other factors.

The baseline model contained the following fixed factors: LENGTH (short vs. long); VOWEL quality (a, i, u, e, o); VOICING (voiced vs. voiceless); primary PLACE of articulation (glottal, coronal, labial, and velar); SONORANCY (sonorant vs. obstruent); and SYLLABLE STRUC(TURE) (open vs. closed syllables). The fixed factors of VOWEL, VOICING, PLACE, SONORANCY, and SYLLABLE STRUC were treatment coded with the first level as the reference category: /a/ for VOWEL; voiced consonants for VOICING; glottal consonants, /h/ and /hy/, for primary PLACE of articulation; sonorants, /w/, /y/, /n/, /r/, /m/, for the SONORANCY factor; and, open syllables for SYLLABLE STRUC. All interactions between VOWEL quality and the other fixed factors were also included in the baseline model. Random intercepts for talker and for word and by-talker random slopes for ENTROPY were also included in the baseline model (the last of which is necessary for model comparison). To correct for the distributional skews apparent in vowel durations (Figure 3), we fitted the models to log-transformed vowel durations. Finally, all continuous predictor variables were *z*-scored. All three models were fit to 274,602 data points (see section 2). The structure of the models is summarized in (1)–(3):

Table 2. Summary of vowel duration distributions across environments.

		/a/			/e/			/i/			/o/			/u/		
Consonant	Form	Mean (*M*)	Standard deviation (*SD*)	Number of observations (*n*)	*M.*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
w	CV	94	52	8,564	85	22	17	62	26	9	129	35	26			0
	CVC	87	44	132	43	.	1			0	70	.	1			0
	CV:	106	29	116	130	28	2	76	.	1	179	17	3			0
hy	CV	70	21	263			0			0	145	90	2	49	8	2
	CVC	64	21	23			0			0	73	3	3			0
	CV:			0			0			0	126	36	684	81	19	13
sy	CV	75	30	843	90	61	2			0	63	25	568	52	21	441
	CVC	83	25	13			0			0	74	27	213	74	18	14
	CV:	260	.	1	108		1			0	114	44	798	98	36	588
y	CV	73	33	1,155	87		1			0	65	24	1,362	43	24	260
	CVC	65	22	167			0			0	84	27	590	45	23	42
	CV:	79	24	3			0			0	102	41	2,631	72	46	2,755
zy	CV	85	50	150	66	15	50			0	81	28	265	63	24	113
	CVC	107	28	30	64	15	45			0	72	27	17	76	22	262
	CV:	139	41	6	147	51	18			0	123	37	1,175	113	45	714
d	CV	81	28	2,276	83	52	10,081	58	14	47	66	30	2,544	52	29	16
	CVC	97	27	537	99	42	130	77	35	10	78	24	102	68	37	2
	CV:	114	41	52	115	39	253	114	42	66	125	38	1,028	180	51	4
z	CV	81	23	448	72	24	178	55	24	1,725	62	17	464	66	45	943
	CVC	93	24	44	88	25	510	65	18	567	89	21	56	80	38	18
	CV:	118	36	62	133	48	18	118	60	44	116	44	268	134	21	3
n	CV	83	31	7,051	99	49	670	76	42	7,541	84	43	16,240	65	26	189
	CVC	83	20	1,104	91	26	335	73	27	514	119	63	64	83	8	3
	CV:	146	57	13	135	65	54	155	93	5	251	153	1,286			
b	CV	84	37	1,338	66	18	648	65	31	284	87	28	410	61	27	576
	CVC	93	28	278	82	20	69	58	18	12	101	34	57	75	22	1,349
	CV:	125	31	53	106	21	104	136	44	110	144	42	49	106	.	1

(*Continued*)

Table 2. (Continued)

		/a/			/e/			/i/			/o/			/u/		
Consonant	Form	Mean (*M*)	Standard deviation (*SD*)	Number of observations (*n*)	*M.*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
s	CV	71	26	3,499	59	18	1,164	43	19	9,291	48	18	3,466	57	34	8,024
	CVC	91	27	925	77	21	1,263	62	21	367	78	20	154	70	29	56
	CV:	122	45	49	113	47	2,524	110	47	264	105	40	741	99	40	330
g	CV	98	50	8,221	69	26	390	69	27	456	73	33	2,861	48	19	669
	CVC	86	26	273	90	24	832	84	26	9	84	35	20	80	28	49
	CV:	152	48	37	124	32	33	101	24	4	114	38	466	102	26	16
k	CV	70	28	10,456	56	22	1,924	42	21	4,713	50	16	7,664	41	22	6,017
	CVC	83	23	2,479	76	22	1,746	65	24	207	70	18	553	59	18	191
	CV:	127	34	53	116	38	1,297	97	28	155	113	38	2,187	112	22	23
t	CV	74	31	9,757	70	45	10,460	45	23	2,612	63	35	15,240	41	22	4,341
	CVC	87	24	817	82	27	700	60	22	42	73	28	200	63	56	31
	CV:	112	38	434	118	37	1,000	96	37	474	160	114	1,062	111	49	118
r	CV	84	37	4,793	63	29	5,525	58	29	3,971	73	29	1,359	59	32	5,343
	CVC	93	27	268	85	31	448	68	18	133	82	24	298	62	20	236
	CV:	151	56	56	133	49	646	115	44	86	115	40	192	104	32	76
h	CV	55	19	1,734	66	27	83	42	23	1,315	59	21	618	39	17	883
	CVC	76	24	596	75	21	362	65	21	81	79	23	854	57	23	14
	CV:	170	91	7	133	60	164	115	34	7	120	45	1,423	76	33	218
m	CV	75	24	9,627	73	26	1,322	58	22	1,743	77	40	5,127	53	25	899
	CVC	83	19	293	84	23	174	67	16	93	79	26	515	62	15	7
	CV:	144	50	653	119	40	566	122	.	1	105	34	322	86	19	7

CV: Consonant–vowel; CVC: Consonant–vowel–consonant.

Table 3. Summary of model comparison.

Fixed factor	Degrees of freedom (*Df*)	Akaike Information Criterion	Bayesian Information Criterion	Log likelihood	Deviance	χ^2	*Df*	Pr(>χ^2)
Baseline	45	259,777	260,251	−129,844	259,687			
Surprisal model	47	259,700	260,195	−129,803	259,606	80.92	2	< 0.0001
Entropy model	57	259,436	260,036	−129,661	259,322	284.26	10	< 0.0001

(1) Baseline model: *duration ~ vowel*length + vowel*voicing + vowel*place + vowel *sonorancy + vowel*syllable_struc +(1 + Entropy|talker) + (1|word)*

(2) Surprisal model: *duration ~ surprisal*length + vowel*length + vowel*voicing + vowel*place + vowel *sonorancy + vowel*syllable_struc + (1 + Entropy|talker) + (1|word)*

(3) Entropy model: *duration ~ surprisal*length + vowel*Entropy*length + vowel*length + vowel*voicing + vowel*place + vowel*sonorancy + vowel*syllable_struc + (1 + Entropy|talker) + (1|word)*

Model comparisons by likelihood ratio tests[7] are summarized in Table 3. Incorporating SURPRISAL (and the interaction between SURPRISAL and LENGTH) results in significant improvement over the baseline model. The lower Akaike Information Criterion (AIC) for the Surprisal model relative to the baseline indicates that model complexity resulting from the addition of SURPRISAL and SURPRISAL*LENGTH is justified by the increased log likelihood of the data. The Entropy model offers further improvement above and beyond the Surprisal model. Adding ENTROPY and the full set of interactions between ENTROPY*VOWEL*LENGTH increases the degrees of freedom from 47 in the Surprisal model to 57 in the Entropy model. This increased complexity is justified by the increased likelihood of the data. Model improvement is reflected in a lower AIC for the Entropy model than for the Surprisal model. The best model of the data is therefore the model that includes both SURPRISAL and ENTROPY as independent predictors.

We now turn to the description of the best-fitting model, the Entropy model in (3). A complete summary of the fixed factors is provided in Table 4. The intercept represents an abstract reference category (phonemically short /a/ in an open syllable before a consonant that is voiced, glottal, sonorant, average Entropy, average Surprisal, etc.), and the fixed factors show a mix of negative and positive effects, explaining the substantial variation around the mean (as observed in Table 1 and Table 2). The direction of many of the effects is as expected given the descriptions of the raw data. For example, the coefficient for VOICING is negative, indicating that vowels are shorter following voiceless consonants than voiced consonants (cf. Figure 3). Similarly, LENGTH and SYLLABLE STRUCTURE have strong positive effects on vowel duration with a larger coefficient for LENGTH than for SYLLABLE STRUCTURE. This reflects the trend observed in the raw data for phonemically short vowels to be longer in closed syllables than in open syllables and for phonemically long vowels to be longer still (e.g., Figure 6). However, there are numerous interactions with vowel quality—every factor in the model shows a strong interaction with at least one of the vowel types in the corpus. This is telling, as it implies that we need to take care when interpreting effects on vowel duration because they may affect different vowels to different degrees or even in different directions. We proceed by plotting the predicted durations for the control variables that interact with vowel and then bore down to vowel-specific models to further assess the locus of Surprisal and Entropy effects.

Table 4. Entropy model: duration ~ surprisal*length + vowel*Entropy*length + vowel*length + vowel*voicing + vowel*place + vowel*sonorancy + vowel*syllable_struc + (1+Entropy|talker) + (1|word).

