English Language and Linguistics, 26.3: 559–582. © The Author, 2022. Published by Cambridge University Press
doi:10.1017/S1360674322000156

Syllable weight and natural duration in textsetting popular music in English[1]

KEVIN M. RYAN
Harvard University

(Received 26 August 2021; revised 20 February 2022)

Hayes & Kaun (1996) argue that the mapping of syllables onto a metrical grid in textsetting is sensitive to natural duration, not just categorical weight (heavy or light). Most of their evidence, however, derives the final lengthening effects, which admit of another possible analysis (Halle 2004). Drawing on a corpus of 2,371 popular songs in English, I confirm that even when one controls for final lengthening and other factors, the setting of syllables to a discrete grid is sensitive to natural duration. Moreover, onset effects reveal that the domain of weight for textsetting is not the syllable, rime, or vowel-to-vowel interval, but rather the interval between p-centers (perceptual centers). Finally, I argue that the textsetting grammar invokes both natural duration and categorical weight; weight mapping cannot be reduced to one or the other.

Keywords: textsetting, language and music, syllable, syllable weight, p-center

When language is set to music or chanted, syllables are coordinated with a relatively isochronous metrical grid. This mapping, though not inflexible, is highly systematic, as reflected in singers' preferences for certain possible maps over others. At the broadest level of generalization, preferred maps between text and music involve correspondences between analogous structures of language and music, such as the matching of prominence, constituency, tone and duration.[2] For example, prominent linguistic elements such as stressed syllables tend to align with metrically prominent events in music such as downbeats (Lerdahl & Jackendoff 1983; Dell 1989; Palmer & Kelly 1992; Halle & Lerdahl 1993; Hayes & MacEachern 1998; Jackendoff & Lerdahl 2006; Kiparsky 2006; Dell & Halle 2009; Hayes 2009a, 2009b; Proto & Dell 2013; Temperley & Temperley 2013; Proto 2015; Girardi & Plag 2019; Tan *et al.* 2019; Kiparsky 2020). Second, linguistic constituency of various levels tends to align with analogous levels of musical phrasing (Halle 2004; Patel 2008; Starr & Shih 2017). For example, intervals between attacks in music tend to scale with prosodic boundary strength. Third, linguistic pitch phenomena such as lexical tone and intonation often correspond with melodic contours (Devine & Stephens 1994; Wee 2007; Schellenberg 2012; Villepastour 2014; McPherson 2018, 2019; McPherson & Ryan 2018; Ladd & Kirby 2020).

[1] I wish to thank the anonymous reviewers at *ELL* for their insightful feedback on an earlier draft of this article, which inspired several improvements.

[2] Other dimensions may correspond as well, such as timbre (Proto 2015: 118–19).

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

A fourth dimension of correspondence between language and music in textsetting, and the focus of this article, is duration matching. A number of textsetting traditions distinguish heavy from light syllables, such that the former are allotted more grid space. In some cases, such correspondence is demonstrably independent of the other dimensions of correspondence just enumerated. In Ancient Greek music, for instance, heavy syllables – regardless of accent and position – are set to long notes in the music, while light syllables are set to short notes, an opposition in note value usually nowadays transcribed using quarter and eighth notes (West 1992: 130–3; Hill 2008; see, for instance, the Delphic hymns per West 1992: 288–300). Similarly, a number of Afro-Asiatic languages exhibit living quantitative textsetting traditions in which heavy syllables are mapped onto more grid space than light syllables, including Bole (Schuh 2001), Hausa (Schuh 2011; Hayes & Schuh 2019), Somali (Banti & Giannattasio 1996) and Tashlhiyt Berber (Dell & Elmedlaoui 2008, 2017; Dell 2011). Of course, quantitative textsetting is not confined to Afro-Asiatic (e.g. Ross & Lehiste 2011 on Estonian; Proto & Dell 2013: 9–10 on Italian; Kiparsky 2020 on Urdu, among others); it is also found in English, as discussed presently. McPherson (2021) analyzes a xylophone surrogate of Seenku (Mande) in which syllable weight is mapped onto musical articulation: heavy syllables such as CV are flammed (double-struck), unlike light syllables such as level-toned CV̌.[3]

In English, too, textsetting is sensitive to weight, even when controlling for stress level and other factors. Consider the two lines in figure 1, the second a constructed comparandum (modified from Hayes & Kaun 1996: 260). The lines exhibit identical stress profiles and constituencies. In both cases, the stressed syllables map onto the tallest grid columns, which mark strong metrical positions. The lines differ in the weights of the initial syllables of *city* and *township*, a difference reflected in the textsetting: light *ci-* is mapped onto one grid space (equivalent to an eighth note or quaver), while heavy *town-* occupies two spaces (a quarter note or crotchet). Hayes & Kaun (1996) refer to these two settings as short first and long first, respectively. To be sure, other settings are possible; indeed, it would not be unmetrical to set (a) long first or (b) short first. But the long-first setting is more frequent and more felicitous with a heavy syllable.[4]

A grid representation such as in figure 1 shows only the alignment of syllables' beginnings (or attacks[5]) with metrical positions, not the alignment of syllables' offsets. In other words, it does not make explicit how long notes are held within their allotted

[3] Any syllable with a contour tone is flammed. Syllables ending with diphthongs or nasal codas are usually unflammed, but vary (McPherson 2021).

[4] An anonymous reviewer observes that the comparison between *ci-* and *town-* in figure 1 might be confounded by differences between *-ty* and *-ship*. Nevertheless, if *township* is replaced by *foundry*, my intuition of a difference in felicity between (a) and (b) remains. At any rate, regardless of intuitions, the corpus studies in section 2 and beyond furnish ample data showing that the first syllable of a disyllable is more likely to be allocated more grid space when it is heavy.

[5] Strictly speaking, it is not the beginning of a syllable that is the target for alignment with an isochronous grid, but a point closer to the beginning of the vowel, namely, the p-center (see section 4).

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

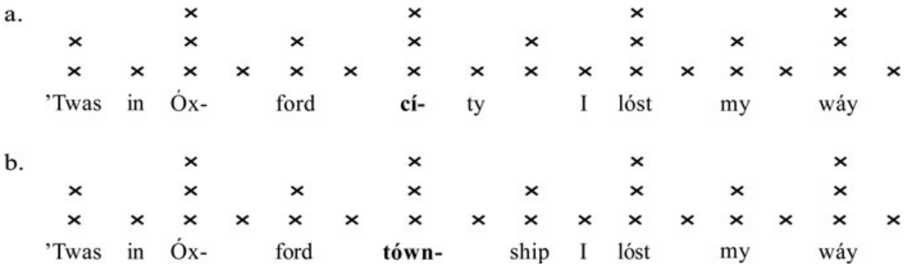
a.
'Twas in Óx- ford cí- ty I lóst my wáy
b.
'Twas in Óx- ford tówn- ship I lóst my wáy

Figure 1. Possible settings of two lines differing only in syllable weight. While both settings (a–b) are possible for both lines, (b) is more likely if the boldface syllable is heavy.

grid spaces. For example, *town* in (b) might fill both available timing slots, or the singer might pause slightly between *town-* and *ship*. This article, along with nearly all research in generative textsetting (e.g. Halle & Lerdahl 1993: 15; Hayes & Kaun 1996: 245; Keshet 2006: 3–4; Dell & Halle 2009: 65), treats the alignment of attacks with the grid and the allotment of grid space between attacks, not the orthogonal question of how syllables are performed for a given setting.[6] As Hayes & Kaun (1996: 245) explain, 'for rhythmic purposes it is not particularly crucial where a note ends'. In order to emphasize that it is grid space rather than note value that is being modeled, the present article employs the terms 'quarter space' and 'eighth space' rather than 'quarter note' or 'eighth note'. ('Space' should not be confused with 'empty space', that is, a rest.) For instance, in figure 1, *town-* occupies a quarter space, regardless of how legato or staccato it is articulated. Both a quarter note and an eighth note followed by an eighth rest constitute quarter spaces.