	Estimate	Standard error	*t*-value
(Intercept)	4.121	0.018	228.67
SURPRISAL	–0.032	0.003	–9.48
LENGTH	0.517	0.016	31.74
VOWEL_e	0.091	0.027	3.35
VOWEL_i	–0.187	0.022	–8.54
VOWEL_o	0.175	0.018	9.89
VOWEL_u	0.056	0.051	1.11
VOICING	–0.134	0.007	–19.45
VELAR	0.139	0.011	12.59
CORONAL	0.250	0.011	22.32
LABIAL	0.053	0.013	4.13
SONORANT	0.053	0.008	6.89
SYLLABLE_STRUC	0.228	0.007	34.69
ENTROPY	–0.004	0.003	–1.36
LENGTH: SURPRISAL	–0.009	0.006	–1.44
VOWEL_e: LENGTH	0.116	0.019	6.04
VOWEL_i: LENGTH	0.341	0.027	12.44
VOWEL_o: LENGTH	0.136	0.018	7.65
VOWEL_u: LENGTH	0.211	0.025	8.52
VOWEL_e: VOICING	0.116	0.013	8.71
VOWEL_i:VOICING	–0.146	0.015	–9.94
VOWEL_o: VOICING	–0.060	0.010	–6.05
VOWEL_u: VOICING	–0.013	0.012	–1.08
VOWEL_e:VELAR	–0.148	0.025	–5.87
VOWEL_i: VELAR	0.052	0.019	2.75
VOWEL_o: VELAR	–0.251	0.016	–15.41
VOWEL_u: VELAR	–0.513	0.047	–10.88
VOWEL_e: CORONAL	–0.260	0.025	–10.33
VOWEL_i: CORONAL	–0.131	0.018	–7.08
VOWEL_o: CORONAL	–0.292	0.016	–18.39
VOWEL_u: CORONAL	–0.434	0.049	–8.81
VOWEL_e: LABIAL	0.036	0.028	1.31
VOWEL_i: LABIAL	0.049	0.024	2.10
VOWEL_o: LABIAL	–0.184	0.020	–9.42
VOWEL_u: LABIAL	–0.248	0.049	–5.04
VOWEL_e: SONORANT	–0.114	0.015	–7.79
VOWEL_i: SONORANT	–0.163	0.015	–10.67
VOWEL_o: SONORANT	–0.109	0.011	–9.96
VOWEL_u: SONORANT	–0.067	0.014	–4.94
VOWEL_e: SYLLABLE_STRUC	0.003	0.010	0.26
VOWEL_i: SYLLABLE_STRUC	0.021	0.013	1.57
VOWEL_o: SYLLABLE_STRUC	0.055	0.012	4.78
VOWEL_u: SYLLABLE_STRUC	0.025	0.014	1.78
LENGTH: ENTROPY	0.000	0.010	0.00
VOWEL_e: ENTROPY	0.064	0.012	5.23
VOWEL_i: ENTROPY	0.108	0.019	5.67
VOWEL_o: ENTROPY	0.028	0.006	4.71
VOWEL_u: ENTROPY	–0.009	0.009	–1.06
LENGTH:VOWEL_e: ENTROPY	–0.127	0.029	–4.39
LENGTH:VOWEL_i: ENTROPY	–0.022	0.049	–0.44
LENGTH:VOWEL_o: ENTROPY	0.058	0.012	4.76
LENGTH:VOWEL_u: ENTROPY	0.072	0.019	3.88

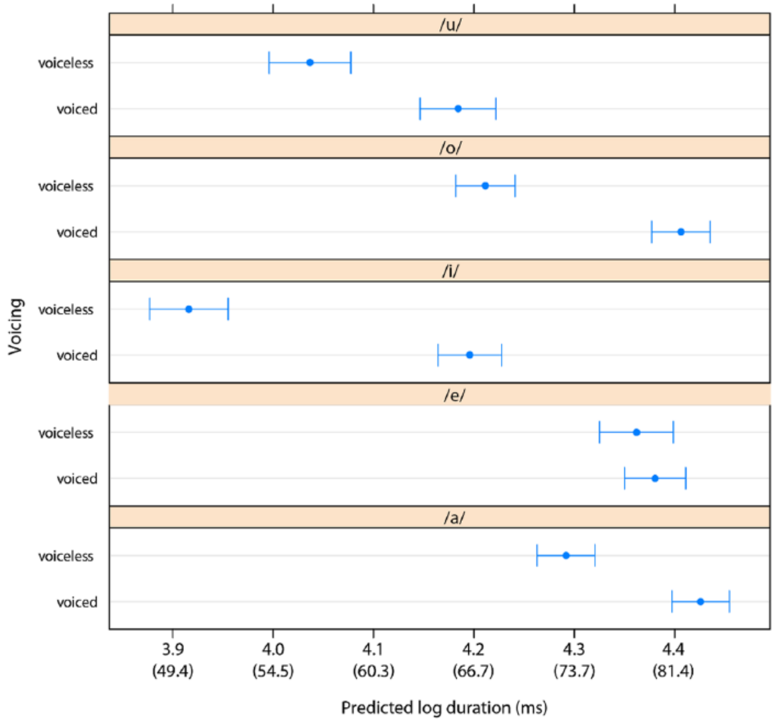

Figure 7. Model predictions for voicing by vowel. Error bars indicate 95% confidence intervals.

To visualize interactions, Figures 7–10 plot predictions of the best-fitting model by vowel for VOICING (Figure 7), SYLLABLE STRUCTURE (Figure 8), SONORANCY (Figure 9), and PLACE (Figure 10) factors. The plots were created using the lsmeans package in R (Lenth, 2016). The full Entropy model, defined in (3) and summarized in Table 4 served as the reference grid. The plots show the average predicted duration for each level of a factor across the levels of other factors. Error bars indicate 95% confidence intervals. Non-overlapping error bars can be interpreted as statistically significant at $\alpha = 0.05$. The units of the plots are in log duration, the dependent variable modeled; however, for ease of interpretation, duration values in milliseconds are provided in parentheses below the log duration labels on the x-axis of each figure.

Figure 7 shows that the effect of VOICING is consistent in direction across vowels—vowels are shorter following voiceless consonants than voiced consonants—but that the magnitude of the effect varies across vowels. The effect is largest for /i/, which may follow from the articulatory configuration for this vowel. The palatal constriction characteristic of /i/ creates aerodynamic conditions that may prolong the period of turbulent airflow following the release of a voiceless consonant (Ohala, 1983). The smallest effect of VOICING was found for /e/, the most central vowel, which was only slightly shorter following voiceless stops than when following voiced stops.

Figure 8 shows predicted durations for the levels of SYLLABLE STRUCTURE, open syllable versus closed. All five vowels are longer in closed syllables than in open syllables, although the effect is strongest for /o/. Figure 9 shows the predictions for SONORANCY. Vowels tend to be longer after sonorants than after obstruents, but /a/ is an exception and shows the reverse pattern. SONORANCY

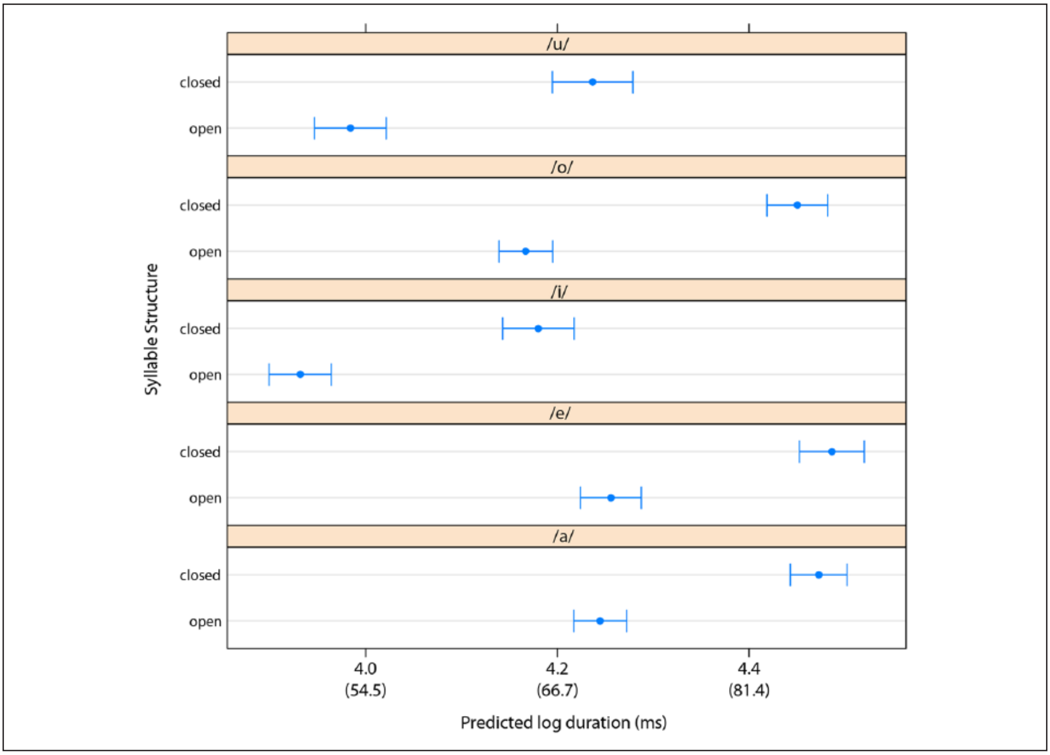

Figure 8. Model predictions for syllable structure by vowel. All vowels tend to be longer in closed syllables than open syllables, but the difference is greater for /o/ than for the other vowels. Error bars indicate 95% confidence intervals.

was significant for /a/, /i/, /u/, and /o/, but, as we also saw in the full model, the direction of the effect on /a/ (positive) is different from the other vowels. Only /a/ is shorter following sonorants than when it is following obstruents.