Returning to the role of weight in English textsetting, Hayes & Kaun (1996: 263–7) demonstrate through textsetting experiments with ten participants (in addition to a set of original scores) that, first, non-word-final heavy syllables are significantly more likely than non-word-final light syllables to be set to quarter as opposed to eighth spaces (to use the present terminology). Although this first finding is compatible with both categorical and gradient weight, they go on to argue that textsetting is sensitive to gradient duration based on the behavior of word-final heavy syllables. As they note, the degree of lengthening of a phrase-final syllable correlates monotonically with the strength of the prosodic juncture that it precedes. For instance, the ultima of an intonation group tends to be more prolonged than the ultima of a phonological phrase that is not final in its intonation group. Likewise, in textsetting, Hayes & Kaun (1996) find that the juncture strength following a syllable correlates monotonically with the syllable's space apportionment. An interpretation of these results is that textsetting responds to the fine-grained, natural durations of syllables.

[6] The question of how syllables are rendered is also linguistically interesting (see e.g. Katz 2010: 127–33; Nwe *et al.* 2010; Hayes & Schuh 2019: 284–93).

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

Halle (2004), however, discusses a consideration that potentially undermines this inference (see also Oehrle 1989: 104–17 for a similar argument): space allocation might respond directly to juncture strength without being mediated by syllable duration. There is, after all, evidence for constituency matching in textsetting that is independent of syllable duration mapping. Indeed, in some cases, the need for direct constituency mapping is trivial. For instance, a sentence ending with a schwa might be followed by a measure of rest in the music. In this case, the musical phrasing corresponds with the linguistic phrasing, but the spacing cannot be attributed to the duration of the schwa. Halle (2004: 6) offers the example of oronyms, that is, phonetically near-identical phrases that differ in constituency (see also Oehrle 1989: 104). For example, *need* in *we need a decanter* is more felicitously set to a quarter space than the same morpheme in *we needed a cantor*.[7] As another illustration (not Halle's) of the necessity of direct constituency mapping, consider *fountain of youth* versus *tip of the hat* in figure 2, in which *tip* is more felicitously set to a quarter space than *foun-*. Because *tip* is shorter than *foun-*, the bias in grouping must be due to constituency mapping rather than syllable duration. Thus, it is possible that Hayes & Kaun's (1996) final lengthening results are driven by constituency matching rather than natural syllable duration. That said, Hayes & Kaun (1996) present another test that is not susceptible to this confound: stressed, heavy syllables are significantly more likely to be set to short spaces as antepenults than as penults, a difference that they attribute to the gradient effect of polysyllabic shortening (1996: 297).

Besides Hayes & Kaun (1996), few studies probe whether durational correspondence in textsetting is sensitive to natural duration. Girardi & Plag (2019), for instance, find that note length in English textsetting correlates with both stress level and poetic-metrical strength, but they do not analyze syllable weight. Hayes (2009b) and Keshet (2006) employ a constraint Strong-is-Long, which penalizes short spaces after prominent beats in the music, but this constraint does not refer to syllable weight.[8] The converse constraint Long-is-Strong does invoke weight, but only binary weight (Hayes & Schuh 2019). San & Turpin (2021) employ Final-is-Long, which requires a measure-final syllable to span three positions, regardless of weight. Beyond English, most research on quantitative textsetting treats weight as categorical. For example, Hayes & Schuh (2019: 284–93) model the durations with which syllables are sung in Hausa, but the model takes as inputs only discrete moras and syllables, not natural duration. McPherson (2021: 11–14) goes further, finding that performative duration in Seenku is affected by gradient prosodic effects such as phonetic closed-syllable shortening. Finally, Gilroy (2021) reports two experiments with 87 participants that test whether vowel tenseness and coda voicing influence textsetting in English. In one of the experiments (2021: 45– 51), both effects are significant. Because all the stimuli are monosyllables, the effects must be attributed to intra-heavy durational differences.

[7] The example is not perfect, since *need* as a monosyllable is longer than the first syllable of *needed*.

[8] Keshet (2006: 25–6) does, however, claim that superheavy syllables behave as if they are stressed even when they are unstressed, another possible effect of intrinsic syllable quantity on textsetting.

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

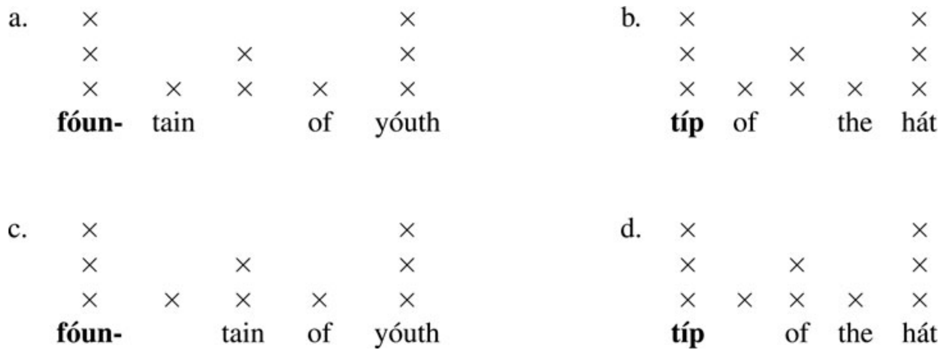

Figure 2. An illustration of the need for constituency matching independent of weight mapping. Despite the syllable *foun-* being greater in natural duration than the syllable *tip*, the former is more likely to be set to a shorter space in (a–d). Especially at a slow tempo, (a) is more felicitous than (b), and (d) more than (c).

Natural duration, which might also be called intrinsic or inherent duration, refers to the timing of spoken language as independent from textsetting. With Hayes & Kaun (1996), Hayes & MacEachern (1998) and Kiparsky (2006), I assume a modular approach to textsetting, such that correspondence constraints have access to natural prosody, allowing natural prosody to influence the realization of language as sung or chanted, though textsetting imposes its own constraints, modifying that prosody (e.g. Hayes & Schuh 2019: 284–93). Natural duration in principle subsumes all systematic aspects of timing in spoken language, including phonemic length, intrinsic segmental duration (e.g. duration correlating with vowel height), adjustments due to segmental context (e.g. vowel shortening in closed syllables), lengthening under stress, final lengthening, polysyllabic shortening, gradient compensatory effects, and so forth. It is an empirical question whether all these properties in fact influence textsetting. This article finds that at least some of them do, and therefore endorses the position that weight mapping for textsetting relates a continuous dimension (duration) to a categorical dimension (the discrete grid).

In order to probe the nature of weight mapping while avoiding any possible interference from constituency matching, the tests in this article focus on syllables in a fixed word-internal position, namely, the stressed initials of disyllables. Each test examines whether a given factor impacts the rate with which syllables are allocated quarter versus eighth spaces, a categorical distinction. Hayes & Kaun's (1996) claim that textsetting invokes natural duration is supported throughout. The present article newly documents several ways in which subcategorical weight affects textsetting, including intrinsic vowel duration (section 2), coda complexity (section 2), onset complexity (sections 3–4) and the compression of vowels after filled onsets (section 4). These findings also bear on the domain of weight, conventionally assumed to be the syllable. I argue, however, that weight must be assessed over p-center intervals (the spans

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

between successive perceptual centers), as this motivates the array of onset effects documented in sections 3–4. Finally, in sections 5–6, I maintain that categorical weight cannot be dispensed with altogether: even though weight for textsetting is sensitive to natural duration, categories are more polarized than a linear effect of duration alone would predict. Grammars are thus hybrid, incorporating both categorical and gradient aspects of weight.

1 The corpus and annotation of grid spacing

As a large annotated corpus of contemporary English popular music, I use DALI 2 (***D****ataset of synchronised* ***A****udio,* ***L****yr****I****cs and notes*; Meseguer-Brocal *et al.* 2018). The anglophone portion of DALI 2, as employed here, comprises 5,913 songs from 2,020 artists, mostly American. Release dates range from 1938 to 2017 with a median of 2003. The songs are mostly mainstream releases, with pop, rock and alternative being the most frequent genres. The three artists with the most songs in the corpus are Glee Cast, the Beatles and Demi Lovato.