Figure 10 shows the predictions for PLACE of articulation, which show a complicated pattern of interaction with vowels. There is a distinct pattern of PLACE effects for each vowel. Starting with vowel duration following the glottals (the reference category for PLACE), /a/ and /i/ are shortest following glottals than other places of articulation while /o/ is longer following glottals than any other place. This vowel-specific patterning may be due to aerodynamic factors arising from lingual articulatory postures. Retraction of the tongue body for /a/ narrows the pharyngeal cavity, facilitating sustained turbulence for /h/, effectively delaying the onset of voicing. More generally, the vowels with narrower constrictions may have this effect to different degrees at different constriction locations: /a/, pharyngeal; /i/, hard palate; /u/, soft palate/uvula. Another factor contributing to PLACE effects on vowel duration could be the distance traveled from the consonant to the following vowel target. For example, /o/, which has a posterior lingual target, is longer before coronal consonants, which require the tongue to be in an anterior position, than before velar consonants. The same goes for /u/ and /a/, but /i/, a front vowel, shows the reverse pattern and /e/, the most central vowel, shows no effect of lingual place (coronal vs. velar).

Overall, Figures 7–10 illustrate complicated patterns of interaction between control variables and vowel quality. Notably, many of the patterns are explicable in light of basic

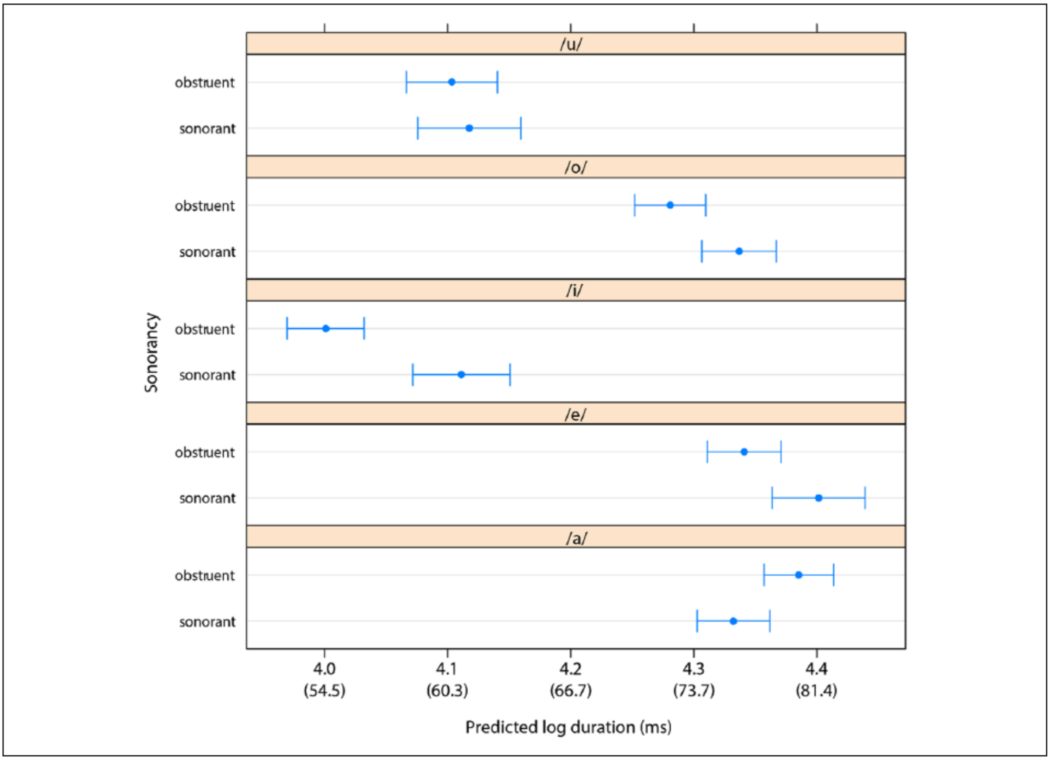

Figure 9. Model predictions for sonorancy by vowel. /o/, /i/, and /e/ are longer after sonorants than after obstruents; /a/ shows the opposite pattern. Error bars indicate 95% confidence intervals.

considerations of aerodynamics and articulatory kinematics. A similarly complex pattern of interaction surfaced for our Information-Theoretic predictability factors, which we explore in the next section.

3.5.2 Vowel-specific models. The model comparison in Table 3 establishes both that SURPRISAL improves on our baseline model of Japanese vowel duration and that ENTROPY delivers further improvement. However, the directions of these effects are rather complicated. There are multiple significant interactions between ENTROPY and VOWEL as well as significant interactions between VOWEL and other control factors (PLACE, and SONORANCY), and three-way interactions between LENGTH, ENTROPY, and VOWEL. For these reasons, we sought to probe the effects of SURPRISAL and ENTROPY further by fitting the same series of models, Baseline, Surprisal, and Entropy models in (1)–(3), to each vowel separately.

Table 5 reports AIC values as the basis for model comparison. Adding SURPRISAL to the BASELINE led to significant improvement for all vowels and a lower AIC. Adding ENTROPY led to further improvement for /a/ and /o/, marginal improvement for /i/ and no improvement for /u/ or /e/. Thus, the contribution of ENTROPY to model improvement can be localized in three of the five Japanese vowels. Specifically, ENTROPY has an impact on the vowels that have the most extreme (least centralized) articulatory positions. SURPRISAL, in contrast, affects the duration of all vowels, but the direction of the effect varied across vowels and interacted with length for /a/ and /o/. The effects found for control variables in the vowel-specific models largely paralleled those found in the full

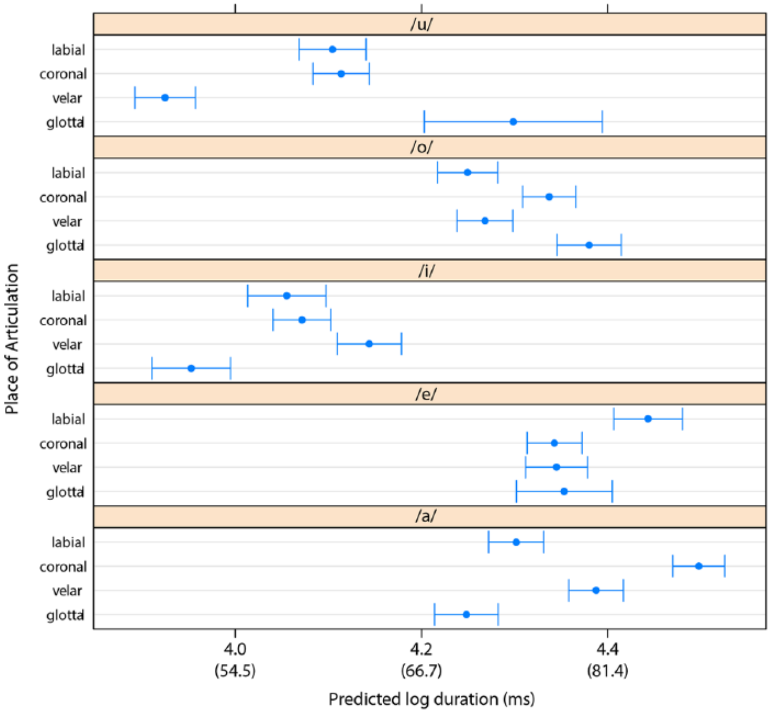

Figure 10. Model predictions for primary PLACE of articulation by vowel. Error bars indicate 95% confidence intervals.

model and displayed in Figures 7–10. A full summary of the fixed effects for vowel-specific models is provided in the Appendix.

To pinpoint the directions of vowel-specific effects and interactions with length, we re-fit the best vowel-specific models to short and long vowel subsets. For /a/, /i/, and /o/, the best model included both ENTROPY and SURPRISAL as factors; for /u/ and /e/ the best model included SURPRISAL but not ENTROPY. The β coefficients and t-values for SURPRISAL and ENTROPY across phonemically long and short vowels of different quality are summarized in Table 6. As expected from the complicated interactions in the full model, the pattern of effects is not uniform across long and short vowels of different quality. To visualize the differences, Figure 11 plots the β coefficients for ENTROPY (top row) and SURPRISAL (bottom row) across vowel qualities and length. Each column corresponds to a vowel; dotted lines represent phonemically short vowels (V) and solid lines represent phonemically long vowels (V:).

For phonemically short vowels, the effects of ENTROPY are positive—more competition results in longer vowels. Phonemically long vowels have either a positive effect of ENTROPY that may be attenuated relative to short vowels (as is the case for /o:/) or a negative effect, the case for /i:/. The direction of the SURPRISAL effect on phonemically short vowels is negative for back vowels (/o/ and /u/) and positive for the others (/a/, /u/, /e/). Phonemically long vowels also show both positive and negative effects of SURPRISAL, depending on vowel quality. From the standpoint of contrast maintenance, increases in SURPRISAL lead to enhanced short-long phonemic contrast for some vowels, /a/, /u/, /o/, and reduced contrasts for others, /i/ and /e/. This is shown visually in the Surprisal panels

Table 5. Comparisons between the baseline model and the Entropy model using likelihood ratio tests for each vowel.