DALI 2 contains data based on song recordings, not scores. Among other features, it provides the start time of each syllable as well as the duration with which it is held, both with nominal millisecond precision (Simpson *et al.* 2015; Meseguer-Brocal *et al.* 2018). DALI 2 does not provide annotations for grid alignment or note value. I add annotations for grid space as follows. An inter-onset interval (IOI) measures the time elapsed between the starts of two successive syllables, including any intervening silence. In most songs, especially those with fixed tempos – drum machines are commonly used in pop recordings – IOIs cluster around certain recurrent values. For example, figure 3 shows the density distribution of IOIs in the song ‘I hate this part’ by the Pussycat Dolls. The peaks of the distribution fall near multiples of 266 ms, with

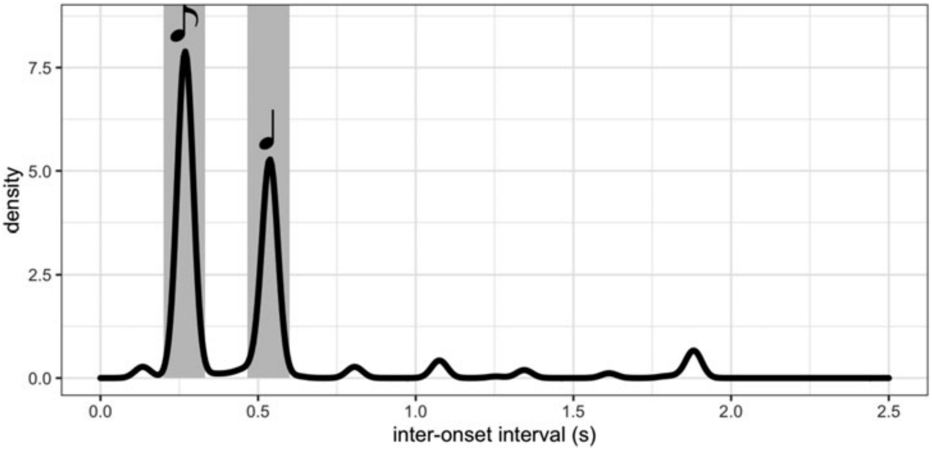

Figure 3. Distribution of inter-onset intervals (IOIs) in the Pussycat Dolls’ ‘I hate this part’. The two highest peaks correspond to eighth and quarter spaces, with shading indicating accepted ranges.

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

the two tallest peaks corresponding to eighth and quarter spaces. Smaller peaks correspond to sixteenth spaces (0.5 × 266 ms), dotted quarter spaces (3 × 266 ms), half spaces (4 × 266 ms), and so forth. Variance around each peak – the width of the mound – reflects the singer not hitting the beat precisely due, for instance, to stylistic choices or changing tempo. Also contributing to variance is the fact that singers do not generally seek to align the beginnings of syllables with beats, but rather their p-centers, which are closer to the beginnings of nuclei (section 4).

To annotate quarter and eighth spaces for the whole music corpus, a density curve like the one in figure 3 is generated for each song. The two tallest peaks are retrieved. If the peak on the left is approximately half (45% to 55%) the duration of the peak on the right, these two maxima are taken to represent median eighth and quarter spaces, respectively. If no such ratio is evident, the song is excluded, leaving 2,371 usable songs in the corpus. Each usable song's spaces are then classified as eighth, quarter, or other, allowing some variance around the median (±25% the duration of the eighth) to accommodate inexact syllable-beat alignment. In figure 3, these bands are shaded. Space categories are distributed as follows: 46% eighth, 24% quarter and 30% other (excluded).

The distribution of IOIs in the songs used in this study, which exhibit relatively rigid grids, contrasts sharply with the distribution of IOIs in English as naturally spoken. Figure 4 depicts the latter, based on the first speaker in the Buckeye corpus of conversational English (Pitt *et al.* 2007). In natural speech, IOIs are unimodally distributed. The contrast between figures 3 and 4 makes explicit that the regularly spaced peaks found in the musical corpus are due to the demands of textsetting, not the mechanics of natural speech. Singing, with its more rigid grid, discretizes the distribution: natural duration is shoehorned into categories corresponding to regular grid spaces.

One caveat about this methodology is that the labels 'quarter' and 'eighth' are somewhat arbitrary absent sheet music. What stops us from treating the two tallest

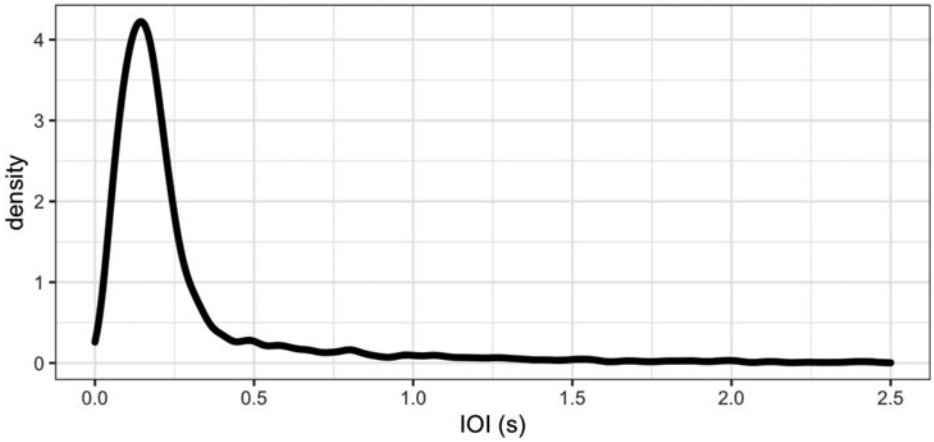

Figure 4. Distribution of IOIs in natural English speech

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

peaks in figure 3 as, say, quarter and half spaces? For one thing, quarter and eighth notes predominate in the genre. For another, if the peak at 266 ms corresponded to a quarter note, the beats per minute would be a genre-defying 226. Finally, and most importantly, the labels are not critical to the tests that follow. What matters is that the target spaces are categories standing in a 1:2 timing relation. Indeed, this point holds implicitly of any study of a primarily oral genre, in which transcriptions are largely post hoc, including the English folk songs analyzed by Hayes & Kaun (1996), Hayes & MacEachern (1998) and Kiparsky (2006).[9] 'Quarter' and 'eighth' in this article might also be labeled 'long' and 'short'.

A second caveat is that, as discussed above, grid space values do not encode how long notes are held. In most instances, note value and grid space are the same. However, for a note before a rest, grid space exceeds note value. For example, a quarter note before three beats of rest occupies a whole-measure space. In the tests below, this divergence between note values and grid spaces is unimportant, for two reasons. First, nearly all the tests consider only non-word-final syllables. In the present corpus, word-internal pauses are rare, meaning that note value and grid space are nearly always equivalent for the syllables in question. Second, the tests consider only quarter and eighth spaces, effectively excluding notes at the ends of musical phrases, which often precede rests. At any rate, grid space is not intended to be a proxy for note value. As discussed, rhythm is instantiated principally by the timing of attacks, not the timing of offsets.

2 Beyond binary weight

All else being equal, heavy syllables tend to be mapped onto larger grid spaces than light syllables. With this first test, I confirm this tendency while controlling for stress level and position in the word by considering only stressed initials of disyllables. The division between light and heavy syllables is nuanced in English (Moore-Cantwell 2021), but for present purposes, I follow Olejarczuk & Kapatsinski (2018: 385) in treating syllables ending with monophthongs except /i, u/ as light and those ending with /i, u/, codas, diphthongs and syllabic consonants (including the rhotic) as heavy. In DALI 2, heavy syllables as initials of disyllables are mapped onto quarter (i.e. long) as opposed eighth (i.e. short) spaces 42% of the time, a rate almost twice that of light syllables in the same context (24%) and a significant difference (Fisher's exact test $p < 0.0001$). The result is the same regardless of how less securely long or short vowels such as /i, u, ɑ, ɔ, ɝ/ are classified.

Beyond binary weight, natural duration influences textsetting. For example, take vowel identity in stressed initials of words of the form $C\acute{V}CVC_0$.[10] Figure 5 shows the rate at

[9] Indeed, even within the realm of scored compositions, the choice between certain time signatures (e.g. 3/4, 3/8, 6/4, 6/8, etc.), and hence between which notes are written as quarter versus eighth, can often be arbitrary. I thank an anonymous reviewer for discussion of this point.

[10] C stands for consonant; V for vowel; the acute indicates stress; and subscripts indicate lower bounds (e.g. C_0 is any number of consonants).

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

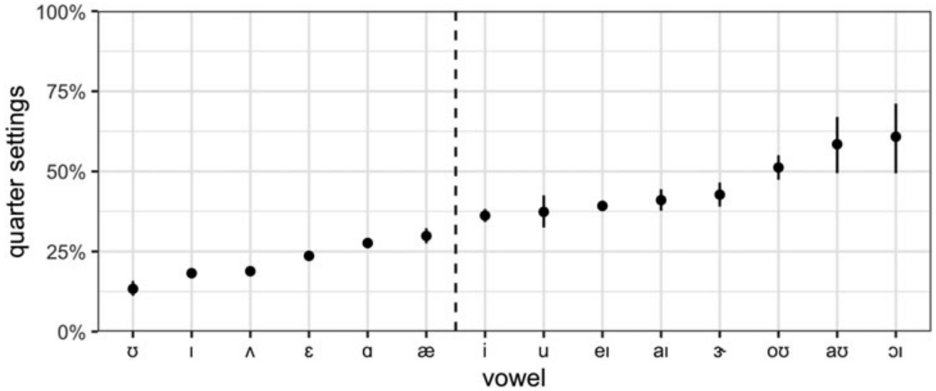

Figure 5. The percentage of the time that CV syllables (as stressed initials of disyllables) are mapped onto quarter as opposed to eighth spaces, by vowel. Error bars are Wilson scores. The dashed divider is a commonly assumed cutoff for light versus heavy CV syllables.

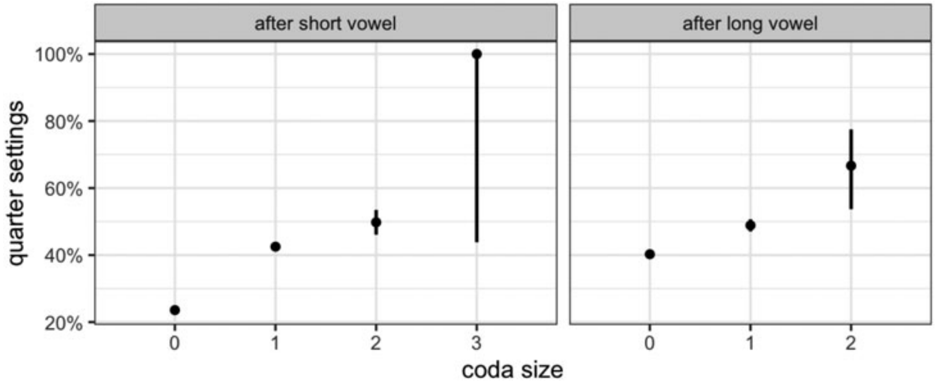

Figure 6. The percentage of quarter and eighth spaces that are quarter spaces as a function of coda size in syllables with short vowels (left) and long vowels (right)

which vowels in this context are mapped onto quarter versus eighth spaces. On the one hand, the distinction between heavy and light syllables (or tense and lax vowels, as in Gilroy 2021) is once again in evidence. Long vowels, which render open syllables heavy, are consistently more quarter space-aligned than short vowels. On the other hand, a role for gradience is obvious. Within each weight class, phonetically long vowels are increasingly likely to be mapped onto quarter spaces. Although some vowels, such as /æ/, vary substantially by dialect, some generalizations about vowel duration are stable across dialects, including /ɪ, ʊ/ being the shortest of the short and /aʊ, ɔɪ/ being the longest of the long (Umeda 1975; Jacewicz *et al.* 2007). These trends are reflected as such in figure 5 and cannot be motivated by a textsetting model that invokes only binary weight. Indeed, given the apparently seamless transition from light to heavy syllables, figure 5 creates the impression that textsetting might ignore binary weight altogether, heeding

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

only natural duration. Nevertheless, the question of whether both categoricity and gradience are motivated requires statistical scrutiny, to which I return in section 5.

Additional evidence for the insufficiency of binary weight can be found in coda complexity. Consider once again stressed initials of disyllables, now relaxing the frame to ignore margin complexity: $C_0\acute{V}C_0VC_0$. As figure 6 reveals, with each consonant added to the coda, quarter maps increasingly predominate eighth maps. For instance, among syllables with long vowels (right panel), which are all categorically heavy, both contrasts in the chain Ø < C < CC are significant, as reinforced by the Wilson intervals. These contrasts add to the contrasts among open syllables in figure 5, which are subsumed by Ø in figure 6.

3 Syllables versus intervals in textsetting

The previous section established that the metrical categorization of syllables is not solely a function of binary weight. This section turns to the domain of weight, which is conventionally assumed to be the syllable or rime. An alternative possibility is that weight reflects the total vowel-to-vowel interval (Steriade 2008, 2012, 2019: 174–6; Hirsch 2014; Ryan 2016: 725–6; Lunden 2017). As an illustration, in the word *pregnancy*, the syllables are [pɹɛg, nən, si], while the intervals are [ɛgn, əns, i], as in figure 7 (cf. Meyer *et al.* 2012: 688).

Insofar as syllable weight is based on the rime, the beginning of the weight domain is the same for syllables and intervals, being the beginning of the vowel/nucleus. The main difference between the proposals concerns the right edge of the domain: with syllables, only some intervocalic consonants are parsed into the preceding vowel's domain, namely, those that can be syllabified as (parts of) codas, whereas with intervals, all intervocalic consonants are parsed into the preceding vowel's domain, regardless of phonotactics. Several arguments for intervals have been put forth (cited above), though the evidence is not clearcut (see Ryan 2016: 725–6 for a synopsis). As one argument, intervals are claimed to account better for the domain of durational invariance: the duration of a vowel trades not just with the duration of its coda, but also with that of the following onset. In this section, I present four statistical tests pitting syllables against intervals as domains of weight for English textsetting. All four favor intervals.

First, I examine grid space allocation as a function of onset size of the following syllable, limiting the data to stressed, non-word-final syllables. As figure 8 shows, even while holding the first (stressed) syllable's rime shape constant, each consonant added

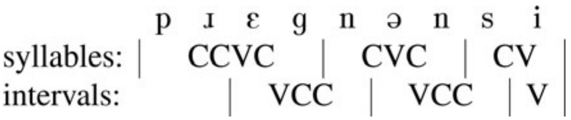

Figure 7. Syllable versus vowel-to-vowel interval parses of *pregnancy*. Under both approaches, the word comprises three spans.

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

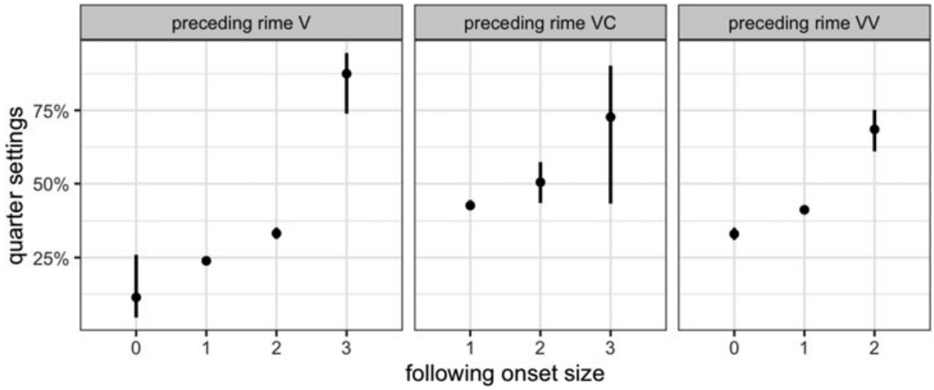

Figure 8. Percentage of quarter as opposed to eighth maps as a function of the size of the trailing onset (onset of the following syllable), grouped by the rime type of the preceding syllable. Only the three most frequent rime types are shown.

to the onset of the following syllable increases the first syllable's propensity to be mapped onto a quarter space. This result is expected with intervals, where the whole interlude counts towards the weight of the first vowel's span, but not with syllables. Indeed, syllables make the opposite prediction: as consonants are added to the following onset, the preceding syllable tends to be compressed, and would therefore be expected to be, if anything, lighter.[11] Syllable divisions here follow CELEX (Baayen *et al.* 1993). Depending on one's theory of English syllabification, one might take issue with certain syllabifications provided by CELEX. For example, where CELEX provides [ɛk.stɹə], one might instead consider [ɛks.tɹə] or ambisyllabicity. CELEX uses onset maximization, as is widely employed for English (Kahn 1976; Selkirk 1982). Other dictionaries use other schemes (Bartlett *et al.* 2009: 313–14). This test therefore favors intervals over syllables with onset maximization, but does not rule out other possible syllabification algorithms. The remaining three tests are more tightly controlled so as not to be ambivalent in this way.