	Akaike Information Criterion (AIC) Baseline model	AIC Surprisal model	AIC Entropy model	χ^2	$Pr(>\chi^2)$
/a/	57,129	57,113***	57,065	52.07	<0.0001
/i/	35,963	35,956**	35,955	5.24	0.07
/u/	39,644	39,636**	39,637	2.84	0.24
/e/	46,457	46,447***	46,449	1.37	0.50
/o/	69,480	69,303***	69,168	105.76	< 0.0001

(bottom row) of Figure 11 in the patterns of solid lines (for V:) and dotted lines (for V). When the solid line is above the dotted line, it indicates that contrast is enhanced with Surprisal. For each unit change in Surprisal, long vowels either lengthen more (/a/) or shorten less (/u/ and /o/) than short vowels of the same quality. For /i/ and /e/, Surprisal has the opposite effect—contrast is maximized in low Surprisal environments.[8]

Overall, then, SURPRISAL and ENTROPY contribute significantly to our model of vowel duration, but the effects are not at all uniform across vowels. In the discussion, we take up the fine details of these effects and aim to reconcile our results with the hypotheses laid out in the introduction.

4 Discussion

Japanese differs rhythmically from the other languages for which predictability effects on phonetic duration have been reported. On the way to evaluating whether such effects generalize to Japanese, we also confirmed several factors influencing vowel duration that have been reported in more controlled laboratory speech studies of Japanese. Features of the preceding consonant, including voicing, sonority, and place of articulation all have systematic influences on vowel duration, as do syllable structure and phonemic length. These served as necessary control variables for our assessment of predictability effects. Before moving to our discussion of the effects of predictability, we first elaborate on one discrepancy between the corpus results and the experimental literature on Japanese vowel duration.

4.1 Differences in duration ratio between corpus and experimental studies

The duration ratio between phonemically long and phonemically short vowels in the corpus was substantially smaller than what has been reported in laboratory experiments. Laboratory speech recordings report that the phonemically long vowels in Japanese are between 2.4 and 3.2 times longer than their phonemically short vowel counterparts (Han, 1962; Hirata, 2004; Tsukada, 1999). In contrast, our corpus study found that long vowels were only 1.6 or 1.9 times longer, depending on vowel quality. There could be several reasons for this discrepancy related to the nature of spontaneous speech versus controlled elicitation in experiments. A typical experimental design investigating the duration of long and short vowels makes use of minimal pairs as stimulus items, so as to control for the numerous other influences on vowel duration. However, the presence of a minimal pair in the lexicon may result in more robust expressions of contrast (Baese-Berk & Goldrick, 2009; Wedel, 2012; Wedel, Kaplan, & Jackson, 2013). Most words with a short vowel in Japanese

Table 6. Fixed factor estimates for ENTROPY and SURPRISAL in vowel-specific models. Effects with *t*-values greater than 2.0 are shown in bold.

	ENTROPY				SURPRISAL			
	Short		Long		Short		Long	
	β	*t*-value	β	*t*-value	β	*t*-value	β	*t*-value
/a/	**0.03**	**7.27**	**0.04**	**2.66**	0.02	1.75	0.04	0.57
/i/	0.01	0.30	–0.14	–1.23	**0.03**	**2.35**	–0.05	–1.73
/u/					**–0.04**	**–5.00**	–0.01	–0.38
/e/					**0.03**	**2.78**	0.01	1.10
/o/	**0.07**	**7.98**	**0.03**	**2.18**	**–0.11**	**–12.1**	**–0.07**	**–3.22**

do not in fact have minimal pairs differentiated only by vowel length. In spontaneous speech, short vowels are much more frequent than long vowels—there is an order of magnitude difference in the corpus investigated here. Words likely to be selected as materials in experiments may over-represent the duration difference between long and short vowels more generally. Moreover, experiments including words with both long and short vowels at equal or roughly equal frequency may also promote enhancement of durational contrast within the experimental session. On the other hand, the CSJ corpus contains spontaneous speech representing a greater range of speech registers, rates, and talkers than are typically represented in experimental studies. Each of these factors could contribute to diminished duration ratio across the corpus. Although it was not practical to factor speech rate and register into the statistical analysis in this study, we consider this fruitful ground for future research. The discrepancy in duration ratio results notwithstanding, the patterns found in the corpus largely replicated past reports of our control variables.

4.2 Vowel-specific effects of Entropy and Surprisal

We now turn to the Information-Theoretic factors. Overall, we found that both SURPRISAL and ENTROPY have significant influences on Japanese vowel duration. This result indicates that phonotactic predictability has a gradient influence on phonetic duration in Japanese. We motivated these particular measures of predictability as indexes of cognitive processes involved in speech production—Entropy indexes competition in the selection of a vowel; Surprisal indexes motor fluency. We hypothesized that, all else being equal, more practiced CV transitions (low Surprisal) would be shorter, owing to the motor fluency that develops with practice or possible co-selection of CV units (Tilsen, 2016), while vowels in less practiced environments (high Surprisal) would be longer, and that vowels with more competition (high Entropy environments) would be longer, owing to the time required to fully inhibit competitor vowels, while vowels with less competition (low Entropy environments) would be shorter. Our results were somewhat more intricate than these simple predictions. They include vowels that follow this exact trend as well as vowels that do not. We proceed by discussing the results for each vowel in turn, pointing out where the predictions are upheld and possible reasons why they breakdown where they do.

From the standpoint of our hypotheses, /a/ showed the most straightforward effects of SURPRISAL and ENTROPY. Both factors had positive effects on vowel duration for both long and short vowels. The direction of the effects is such that the duration of /a/ increases in contexts where /a/ is unexpected (high SURPRISAL) and also in environments in which vowel identity is uncertain (high ENTROPY). Consistent with the hypotheses above, the positive effect of ENTROPY may indicate that it

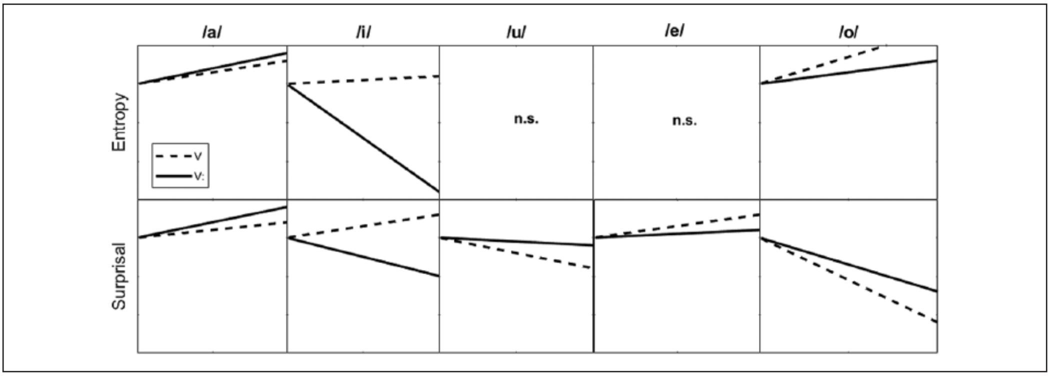

Figure 11. Lines plot the slopes of β coefficients for Entropy (top row) and Surprisal (bottom row) factors in the best-fitting vowel-specific model. Dotted lines show short vowels; and solid lines show long vowels. The direction and steepness of the slopes varies across vowel quality and vowel length.

takes more time to resolve competition when there are a greater number of contextually likely alternatives. The positive effect of surprisal is consistent with the interpretation that under-practiced (high surprisal) CV sequences require more time to actuate.

The other two vowels influenced significantly by entropy were /i/ and /o/. We first discuss /i/. Both surprisal and entropy had negative effects on long /i:/ and positive effects on short /i/. The positive effect of entropy on short /i/ is weaker than the effect for /a/, but it is in the same direction, which supports our general hypothesis. It is not clear why entropy has a negative effect on long /i:/. From the standpoint of our competition-based interpretation of entropy, this implies that /i:/ is acoustically shorter when there are more active vowel competitors (within the CV mora) and lengthened when there are fewer active competitors. One way that /i/ differs from other vowels that may be relevant to this result is in its coarticulatory aggression (Chen, Chang, & Iskarous, 2015; Iskarous et al., 2013; Recasens & Espinosa, 2009). Compared with other vowels, /i/ has a strong coarticulatory influence on neighboring segments (Chen et al., 2015). In Japanese specifically, /i/ induces allophony on preceding consonants, for example, /t/➔[tʃ]/__i (Shaw, 2007; Shaw & Balusu, 2010). The CSJ transcribes distinct allophones for numerous consonants before /i/. Although we have ignored such allophonic variation in our computation of Entropy (focusing on phonological contrast), it is possible that articulatory enhancement in the presence of competitors for /i/ takes the form of increased overlap with the preceding consonant. This could be akin to how phonological neighborhood density has been shown to condition coarticulation—words with more lexical neighbors showing more coarticulation (Scarborough, 2004). If high Entropy induces increased coarticulation of /i/ with the preceding consonant, the acoustic duration of /i/ could decrease (even as the total articulatory duration of /i/ remains constant or even increases). Increased temporal overlap with preceding consonants (in the presence of competition) may not be available to the same degree for vowels that have less coarticulatory aggression. However, this hypothesis is difficult to test without articulatory data, and it is not clear—if increased temporal overlap is responsible for shorter acoustic duration—why the overlap effect should be so much larger for long /i:/ than for short /i/.

Turning now to Surprisal, we found that short /i/ has a positive effect, similar to /a/, while the effect of Surprisal on long /i:/ goes in the opposite direction. In more practiced consonantal contexts (low surprisal), /i:/ actually tends to be longer than in less practiced (high surprisal) contexts. A similar result was found by Tomaschek et al. (2013) for German /i/ (although they used word frequency as their index of motor practice). In their study, the duration difference between phonemically long and short words was found to be greater in high frequency words than in low frequency words. The negative effect of SURPRISAL on long /iː/ suggests that the development of motor fluency may be sensitive to phonological structure. Shortening may not be an appropriate index of motor fluency for long vowels, since it pushes towards contrast neutralization. The more practiced /Ciː/ sequences tend towards durations that enhance contrast with phonemically short vowels. In this way, the type of motor efficiency indexed by SURPRISAL appears to be sensitive to phonological contrast. That the contrast enhancing effect of practice shows up more strongly in /iː/ than /aː/ may be related to the different duration ratios that cue length contrasts for these vowels in Japanese—recall that /i/ and /o/ have a higher duration ratio between long and short vowels (as reported in section 3.3: 1.9 for /i/ and /o/, cf. 1.6 for /a/, /e/, /u/).