Second, consider words of the shape C_0V́V(C)VC_0, where V́V is a stressed long vowel or diphthong, such that the initial syllable is always heavy.[12] Intervals predict that the initial of C_0V́V(C)VC_0 should be heavier when the medial consonant is present: V́VC > V́V. This prediction is correct: When the consonant is present, the initial span is mapped onto quarter spaces 41% of the time, versus 33% when it is absent (Fisher's exact test $p < 0.0001$). This difference cannot be explained by syllables. More specifically, with syllables, there are two conceivable explanations for the difference,

[11] In the Buckeye corpus, for instance, non-word-final C_0V syllables, where V is short, have means of 194, 134 and 130 ms, respectively, preceding onsets of zero, one and two consonants. C_0VC syllables in the same frame have means of 190, 173 and 177 ms preceding onsets of one, two and three consonants.

[12] I restrict the first vowel to being long because when no consonant intervenes between the vowels, the first vowel is nearly always long.

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

neither of which turns out to be viable. First, if $C_0\acute{V}V$ were durationally longer in $C_0\acute{V}V(C)VC_0$ when the medial consonant was present, one could claim that the duration of the first syllable was driving the effect. However, the opposite is true. Based on the Buckeye corpus, $C_0\acute{V}V$ is *shorter* in $C_0\acute{V}V(C)VC_0$ when the medial consonant is present (mean 172 when present versus 198 ms, $t = -11.6$, $p < 0.0001$). A second potential defense of syllables begins with the assumption that an intervocalic consonant can be syllabified as a coda. (If one is not willing to make this assumption, this second defense is unavailable.) For $C_0\acute{V}VCVC_0$, such a parse is most plausible when $C_0\acute{V}VC$ is a root and VC_0 is either an inflectional suffix (e.g. *-ing*) or the second member of a compound (e.g. *weekend*). Bearing this in mind, I divide all $C_0\acute{V}VCVC_0$ words into two groups. The first group ($n = 5{,}523$) comprises words ending with a VC_0 inflectional affix or compound member.[13] The second group ($n = 6{,}806$) comprises the remaining $C_0\acute{V}VCVC_0$ words. Words of the first group are more likely to be syllabified as VC.V than words of the second group. Nevertheless, the two groups exhibit essentially the same rate of quarter maps, at 41.3% and 40.8%, respectively (Fisher's exact test $p = 0.61$). This means that the observed difference between $C_0\acute{V}VCVC_0$ and $C_0\acute{V}V.VC_0$ cannot be attributed to a subset of the former being syllabified with a medial coda. Syllables have no way out.

As a third test, consider words of the shape $C_0\acute{V}C_0.C1(C2)VC_0$, where C1 is a plosive and C2 is a liquid, excluding sequences *tl* and *dl* as well as words ending with *-less*. Intervals, unlike syllables, predict that the addition of C2 should contribute to the weight of the first span. The prediction is correct. The initial span of this frame is quarter-set 49% of the time when the following syllable starts with a plosive-liquid cluster, versus 33% of the time when it starts with only a plosive (Fisher's exact test $p < 0.0001$).

Finally, consider the duration of the single intervocalic consonant in the frame $C_0\acute{V}CVC_0$. As an estimate of the natural duration of each consonant in this context, I take its mean duration as the medial consonant of $C_0\acute{V}CVC_0$ in the Buckeye corpus. Means thus obtained range from 30 ms for [ɾ] to 129 ms for [ʃ]. In a logistic regression with grid mapping (quarter versus eighth) of the initial syllable as the dependent variable and intervocalic consonant duration and vowel quality (14 levels) as independent variables, consonant duration is positive and significant ($t = 3.3$, $p < 0.001$). This outcome is expected with intervals, since the intervocalic consonant is always grouped with the preceding vowel. It is unexpected with syllables, where either the consonant serves uniformly as the onset of the following syllable, or, if one admits VC.V parses, is not expected to vary systematically in its coda versus onset status as a function of its intrinsic duration. In fact, the outlook for syllables is grimmer still, since vowels tend to shorten before longer consonants in the frame VCV (Farnetani & Kori

[13] As VC_0 inflectional affixes, I include *-ed*, *-en*, *-er*, *-or*, *-es*, *-est* and *-ing*. As observed VC_0 compound endings, I include *age*, *end* and *up*. It is not critical that this list be exhaustive. Indeed, stochasticity is likely to be involved, meaning that no list would be perfect (Olejarczuk & Kapatsinski 2018). What matters here is that the first group exhibits an aggregately higher propensity for VC.V than the second.

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

1986; Fant & Kruckenberg 1989; McCrary 2004). Thus, if textsetting were sensitive to the durations of syllables, one would expect the effect to go in the opposite direction, with V1 behaving as lighter before a longer intervocalic consonant.

4 Onsets and the p-center interval

Having established that the vowel-to-vowel interval better characterizes the domain of weight for textsetting than does the syllable or rime, this section turns to the role of leading onsets. Under traditional views of both the syllable and the interval, the onset preceding a vowel – that is, the span's *leading* onset – is not expected to affect the weight of the span. Under syllables, this expectation follows from rime-based weight (Halle & Vergnaud 1987) or onsets not serving as heads of moras (Hyman 1985). Similarly, under intervals as characterized by Steriade (2008, 2012, 2019: 174–6) and others, the interval begins with the vowel. Thus, an onset is either grouped with the preceding vowel's span or, if no vowel precedes, is extraprosodic. In short, intervals predict onsets to affect weight, but only trailing onsets, not leading ones.[14]

Nevertheless, evidence has been mounting from across languages and weight-sensitive phenomena that leading onsets can contribute to weight, and not simply in ways that can be attributed to modulations of the following vowel (Ryan 2011b, 2014, 2016: 726–8, 2019; cf. Gordon 2005). For example, they influence stress placement in English, with increasingly long onsets increasingly attracting stress (Ryan 2014). (Note that onset length negatively correlates with the duration of the following vowel; thus, a rime or vowel-to-vowel interval is shorter, not longer, following a longer onset.) In this section, I demonstrate that leading onsets affect weight in textsetting.

As before, consider words of the form $C_0\acute{V}C_0VC_0$. As shown in figure 9, as initial onset size increases, so does the incidence of quarter settings. This is confirmed by a logistic regression including predictors for onset size (four levels), categorical weight (two levels), vowel identity (14 levels), and the interaction of onset size and categorical weight. Regardless of whether categorical weight is computed using syllables or intervals,[15] the onset factor is significant ($t = 6.0$ and $t = 7.2$, respectively, both $p < 0.0001$). The one exception to the trend, as apparent in figure 9, is the contrast between null and simple onsets, with the former patterning as heavier. This reversal can be attributed to the duration of the vowel: vowels are significantly longer after null onsets than after simple onsets in English (Fowler 1983; van Santen 1992; Clements & Hertz 1996; Katz 2010). For example, in words of the shape $(C)\acute{V}C_0VC_0$ in the Buckeye corpus, the initial vowel is on average 21% longer after a null onset. Given the sensitivity to vowel duration seen previously in figure 5, it is not surprising that an

[14] With either syllables or intervals, a leading onset could potentially affect the weight of the following span by virtue of affecting the duration or other psychoacoustic properties of the following vowel (Gordon 2005). Insofar as this workaround is unavailable, leading onsets are predicted not to affect weight.

[15] A heavy interval is one with three or more timing slots. This is the slot-counting metric that best correlates with the rime-based criterion.