We now turn to /o/. The effect of ENTROPY was positive, as predicted, for both phonemically short and long vowels. The interaction between ENTROPY and LENGTH was significant (see Appendix), indicating that the effect of Entropy on long /oː/ is attenuated relative to the effect that ENTROPY has on short /o/. This may be because phonemically long vowels have inherently more time to resolve competition, although in that case it is not clear why the effect of ENTROPY is not similarly attenuated for /a/ (which is produced with similar average duration: Table 1).

There was also a significant effect of SURPRISAL on both long /oː/ and short /o/, but these effects were both negative, which is the opposite direction of our predictions. From a strictly Information-Theoretic perspective, this is unexpected, but it makes sense when we consider also the spatial target of /o/ alongside the spatial distribution (within the vocal tract) of consonants that condition variation in SURPRISAL for /o/. /o/ requires the most extreme posterior position of all of the Japanese vowels, but consonantal environments that condition low SURPRISAL tend to be palatal environments, which require the tongue to be in an anterior position. Thus, it happens that, for /o/, low SURPRISAL environments tend to require larger movements to transition from the preceding consonant to the /o/ target. While it is generally true that the peak velocity of articulators increases with increases in tongue displacement (Ostry & Munhall, 1985; Roon, Gafos, Hoole, & Zeroual, 2007), there are upper bounds on tongue velocity, and the magnitude of peak velocity increase cannot always compensate for greater displacement. That is, larger articulator movements may require more time than smaller movements. Relatedly, as we reported in Figure 5, vowel durations are generally longer following consonants with anterior constrictions. From an acoustic standpoint, /o/ requires sufficient time to achieve a low steady state F2 following the high F2 for palatal consonants. Thus, what seems like an exception—a negative effect of SURPRISAL for short /o/—has a principled explanation when we consider the starting positions for vowel movements associated with contexts of varying SURPRISAL.

We conclude our discussion of vowel-specific patterns with /e/ and /u/. The effects of SURPRISAL on these vowels are similar to those already discussed. /e/ patterns with /a/ in that there is a positive effect of SURPRISAL on both short and long vowels and it patterns with /i/ in that the effect of Surprisal (motor fluency) enhances contrast between short and long vowels. /u/ patterns with /o/ in having negative effects of SURPRISAL on short vowels. The strength of the negative effects on /u/ are not as strong as those on /o/, which follows from our explanation in terms of articulatory movement distance. For /o/, we suggested that the negative effect of SURPRISAL could follow from the coincidence that many low SURPRISAL contexts have anterior articulatory targets and therefore require larger movements to achieve the posterior target for /o/. The same goes for /u/ except that SURPRISAL for /u/ is not quite as low before palatal consonants as it is for /o/ and the distance from the palatal constriction to the target for /u/ is not as large as the distance for /o/ (more on that below). Therefore, the additional time required to achieve a /u/ target from (low SURPRISAL) palatal contexts is also not as great for /u/ as for /o/.

Neither /e/ nor /u/ showed significant effects of ENTROPY, which may also follow from the spatial targets of these vowels. /e/ and /u/ are the most central of the Japanese vowels. /e/ is a mid, front vowel; the International Phonetic Alphabet symbol we have been using for /u/ suggests a high, back place of articulation, but articulatory studies have shown that it is centralized in Japanese and only slightly higher than /e/ (Shaw & Kawahara, 2017); the magnetic resonance imaging images reported in Isomura (2009) point to the same observation, /u/ is not as posterior as /o/. Cross-linguistically as well, /u/ has a tendency to centralize, that is, the phenomenon of “GOOSE fronting” in various dialects of English (e.g., Blackwood-Ximenes, Shaw, & Carignan, 2017; Harrington, Kleber, & Reubold, 2008; Sóskuthy, Foulkes, Haddican, Hay, & Hughes, 2015), and is known to be particularly susceptible to coarticulation from neighboring segments (Recasens & Espinosa, 2009).

The absence of an effect of ENTROPY on more central vowels may provide additional insight into the nature of the ENTROPY effects on the peripheral vowels. We assume with other work in speech production (e.g., Dell, 1986; Roon & Gafos, 2016; Tilsen, 2016) that selection of a gesture for articulation requires inhibition of competitors. The positive effects of ENTROPY on vowel duration for more peripheral vowels may reflect delayed movement towards targets resulting from residual competition. Incomplete inhibition of a competitor vowel in planning is known to have influences on vowel articulation that are detectable from small shifts in formant values (Tilsen, 2009; Whalen, 1990). A shift in formants towards the (not fully) inhibited vowel indicates that the articulators are pulled towards the competitor vowel. Recovery from such spatial perturbation may take time, ultimately elongating the vowel, as per positive effects of ENTROPY on peripheral vowels: /i/, /a/, /o/. As central vowels require, on average, less articulator displacement to obtain their targets, they can recover from competitor-based spatial perturbation without elongating the vowel. In this way, the absence of ENTROPY effects on the duration of more central vowels sheds light on the spatial nature of the effects. When progressing from a consonant articulation to a vowel articulation, proximal vowel targets show greater temporal resilience to slight detours in movement direction than distal vowel targets. Intuitively, a slight detour need not make you late if you have plenty of time to get where you are going to begin with. Consideration of vowels as articulatory movements that unfold in time from a starting point to a target strengthens our interpretation of ENTROPY as an index of competition in vowel selection.

Overall, the discussion of the vowel-specific effects highlights two general themes. One is that motor fluency, as indexed by Surprisal, is sensitive to phonological length contrasts. The second is that we need to take into account the spatial position of the tongue, including its position at the onset and offset (target) of controlled movement, when considering movement duration. The conditions under which the basic hypotheses raised in the introduction break down follow from these two additional considerations. We continue our discussion by elaborating on some implications of these results.

4.3 Phonologically-guided motor optimization

For some vowels of Japanese, the effect of SURPRISAL was more positive on short vowels than on long vowels. This results in enhanced contrast between phonetically long and short vowels in lower Surprisal contexts. For /i/ and /e/, practice with a CV sequence (lower Surprisal) leads to either a shortening effect on phonemically short vowels and a lengthening effect on phonemically long vowels (the case of /i/) or a shortening effect on short vowels and an attenuated shortening effect on long vowels (the case of /e/). These patterns result in greater contrast in low Surprisal contexts. For these vowels, the degree of contrast between long and short vowels increases with practice, suggesting that motor fluency develops with sensitivity to phonological contrast.

Usage-based enhancement of phonological contrast is not necessarily expected. Enhancement of this type is largely *inconsistent* with a popular axiom in phonetic research that speech patterns involve a compromise between speaker-oriented factors, such as ease of articulation, and listener-oriented factors, such as perceptual distinctiveness (e.g., Lindblom, 1990). Listener-oriented factors demand more robust expressions of phonological contrast in less frequent (high Surprisal) contexts, which is not what we observe for length contrasts in Japanese. We had raised the possibility in section 2.2.2 that Surprisal could be positively correlated with vowel duration due to audience design considerations, that is, speakers may phonetically enhance unlikely vowels. This is consistent only for the patterns found for /a/, which showed longer durations and greater length contrast between phonemically long and short /a/ in high Surprisal contexts. The results for the other vowels do not support this view—the length contrast is actually produced less robustly in high Surprisal contexts for /i/ and /e/. For /u/ and /o/ the effect of Surprisal is negative on both short and long vowels. The Surprisal patterns in our data are also not consistent with frequency-based reduction of articulatory effort. Some phonemically long vowels, /i:/, /u:/, /o:/, are actually produced with greater duration in low Surprisal (more practiced) contexts than in high Surprisal contexts. To the extent that phonetic duration is related to articulatory effort, as is sometimes claimed (Zipf, 1949), these results challenge the idea that there is effort-based reduction in more frequent CV sequences.

The effects of Surprisal on long versus short /i/ require an alternative view whereby motor fluency enhances phonemic contrast. One possibility is that speakers receive internal feedback that functions to enhance phonological contrast, possibly by more robust encoding of phonetically distinct tokens. Fine tuning of phonological targets on a token-by-token basis may explain why more frequent CV sequences condition more robust phonetic contrasts between long and short vowels. Another possibility is that practice facilitates the physical actuation of abstract task goals organized to maximize phonological contrast. Tomaschek, Tucker, Wieling, and Baayen (2014) suggest that practice with an articulatory sequence may make particular articulations more precise. If so, the major difference between high and low Surprisal contexts may not be the phonological targets for the vowels but the degree to which they can be reliably produced in spontaneous speech. The differential effects of Surprisal on length preservation across vowels deserves attention in future research.

4.4 The effect of spatial displacement on movement duration

From the standpoint of the Information-Theoretic principles from which Entropy and Surprisal are derived, vowels are random variables that can take, in the case of Japanese, one of five values. Our analysis links the contextual predictability of that random variable to vowel duration, a physical parameter related both to the actuation of phonological form through articulatory movement and to the resulting acoustics. Although we found strong effects of Entropy and Surprisal on vowel duration, the strength and direction of the effects varied significantly across vowels. We have argued that the source of the interaction can be understood when we go beyond vowels as random variables to consider the phonetic details involved in their production.

For two vowels, /o/ and /u/, we observed negative effects of Surprisal. That is, higher Surprisal, which equates to less practice with a particular CV sequence, led to shorter durations. At first blush, this appeared to be an unexpected exception to the Information-Theoretic principles guiding our hypotheses. All else equal, more Surprising vowels (particularly phonemically short vowels) were hypothesized to be longer than less surprising counterparts. When it comes to actual speech articulation, however, all else is *not* equal; there are, for example, spatial differences in the articulation of vowels that influence movement duration. Of the five Japanese vowels, /o/ requires the greatest posterior displacement of the tongue; /u/ is more centralized but also requires some posterior displacement (Isomura, 2009). The CV transitions involving the greatest anterior-posterior displacement of the tongue are thus those that involve a consonant with an anterior lingual constriction followed by /o/ or /u/. Accordingly, we found that /o/ was longer on average when preceded by a coronal consonant than when preceded by a dorsal consonant, and the same was true for /u/ (Figure 10). Due to phonotactic constraints prohibiting front vowels in palatal contexts, these back vowels, /o/ and /u/, have relatively low Surprisal in palatalized contexts, which involve an anterior lingual constriction. The negative effect of Surprisal for these vowels, then, follows from a particular confluence of factors—consonant environments conditioning low Surprisal happen also to require greater tongue displacement to achieve these vowel targets.

We also appealed to the spatial position of vowels in our interpretation of the full range of Entropy effects across vowels. There were positive effects of Entropy on the three most peripheral short vowels. This result is in the expected direction, as it indicates that more uncertainty about vowel quality results in longer vowel duration. Assuming a planning–production cycle involving cascading activation, in which gesture-selection and actuation are temporally overlapping processes, we interpret on-going competition in terms of local perturbation. That is, movement towards a vowel target may be temporarily diverted in the direction of a competitor vowel that is not fully inhibited. Such misdirection effects may be akin to those observed in masked-priming experiments tracking articulation (Davis et al., 2015). Davis et al. (2015) show that the articulators move in the direction of a subliminal masked prime displayed before the target word is produced. When tongue displacement in the direction of the (non-target) prime is antagonistic to the stimulus, the tongue speeds up to achieve its production goals. Increases in peak velocity may fail to compensate fully for spatial perturbation, resulting in increased vowel duration, as the tongue arrives late to its phonetic target. Viewing the effects of Entropy on articulation in spatial terms allows us to make sense of the lack of effect of Entropy on non-peripheral vowels. Vowel targets that are more central in the articulatory space can recover from competitor-based spatial perturbation without elongating the vowel.

4.5 Gradient phonotactic predictability conditions vowel duration

Our extension of the predictability-based hypothesis to phonotactics was inspired in part by the proposal that all levels of linguistic organization may be guided by the distribution of information (Jaeger, 2010). Most of the studies reviewed in the introduction have related phonetic duration to the predictability of higher level units, for example, words, phrases, etc. The domain over which we computed Surprisal and Entropy was restricted to a phonologically relevant unit, the CV mora, in Japanese, as opposed to a unit of meaning, such as the morpheme or word. It remains unclear the degree to which the effects that we have observed at the level of phonotactics are related to variation in duration at the higher levels of linguistic organization in which CV units are embedded. We view explorations into such connections as fertile ground for future research. For the time being, we conclude on the basis of our results that the distribution of information at the level of phonotactics also influences vowel duration.

Theories relating Information-Theoretic measures of predictability to signal robustness (e.g., Aylett & Turk, 2004; Jaeger, 2010; Jurafsky et al., 2001; Son & Pols, 2003) generally operate at the computational level and, as such, are generally consistent with a range of possible mechanisms through which humans may achieve efficient speech (Jaeger & Buz, 2017). Experimental research has identified factors that influence both signal robustness in the phonetics, as indexed by phonetic duration and vowel dispersion, and lexical access in speech production, as indexed by naming latency. Examples include the predictability of metrical stress (Shaw, 2012, 2013), and phonological neighborhood density (Buz & Jaeger, 2016; Munson, 2007). Dual effects of predictability on planning and articulation may suggest an intimate link (Fink & Goldrick, 2015), although care needs to be taken when generalizing these results (Buz & Jaeger, 2016). We motivated the particular Information-Theoretic measures used in this study also in terms of plausible mechanisms in the speech production process. The mechanistic interpretations were particularly relevant for understanding variation in how predictability influences different vowels in the corpus.

This brings us to a second methodological point. We investigated each vowel in our analysis separately, which is also somewhat unusual in the antecedent literature. Attempts to pinpoint probabilistic reduction effects in particular segments have been rare (although see Hall et al., 2016). Recall that in the current work, Entropy showed strong interactions with vowel (Tables 3). Following up those interactions ultimately strengthened our mechanistic interpretation of the Information-Theoretic effects. The broader pattern of results across particular vowels is consistent with our interpretations of Surprisal as inversely related to motor proficiency and of Entropy as an index of selection competition. Non-uniformity in the effects of Surprisal and Entropy across vowels followed from principled phonetic reasons, particular to individual vowels. These interactions underscore the importance of considering the details of the phonetics, including aerodynamics, kinematics, and the mapping between articulation and acoustics, when drawing inferences about speech planning from phonetic parameters.

Ultimately, the totality of our results is consistent with the broader hypothesis that units with more information are produced more robustly. Phonetic robustness in this case follows from phonologically-informed motor optimization and selection competition between vowels, insights that were gained by considering in detail the differential effects of Surprisal and Entropy on the duration of Japanese vowels. Overall, this study highlights the importance of focusing on individual data, which may reveal the interplay of various principles including Information-Theoretic factors that govern phonetic behavior. Maintaining support for the broader hypothesis at this level of detail may require sophisticated incorporation of phonetic factors and commitment to the mechanisms underlying predictability–robustness tradeoffs in speech.

5 Conclusion

To conclude, the current analysis of the CSJ reveals that various factors affect vowel duration in Japanese. Amongst them were two Information-Theoretic factors, Surprisal and Entropy, defined at the level of the CV mora. In doing so, the study revealed: (a) that tradeoffs between predictability and phonetic robustness can be found internal to the phonology; and (b) that effects of predictability on vowel duration generalize to a language with rhythmic properties not found in the Indo-European languages on which much previous research has focused.

To explain the role of Surprisal and Entropy at the level of phonotactics, we appealed to mechanistic interpretations—Surprisal as an index of motor fluency and Entropy as an index of competition in vowel selection. Doing so brought order to the non-uniformity of these effects across vowels of different quality and length. Notably, differential effects of Surprisal on the duration of phonemically long and short vowels of the same quality conspired to enhance phonological length contrast with practice, which we interpreted as evidence for phonologically-guided motor optimization.

Acknowledgements

We are grateful to Robert Daland and an anonymous reviewer for extensive comments on previous versions of the paper.

Funding

The collaborative nature of this research was made possible by Japan Society for the Promotion of Science (JSPS) (grant number 15F15715), and was further supported by JSPS grants to the second author (number 26284059 and number 17K13448).

Notes

1. How to quantify "information" is a non-trivial question, and in fact the previous studies reviewed in this section use a range of different measures. Our own analysis makes use of Information-Theoretic (Shannon, 1948) Surprisal and Entropy, formal definitions of which are reported in section 2.2.
2. For further details, which should not be relevant for the current analyses (such as the whole list of topics in Academic Presentation Style and recording equipment details), see the documentation available at http://pj.ninjal.ac.jp/corpus_center/csj/manu-f/recording.pdf
3. Thanks to Hanae Koiso (personal communication) for answering some of our questions regarding the CSJ-RDB. A reviewer asked how the CSJ-RDB determined a boundary between [w] and [a]. According to the CSJ manual available at http://pj.ninjal.ac.jp/corpus_center/csj/k-report-f/06.pdf (p. 327), they: (a) determined the end of the steady state of the preceding vowel; (b) determined the midpoint of the glide (located based on the formant peak); and (c) determined the onset of the steady state following vowel. The onset of the glide was marked at the middle point between (a) and (b). The end of the glide was marked at the middle point between (b) and (c). Devoiced vowels were often not distinguishable from the preceding consonants, and hence often merged with them, ending up having no interval of their own.
4. We have also calculated token-based Entropy and type-based Surprisal and used them in the analysis. Models using these measures were not any better than those that we report in this paper in that they do not increase the likelihood of the data.
5. A salient alternative to accounting for some facts about production latencies (Mooshammer et al., 2012) and the phonetic details of errors (Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007) involves considering competition between gestural coupling regimes as opposed to competition for gesture selection. There is on-going debate as to which facts about errors can be captured via partial activation in phonological selection (Goldrick & Chu, 2014) versus competitive coupling (Pouplier & Goldstein, 2010, 2014).
6. Consonants that occurred less than 1,000 times (token counts) were: /dy/, /kw/, /ty/, /ny/, /v/, /ry/, /ky/, /cy/, /py/, /by/, /my/, and /p/.
7. To enable model comparison via likelihood ratio tests, model parameters were fit using maximum likelihood method (instead of the lmer default restricted maximum likelihood method).
8. An anonymous reviewer points out that, since our dependent variable is log duration, an additive effect of Surprisal (as opposed to a multiplicative effect) would be expected to have a stronger influence on short vowels than on long vowels, a pattern that we observe for /e/.