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

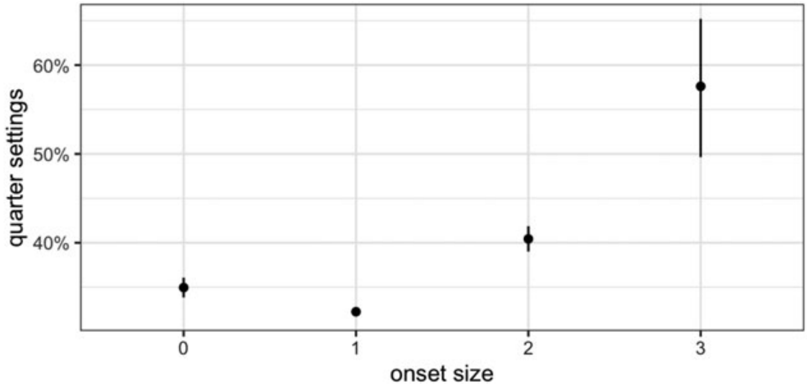

Figure 9. The percentage of the time that initial syllables of words of the shape $C_0\acute{V}C_0VC_0$ are mapped onto quarter as opposed to eighth spaces as a function of initial onset size. As discussed, the trend-defying uptick for null onsets is expected with natural duration mapping.

increase of this magnitude is felt. This increased vowel duration may or may not be further augmented by the prothetic glottal stop that often accompanies a stressed, word-initial vowel in English.[16]

In the previous section (section 3), medial onsets were shown to contribute to the weight of the interval initiated by the preceding vowel. I now test whether they also contribute to the weight of the following interval. Consider the onset of the ultima of words of the form $C_0VC_0VC_0$. As figure 10 reveals, the same effect of onset size that is found for word-initial onsets obtains also for word-medial onsets, both for unstressed (left panel) and stressed (right panel) ultimas. The slight reversal from null to simple similarly recapitulates figure 9. Logistic models, as above, confirm the significance of the trend observed in figure 10 ($t = 6.2$ and $t = 6.3$, both $p < 0.0001$).

Thus, a leading onset contributes to the weight of its span in textsetting, an effect predicted neither by rime-based weight nor by vowel-to-vowel intervals. Given that intervals were found in section 3 to be superior to syllables for modeling weight in textsetting – only intervals capture trailing onset effects — the question is how to reconcile intervals with leading onset effects. One logically possible explanation that does not work is to invoke the effect of onset complexity on the duration of the following vowel. If vowels were longer after longer onsets, an explanation of onset weight would be available without modifying the vowel-to-vowel span. However, as footnote 16 points out, vowels are progressively shorter, not longer, after progressively longer onsets.

[16] Just as vowels are shorter after one consonant than after zero, the trend continues through increasingly complex onsets: vowels are shortened on average by 8.0% when going from one to two consonants in Buckeye, and by another 3.2% when going from two to three (cf. also Fowler 1983, van Santen 1992 and Katz 2010 for agreeing results). Nevertheless, these smaller degrees of compression are insufficient to counteract the perturbation of weight induced by each additional onset consonant.

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

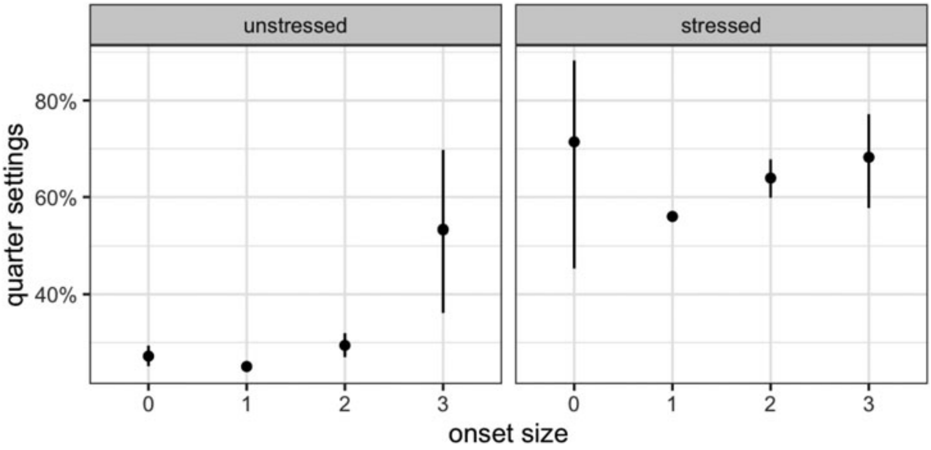

Figure 10. The percentage of the time that final syllables of $C_0VC_0VC_0$ words are mapped onto quarter as opposed to eighth spaces as a function of final onset size, paneled by the stress level of the ultima.

A viable solution is a modification to interval theory that I have elsewhere termed the p-center interval (Ryan 2016: 727, 2019: 239). The p-center, or perceptual center (Morton *et al.* 1976), refers to the time point in a syllable at which the beat is felt, or, from a production standpoint, the target of alignment between tactus and syllable. Even in the absence of an overt grid, p-centers are targets for isochrony. For example, in reciting a list of monosyllables, p-centers are targeted for even spacing. As has long been recognized (e.g. List 1974: 368), when entraining syllables with an isochronous grid, singers do not generally attempt to align the beginnings of syllables with beats. Rather, the target for isochrony is closer to the beginning of the vowel.[17] For instance, Bravi (2016) examines the alignment of syllables with beats in nonsense Italian songs containing only the syllables *ma* or *pa*, finding that the point of alignment is closer to the left edge of the vowel than that of the consonant (2016: 444–5). Likewise, Seifart *et al.* (2018) find that the drum strikes used to imitate the tone and rhythm of the Amazonian language Bora approximate the intervals between vowels rather than syllables. McPherson (2021) finds the same for the xylophone surrogate of Seenku.

Nevertheless, the p-center cannot be identified precisely with the beginning of the vowel (or nucleus): the target deviates from the vowel edge depending on the constitution of the syllable. For one thing, adding consonants or duration to the onset tends to shift p-centers leftwards relative to the vowel (Cooper *et al.* 1986; Villing 2010).[18] For example, the p-center anticipates the vowel more in *spa* more than *ba* (Port 2007: 509). Specifically, the p-center is on average 24 ms earlier relative to the

[17] One can confirm this by uttering syllables to oneself while clapping.

[18] Properties of the rime, including vowel length (Fox & Lehiste 1987a) and coda presence (Fox & Lehiste 1987b), also affect p-center placement.

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

vowel in *spa* than in *ba* in the Harvard-Haskins Database of Regularly Timed Speech (Patel *et al.* 1999). This offset is considerably less than the difference in duration between the two onsets. Marcus (1981: 253) regresses on the durations of the onset and rime to predict p-center placement, finding both to be significant, but with substantially different coefficients. Ryan (2014: 329) shows that p-centers steadily drift leftwards relative to the vowel as onset complexity increases in English; see also Šturm & Volín (2016) on Czech and Barbosa *et al.* (2005) on Brazilian Portuguese. Franich (2018) finds that prenasalized syllables exhibit earlier p-centers in Medʉmba (Bantu).

In sum, the p-center is an event (or probability distribution over events) that approximates the beginning of the vowel, but incorporates some (not all) of the duration of the onset, at least for longer onsets.[19] The p-center interval is therefore a good candidate for the domain of weight in textsetting. On the one hand, it allows for trailing onset effects (section 3): The span between p-centers includes not just a rime, but a portion of the following onset. On the other hand, it allows for leading onset effects (this section): in the presence of a longer onset, the p-center is realized slightly earlier, expanding the domain of weight. This approach is thus superior to the three alternatives of vowel-to-vowel intervals (which fail to capture leading onset effects), whole-syllable weight (which fails to capture trailing onset effects) and rime-based weight (which captures neither leading nor trailing onset effects).