References

Arai, T., Warner, N., & Greenberg, S. (2001). OGI tagengo denwa onsei koopasu-ni okeru nihongo shizen hatsuwa onsei no bunseki [Analysis of spontaneous Japanese in OGI multi-language telephone speech corpus]. *Nihon Onkyoo Gakkai Shunki Happyoukai [The Spring Meeting of the Acoustical Society of Japan]*, *1*, 361–362. [In Japanese.]

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, *23*(4), 495–527.

Arnon, I., & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, *56*(3), 349–371.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56.

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, *119*(5), 3048–3059.

Baayen, R. H., & Milin, P. (2015). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28.

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, *24*(4), 527–554.

Baker, R. E., & Bradlow, A. R. (2007). Second mention reduction in Indian, English, and Korean. *Journal of the Acoustical Society of America*, *122*(5), 2993.

Beckman, M. (1982). Segment duration and the 'mora'in Japanese. *Phonetica*, *39*(2–3), 113–135.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America*, *113*(2), 1001–1024.

Blackwood-Ximenes, A., Shaw, J., & Carignan, C. (2017). A comparison of acoustic and articulatory methods for analyzing vowel variation across American and Australian dialects of English. *Journal of Acoustical Society of America*, *142*(2), 363–377.

Bürki, A., Ernestus, M., Fougeron, C., Gendrot, C., & Frauenfelder, U. H. (2008). Factors influencing French schwa deletion and duration: A corpus-based study. *Journal of the Acoustical Society of America*, *123*(5), 3889–3889.

Buz, E., & Jaeger, T. F. (2016). The (in)dependence of articulation and lexical planning during isolated word production. *Language, Cognition and Neuroscience*, *31*(3), 404–424.

Campbell, N. (1992). Segmental elasticity and timing in Japanese speech. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic structure* (pp. 403–418). Amsterdam, The Netherlands: IOS Press.

Campbell, N. (1999). A study of Japanese speech timing from the syllable perspective. *Journal of the Phonetic Society of Japan*, *3*(2), 29–39.

Chen, W.-r., Chang, Y.-c., & Iskarous, K. (2015). Vowel coarticulation: Landmark statistics measure vowel aggression. *Journal of the Acoustical Society of America*, *138*(2), 1221–1232.

Cohen Priva, U. (2012). Sign and signal: Deriving linguistic generalizations from information utility. Doctoral dissertation. Stanford University, California, USA. Retrieved from https://stacks.stanford.edu/file/druid:wg646gh4444/UrielCohenPriva-Dissertation-augmented.pdf

Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. *Laboratory Phonology*, *6*(2), 243–278.

Daland, R., Oh, M., & Kim, S. (2015). When in doubt, read the instructions: Orthographic effects in loanword adaptation. *Lingua*, *159*(1), 70–92.

Davis, C., Shaw, J., Proctor, M., Derrick, D., Sherwood, S., & Kim, J. (2015). Examining speech production using masked priming. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: University of Glasgow. Retrieved from https://campuspress.yale.edu/jasonshaw/files/2016/12/davis_shaw_etal_2015_ICPHS0560_priming-1k7o7vk.pdf

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.

Everett, C., Miller, Z., Nelson, K., Soare, V., & Vinson, J. (2011). Reduction of Brazilian Portuguese vowels in semantically predictable contexts. Paper presented at the Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong, 17–21 August 2011. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Everett/Everett.pdf

Fink, A., & Goldrick, M. (2015). The influence of word retrieval and planning on phonetic variation: Implications for exemplar models. *Linguistics Vanguard*, *1*(1), 215–225.

Fox, N. P., Reilly, M., & Blumstein, S. E. (2015). Phonological neighborhood competition affects spoken word production irrespective of sentential context. *Journal of Memory and Language*, *83*(1), 97–117.

Frisch, S. A., & Wright, R. (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics*, *30*(2), 139–162.

Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language*, *47*(1), 27–52.

Fujimura, O. (1986). Relative invariance of articulatory movements: an iceberg model. In J. S. Perkell, & D. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 226–242). Hillsdale, NJ; London, UK: Lawrence Erlbaum Associates.

Gafos, A., & Kirov, C. (2010). A dynamical model of change in phonological representations: The case of lenition. In F. Pellegrino, E. Marsico, I. Chitoran, & C. Coupé (Eds.), *Approaches to phonological complexity* (pp. 225–246). Berlin, Germany: Mouton de Gruyter.

Garnham, A., Shillcock, R. C., Brown, G. D., Mill, A. I., & Cutler, A. (1981). Slips of the tongue in the London-Lund corpus of spontaneous conversation. *Linguistics*, *19*(7–8), 805–818.

Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*(6), 649–683.

Goldrick, M., & Chu, K. (2014). Gradient co-activation and speech error articulation: Comment on Pouplier and Goldstein (2010). *Language, Cognition and Neuroscience*, *29*(4), 452–458.

Goldstein, J., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, *103*(3), 386–412.

Hall, K. C. (2009). *A probabilistic model of phonological relationships from contrast to allophony*. PhD Thesis. The Ohio State University, USA.

Hall, K. C., Hume, E., Jaeger, F., & Wedel, A. (2016). The message shapes phonology. *SST satellite workshop*, December 10–11, 2016. Retrieved from http://sst2016.westernsydney.edu.au/wp-content/uploads/2016/05/Hall_Hume_Wedel_Jaeger9-12-16_SST_Symposium2016-for-distribution.pdf

Han, M. (1962). The feature of duration in Japanese. *音声の研究 (onsei no kenkyuu 'Phonetics Research')*, *10*(1), 65–80.

Han, M. S. (1994). Acoustic manifestations of mora timing in Japanese. *Journal of the Acoustical Society of America*, *96*(1), 73–82.

Hanique, I., Schuppler, B., & Ernestus, M. (2010). Morphological and predictability effects on schwa reduction: The case of Dutch word-initial syllables. Paper presented at the 11th Annual Conference of the International Speech Communication Association (Interspeech 2010). Retrieved from http://pubman.mpdl.mpg.de/pubman/item/escidoc:529023:7/component/escidoc:529175/IS10Hanique_Schuppler_Ernestus.pdf

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation,/u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America*, *123*(5), 2825–2835.

Hirata, Y. (2004). Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics*, *32*(4), 565–589.

Homma, Y. (1981). Durational relationship between Japanese stops and vowels. *Journal of Phonetics*, *9*, 273–281.

Hume, E. (2016). Phonological markedness and its relation to the uncertainty of words. *On-in Kenkyu [Phonological Studies]*, *19*, 107–116.

Hume, E., & Mailhot, F. (2013). The role of entropy and surprisal in phonologization and language change. In A. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 29–50). Oxford, UK: Oxford University Press.

Idemaru, K., & Guion, S. G. (2008). Acoustic covariants of length contrast in Japanese stops. *Journal of the International Phonetic Association*, *38*(02), 167–186.

Iskarous, K., Mooshammer, C., Hoole, P., Recasens, D., Shadle, C. H., Saltzman, E., & Whalen, D. (2013). The coarticulation/invariance scale: Mutual information as a measure of coarticulation resistance, motor synergy, and articulatory invariance. *Journal of the Acoustical Society of America*, *134*(2), 1271–1282.

Isomura, K. (2009). *Nihongo-wo Oshieru [Teaching Japanese]*. Tokyo, Japan: Hitsuji. [In Japanese.]

Ito, J. (1989). A prosodic theory of epenthesis. *Natural Language and Linguistic Theory*, *7*(2), 217–259.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62.

Jaeger, T. F., & Buz, E. (2017). Signal reduction and linguistic encoding. In E. M. Fernández, & H. Smith Cairns (Eds.), *The handbook of psycholinguistics* (pp. 38–81). Chichester, UK: Wiley-Blackwell.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). The Netherlands: John Benjamins.

Kawahara, S. (2006). A faithfulness ranking projected from a perceptibility scale: The case of [+ voice] in Japanese. *Language*, *82*(3), 536–574.

Kawahara, S. (2016a). Japanese has syllables: a reply to Labrune. *Phonology*, *33*(1), 169–194.

Kawahara, S. (2016b). Japanese loanword devoicing once again: Insights from Information Theory. *Proceedings of FAJL*, *8*, 43–62.

Kubozono, H. (2003). The syllable as a unit of prosodic organization in Japanese. In C. Fery, & R. van der Vijver (Eds.), *The syllable in optimality theory* (99–122). Cambridge, UK: Cambridge University Press.

Kuperman, V., & Bresnan, J. (2012). The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory and Language*, *66*(4), 588–611.

Kuperman, V., Pluymaekers, M., Ernestus, M., & Baayen, H. (2007). Morphological predictability and acoustic duration of interfixes in Dutch compounds. *Journal of the Acoustical Society of America*, *121*(4), 2261–2271.

Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioural and Brain Sciences*, *22*(1), 1–75.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403-439). Dordrecht, The Netherlands: Springer.

Linzen, T., & Jaeger, T. F. (2015). Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. *Cognitive Science*, *40*(6), 1382–1411.

Maekawa, K., Koiso, H., Furui, S., & Isahara, H. (2000). Spontaneous speech corpus of Japanese. Paper presented at the Proceedings of LREC2000 (Second International Conference on Language Resources and Evaluation).

McCarthy, J. J. (2008). The gradual path to cluster simplification. *Phonology*, *25*(02), 271–319.

McMillan, C. T., & Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition*, *117*(3), 243–260.

McMillan, C. T., Corley, M., & Lickley, R. J. (2009). Articulatory evidence for feedback and competition in speech production. *Language and Cognitive Processes*, *24*(1), 44–66.

Mooshammer, C., Goldstein, L., Nam, H., McClure, S., Saltzman, E., & Tiede, M. (2012). Bridging planning and execution: Temporal planning of syllables. *Journal of Phonetics*, *40*(3), 374–389.

Munhall, K. G., Ostry, D. J., & Parush, A. (1985). Characteristics of velocity profiles of speech movements. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(4), 457–474.

Munson, B. (2007). Lexical access, lexical representation, and vowel production. In J. S. Cole & J. I. Hualde (Eds.), *Laboratory Phonology* (Vol. 9, pp. 201–228). Berlin/New York: Mouton.

Ohala, J. (1983). The origin of sound patterns in vocal tract constraints. In P. MacNeilage (Ed.), *The production of speech* (pp. 189–216). New York, NY: Springer-Verlag.

Ostry, D., & Munhall, K. (1985). Control of rate and duration of speech movements. *Journal of the Acoustical Society of America*, *77*(2), 640–648.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(9), 3526–3529.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America*, *118*(4), 2561–2569.

Port, R. F., Al-Ani, S., & Maeda, S. (1980). Temporal compensation and universal phonetics. *Phonetica*, *37*(4), 235–252.

Port, R., Dalby, J., & O'Dell, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America*, *81*(5), 1574–1585.

Pouplier, M., & Goldstein, L. (2010). Intention in articulation: Articulatory timing in alternating consonant sequences and its implications for models of speech production. *Language and Cognitive Processes*, *25*(5), 616–649.

Pouplier, M., & Goldstein, L. (2014). The relationship between planning and execution is more than duration: Response to Goldrick & Chu. *Language, Cognition and Neuroscience*, *29*(9), 1097–1099.

Raymond, W. D., Dautricourt, R., & Hume, E. (2006). Word-internal/t, d/deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change*, *18*(1), 55–97.

Recasens, D., & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *Journal of the Acoustical Society of America*, *125*(4), 2288–2298.

Roon, K. D., & Gafos, A. I. (2016). Perceiving while producing: Modeling the dynamics of phonological planning. *Journal of Memory and Language*, *89*, 222–243.

Roon, K. D., Gafos, A. I., Hoole, P., & Zeroual, C. (2007). Influence of articulator and manner on stiffness. In *Proceedings 16th ICPhS 2007, Saarbrücken, 6–10 August 20*, pp. 409–412. Retrieved from https://www.researchgate.net/publication/228864646_Influence_of_articulator_an_manner_of_stiffness

Sagisaka, Y. (1985). *Onsei Gousei-no Tame-no Inritsu Seigyo-no Kenkyuu [A study on Prosodic Features for Speech Synthesis].* PhD Thesis. Waseda University, Tokyo, Japan. [In Japanese.]

Sagisaka, Y., & Tohkura, Y. (1984). Kisoku ni yoru onsei gousei no tame no on'in jikanchou seigyo [Phoneme duration control for speech synthesis by rule]. *Denshi Tsuushin Gakkai Ronbunshi [The transactions of the Institute of Electronics, Information and Communication Engineers], 67*(7), 629–636. [In Japanese.]

Scarborough, R. A. (2004). *Coarticulation and the structure of the lexicon*. PhD Thesis. University of California, Los Angeles, USA.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, *133*(1), 140–155.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(4), 379–423.

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. MacNeilage (Ed.), *The production of speech* (pp. 109–136). New York, NY: Springer-Verlag.

Shaw, J. (2007). ti/~/tʃi/contrast preservation in Japanese loans is parasitic on segmental cues to prosodic structure. In *Proceedings of ICPhS XVI*, pp. 1365–1368.

Shaw, J. A. (2012). *Metrical rhythm in speech planning: priming or predictability*. Paper presented at the Proceedings of the 14th Australasian International Conference on Speech Science and Technology (SST). Retrieved from https://campuspress.yale.edu/jasonshaw/files/2017/01/Shaw_2012_SST-2niarcq.pdf

Shaw, J. A. (2013). The phonetics of hyper-active feet: Effects of stress priming on speech planning and production. *Laboratory Phonology*, *4*(1), 159–190.

Shaw, J., & Balusu, R. (2010). Language contact and phonological contrast: the case of coronal affricates in Japanese loans. In M. Norde, B. de Jonge, & C. Hasselblatt (Eds.), *Language contact: New perspectives, Volume 28* (pp. 155–180). Amsterdam, The Netherlands: John Benjamins.

Shaw, J. A., Han, C., & Ma, Y. (2014). Surviving truncation: informativity at the interface of morphology and phonology. *Morphology*, *24*(4), 407–432.

Shaw, J. A., & Kawahara, S. (2017). The lingual articulation of devoiced /u/ in Tokyo Japanese. *Journal of Phonetics*, 1–20.

Smith, C. L. (1995). Prosodic patterns in the coordination of consonant and vowel gestures. In B. Connell, & A. Arvaniti (Eds.), *Papers in laboratory phonology IV: phonology and phonetic evidence* (pp. 205–222). Cambridge, UK: Cambridge University Press.

Son, R., & Pols, L. C. (2003). *How efficient is speech?* Paper presented at the Instituut voor Fonetische Wetenschappen (Institute of Phonetic Sciences), University of Amsterdam.

Sóskuthy, M., Foulkes, P., Haddican, B., Hay, J., & Hughes, V. (2015). Word-level distributions and structural factors codetermine GOOSE fronting. Paper presented at the 18th International Congress of Phonetic Sciences, Glasgow, 10–14 August 2015. Retrieved from http://www-users.york.ac.uk/~ms1341/soskuthy-et-al-goose-fronting-2015.pdf

Tilsen, S. (2009). Subphonemic and cross-phonemic priming in vowel shadowing: Evidence for the involvement of exemplars in production. *Journal of Phonetics*, *37*(3), 276–296.

Tilsen, S. (2014). Selection and coordination of articulatory gestures in temporally constrained production. *Journal of Phonetics*, *44*(1), 26–46.

Tilsen, S. (2016). Selection and coordination: the articulatory basis for the emergence of phonological structure. *Journal of Phonetics*, *55*(1), 53–77.

Tomaschek, F., Wieling, M., Arnold, D., & Baayen, R. H. (2013). *Word frequency, vowel length and vowel quality in speech production: an EMA study of the importance of experience*. Paper presented at the Interspeech, Lyon.

Tomaschek, F., Tucker, B., Wieling, M., & Baayen, H. (2014). Vowel articulation affected by word frequency. Paper presented at the 10th International Seminar on Speech Production, Cologne, 5–8 May 2014. Retrieved from http://www.martijnwieling.nl/files/Tomaschek2014.pdf

Torreira, F., & Ernestus, M. (2009). *Probabilistic effects on French [t] duration*. Paper presented at the 10th Annual Conference of the International Speech Communication Association, Dresden, Germany, 6–10 September 2015 (Interspeech 2009). Retrieved from http://www.mirjamernestus.nl/Ernestus/NCCFr/ProbEffectsFrench.pdf

Tsukada, K. (1999). *An acoustic phonetic analysis of Japanese-accented English*. PhD Thesis. Macquarie University, Sydney, Australia.

Vance, T. J. (2008). *The sounds of Japanese with audio CD*. Cambridge, UK: Cambridge University Press.

Warner, N., & Arai, T. (2001). Japanese mora-timing: A review. *Phonetica*, *58*(1–2), 1–25.

Watson, D. G., Buxó-Lugo, A., & Simmons, D. C. (2015). The effect of phonological encoding on word duration: Selection takes time. In L. Frazier, & E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor* (pp. 85–98). New York, NY:Springer.

Wedel, A. (2012). Lexical contrast maintenance and the organization of sublexical contrast systems. *Language and Cognition*, *4*(4), 319–355.

Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, *128*(2), 179–186.

Whalen, D. (1990). Coarticulation is largely planned. *Journal of Phonetics*, *18*(1), 3–35.

Wilson, C. (2001). Consonant cluster neutralization and targeted constraints. *Phonology*, *18*(1), 147–197.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley Press.

Appendix

***β* estimates and *t*-values for fixed factors in mixed models fit separately to each vowel.**

	/a/		/i/		/u/		/e/		/o/	
Fixed factor	*β*	*t*-value	*β*	*t*-value	*β*	*t*-value	*β*	*t*-value	*β*	*t*-value
(Intercept)	4.13	206.68	3.96	121.60	4.21	41.74	4.18	111.25	4.24	193.57
ENTROPY	0.03	7.17	0.00	–0.07	–0.01	–0.69	0.02	0.94	0.05	5.51
LENGTH	0.53	29.81	0.82	22.46	0.80	32.75	0.58	31.08	0.69	69.15
SURPRISAL	0.03	2.05	0.03	2.47	–0.03	–3.52	0.04	3.70	–0.11	–13.16
VOICING	–0.13	–17.57	–0.24	–12.21	–0.20	–14.04	–0.13	–8.29	–0.24	–24.69
VELAR	0.12	9.80	0.07	2.90	–0.32	–3.31	–0.04	–1.29	–0.02	–1.08
CORONAL	0.22	15.69	0.08	3.30	–0.23	–2.37	0.02	0.53	0.04	2.46
LABIAL	0.20	13.06	0.04	1.32	–0.17	–1.78	–0.06	–1.65	–0.09	–4.88
SONORANCY	0.02	2.66	–0.07	–3.58	–0.05	–3.56	0.03	1.65	–0.06	–5.44
SYLL_STRUC	0.20	28.30	0.23	16.30	0.25	16.44	0.22	19.53	0.28	23.73
ENTROPY: LENGTH	0.03	1.81	0.15	2.21	0.04	1.69	0.01	0.40	0.07	5.79
LENGTH: SURPRISAL	–0.14	–2.99	–0.05	–1.78	0.00	0.50	0.00	0.15	–0.06	–2.44