5 On the coexistence of categoricity and gradience in weight mapping for textsetting

While weight mapping for textsetting is sensitive to natural duration, that does not preclude it from also being sensitive to categorical weight. Indeed, this section presents two tests whose results suggest that textsetting invokes both duration and categorical weight. First, consider once again the effect of vowels' intrinsic durations on grid space allocation. Figure 11 is based on the stressed vowels of $C\acute{V}CVC_0$ words. As before, duration correlates with grid space: the longer the vowel, the more likely it is to be set to a quarter as opposed to an eighth space. But the figure now also separates vowels by categorical weight (as defined in section 2), revealing that long and short vowels (i.e. vowels that render CV syllables heavy or light, respectively) are more polarized than duration alone would predict. Intrinsic duration is estimated using the Buckeye corpus, taking only stressed, initial vowels in $C\acute{V}CVC_0$ words. The independent effect of categorical weight is supported by logistic regression. A model with predictors for both categorical weight and duration significantly outperforms the subset model with duration alone ($\Delta\text{AIC} = -475$; $p < .0001$). Categorical weight and duration have roughly similar effect sizes, with scaled coefficients of 0.32 and 0.21, respectively. Their interaction is non-significant ($p = .38$), meaning that duration has a similar effect within each category.

[19] On theories about how p-centers are computed for acoustic signals, see, for example, Villing (2010). For present purposes, it suffices to highlight some broad empirical properties of p-centers that make them appealing as an approach to weight in textsetting.

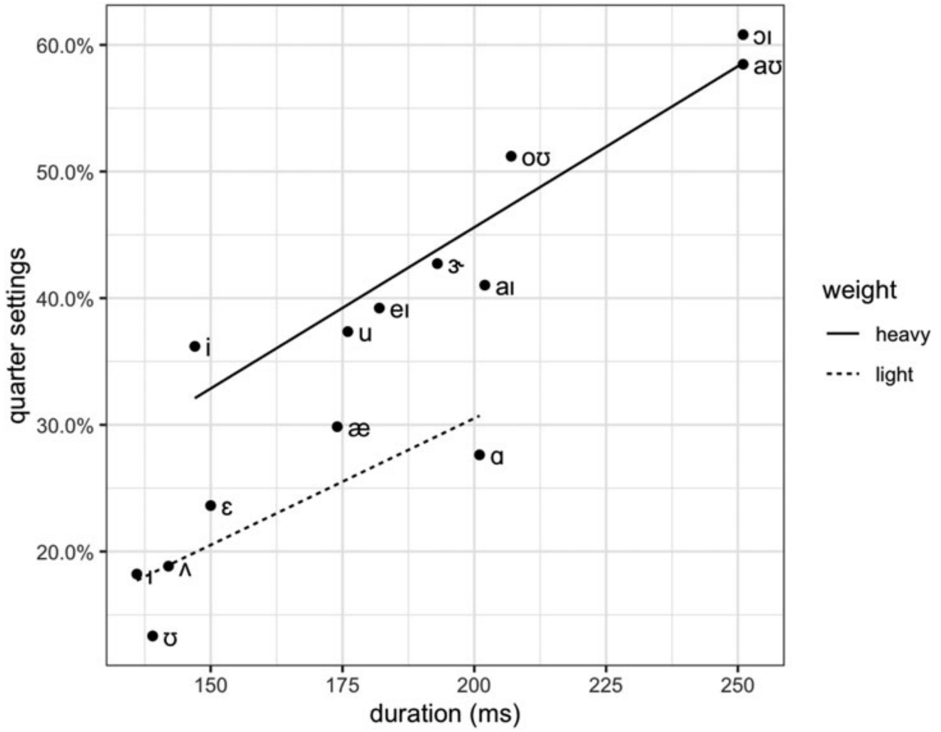

Figure 11. Categorical weight and intrinsic segmental duration exert independent effects, based on stressed, initial CV syllables in disyllables

A second, more omnibus test supports the same conclusion. Consider initial, stressed intervals of the more inclusive frame $C_0\acute{V}C_0VC_0$. For each vowel-to-vowel interval $\acute{V}C_0$ in this context, I compute the percentage of the time that it is mapped onto quarter as opposed to eighth spaces, as well as the mean duration of that interval in the same word context in Buckeye.[20] I exclude any interval that is unattested or attested only once in either corpus. As figure 12 illustrates, duration correlates gradiently with quarter settings among both heavy and light intervals. Moreover, there is a significant additive effect of category membership (ΔAIC=$-$522; $p < .0001$). As usual, light versus heavy is operationalized for intervals in terms of timing slots, such that any interval with three or more slots is heavy. Once again, categorical weight and duration have roughly similar effect sizes, with scaled coefficients of 0.26 and 0.35, respectively. Additionally, the interaction of categorical weight and duration is now significant ($p < .0001$), with the effect of duration being greater among light syllables.

[20] I take vowel-to-vowel intervals here because these are more accurate than syllables or rimes as the domain of weight (section 3). Although section 4 argued for p-center intervals over vowel-to-vowel intervals, the former cannot be measured directly from an acoustic signal; thus, the test employs the latter as a proxy.

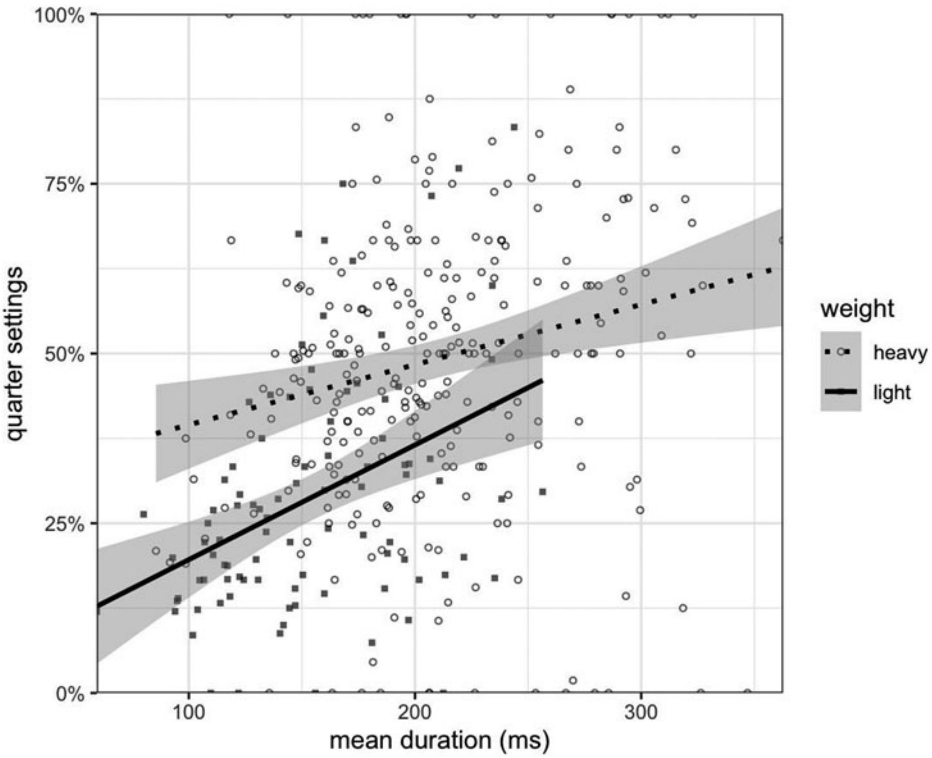

Figure 12. Quarter as opposed to eighth settings as a function of the mean duration of the interval, split by categorical weight

This is, perhaps, to be expected, given that each millisecond added has a proportionally smaller impact on a heavy syllable.

6 Conclusion

In the setting of text to music, Hayes & Kaun (1996) posit that the mapping of syllables to a discrete metrical grid is based on the phonetic durations of those syllables: 'Reflect the natural phonetic durations of syllables in the number of metrical beats they receive' (1996: 260). As discussed, most of their evidence for weight as a continuous as opposed to categorical variable derives from the correlation of grid space with final lengthening, evidence that is subject to another possible analysis: independent of syllable weight, textsetting is known to match linguistic constituency (Oehrle 1989; Halle 2004; above, figure 2). The larger gaps after higher-level constituents (sites of greater final lengthening) might be due to the parallelism of linguistic constituency and musical phrasing (Halle 2004).

That said, the present study supports Hayes & Kaun's (1996) conclusion through several tests that avoid this confound by controlling for constituency, taking only the

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

stressed initials of disyllables. Moreover, I expand the repertoire of phonetic factors known to influence textsetting, adding vowel height,[21] coda size, leading onset size, trailing onset size, and vowel lengthening after a null onset. Furthermore, onset effects reveal that weight is based not on the rime, syllable, or vowel-to-vowel interval, but rather on the p-center interval. P-centers approximate the beginnings of vowels, but are perturbed by properties of onsets and other factors. P-center intervals are only a slight modification of vowel-to-vowel intervals, but one that permits the incorporation of leading onset effects of the type documented in section 4. The role of p-centers as points of alignment between syllables and beats has long been recognized. Given this role, it is perhaps not surprising that p-centers should also serve to delimit the domain of weight for textsetting, as proposed here.

Finally, while textsetting is based in part on natural duration, I argue that there remains a role for categorical weight. In particular, syllables/intervals are more polarized in textsetting than durational differences alone would predict. A grammatical model with both categorical and gradient weight as factors significantly outperforms one with gradient weight alone.[22] Indeed, Ryan (2011a: 440–6) postulates the same need for hybridity in metrics.[23] For example, in the Finnish Kalevala epic, heavy and light syllables behave as nearly categorically distinct; however, a cline of weight is evident within heavy syllables. Categoricity is also supported by several textsetting traditions in which syllable weight translates more or less directly to metrical grid positions (e.g. Ancient Greek, Hausa and others cited in the introduction). I therefore maintain that weight mapping for textsetting has both categorical and gradient aspects, as can be implemented by the combination of categorical correspondence (say, Heavy-is-Long: 'a heavy interval must be allocated at least a quarter measure') and gradient correspondence (say, Heavier-is-Long: 'for every interval mapped onto an eighth measure, assign a penalty equivalent to that interval's natural duration'). While these constraints are merely sketches inviting further development, this combined-constraint approach is essentially what was implemented by the logistic models in section 5.

Another area of ongoing research concerns the degree of phonetic detail to which the textsetting grammar has access. This article supports the position that at least some phonetic detail is available (e.g. the intrinsic durations of vowels), building on phonetic effects in textsetting posited by other studies (e.g. polysyllabic shortening in Hayes & Kaun 1996, prevoiceless compression in Gilroy 2021). But numerous aspects of prosody and segmental timing are yet to be probed. More generally, what is at issue is how textsetting grammars operationalize natural duration. One simple hypothesis is that the grammar has access to all systematic phonetic detail in the spoken language.

[21] Hayes & Kaun (1996: 300) test for an effect of vowel height, but the result is not significant.

[22] Strictly speaking, I have shown only that a linear effect of duration underperforms a hybrid model. It may also be possible to incorporate the effect of categoricity directly into the gradient factor by redefining that factor to refer to a warped perceptual space. Either way, however, the conclusion is the same: categoricity affects weight mapping for textsetting; raw duration is insufficient.

[23] See also Ryan (2014: 330–1, 2016: 728) on stress.

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

However, it may turn out that some aspects of spoken language timing are not taken into account when allocating discrete grid space to syllables. For example, timing in spoken language is sensitive to factors such as word frequency, contextual predictability and speech rate, to name a few; textsetting may or may not respond to such factors. Weight mapping is rich, but how rich?

Author's address:

Department of Linguistics
Harvard University
317 Boylston Hall
Cambridge, MA 02138
USA
kevinryan@fas.harvard.edu

References

Aroui, Jean-Louis & Andy Arleo (eds.). 2009. *Towards a typology of poetic forms*. Amsterdam: John Benjamins.

Baayen, Rolf Harald, Richard Piepenbrock & Léon Gulikers. 1993. *The CELEX lexical database* [CD-ROM]. Philadelphia, PA: Linguistics Data Consortium, University of Pennsylvania.

Banti, Giorgio & Francesco Giannattasio. 1996. Music and metre in Somali poetry. In Richard J. Hayward & Ioan M. Lewis (eds.), *Voice and power: The culture of language in North-East Africa*, 83–127. London: School of Oriental and African Studies.

Barbosa, Plínio, Pablo Arantes, Alexsandro R. Meireles & Jussara M. Vieira. 2005. Abstractness in speech-metronome synchronisation: p-centres as cyclic attractors. *Interspeech* 2005, 1441–4.

Bartlett, Susan, Grzegorz Kondrak & Colin Cherry. 2009. On the syllabification of phonemes. *Human language technologies: The 2009 annual conference of the North American chapter of the ACL*, 308–16. Boulder, CO: Association for Computational Linguistics.

Bravi, Paolo. 2016. Sung syllables: Structure and boundaries of the metrical unit in sung verse. In Domenico Russo (ed.), *The notion of the syllable across history, theories and analysis*, 436–52. Newcastle upon Tyne: Cambridge Scholars Publishing.

Clements, G. N. & Susan Hertz. 1996. An integrated model of phonetic representation in grammar. Lisa Lavoie & William Ham (eds.), *Working papers of the Cornell Phonetics Laboratory*, vol. 11, 43–116. Ithaca, NY: CLC Publications.

Cooper, André Maurice, D. H. Whalen & Carol Ann Fowler. 1986. P-centers are unaffected by phonetic categorization. *Perception & Psychophysics* 39, 187–96.

Dell, François. 1989. Concordances rythmiques entre la musique et les paroles dans le chant: l'accent et l'*e* muet dans la chanson française. In Marc Dominicy (ed.), *Le souci des apparences*, 121–36. Brussels: Éditions de l'Université de Bruxelles.

Dell, François. 2011. Singing in Tashlhiyt Berber, a language that allows vowel-less syllables. In Charles Cairns & Eric Raimy (eds.), *Handbook of the syllable*, 173–93. Leiden: Brill.

Dell, François & Mohamed Elmedlaoui. 2008. *Poetic meter and musical form in Tashlhiyt Berber*. Cologne: Köppe.

Dell, François & Mohamed Elmedlaoui. 2017. Syllabic weight in Tashlhiyt Berber. In Paul Newman (ed.), *Syllable weight in African languages*, 83–96. Amsterdam: John Benjamins.

Dell, François & John Halle. 2009. Comparing musical textsetting in French and in English songs. In Aroui & Arleo (eds.), 63–78.

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

Devine, Andrew M. & Laurence Stephens. 1994. *The prosody of Greek speech*. Oxford: Oxford University Press.

Fant, Gunnar & Anita Kruckenberg. 1989. Preliminaries to the study of Swedish prose reading and reading style. *STL-QPSR* 2, 1–83.

Farnetani, Edda & Shiro Kori. 1986. Effects of syllable and word structure on segmental durations in spoken Italian. *Speech Communication* 5, 17–24.

Fowler, Carol Ann. 1983. Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General* 112, 386–412.

Fox, Robert A. & Ilse Lehiste. 1987a. The effect of unstressed affixes on stress-beat location in speech production and perception. *Peceptual and Motor Skills* 65, 35–44.

Fox, Robert A. & Ilse Lehiste. 1987b. The effect of vowel quality variations on stress-beat location. *Journal of Phonetics* 15, 1–13.

Franich, Kathryn. 2018. Tonal and morphophonological effects on the location of perceptual centers (p-centers): Evidence from a Bantu language. *Journal of Phonetics* 67, 21–33.

Gilroy, Nicole. 2021. English vowel duration in textsetting. Master's thesis, Carleton University.

Girardi, Elena & Ingo Plag. 2019. Metrical mapping in text-setting: Empirical analysis and grammatical implementation. MS, Heinrich-Heine-Universität Düsseldorf.

Gordon, Matthew. 2005. A perceptually-driven account of onset-sensitive stress. *Natural Language and Linguistic Theory* 23, 595–653.

Gussenhoven, Carlos & Aoju Chen (eds.). 2020. *The Oxford handbook of language prosody*. Oxford: Oxford University Press.

Halle, John. 2004. Constituency matching in metrical texts. MS, submitted to the *Proceedings of the Conference Words and Music*, University of Missouri, Columbia.

Halle, John & Fred Lerdahl. 1993. A generative textsetting model. *Current Musicology* 55, 3–23.

Halle, Morris & Jean-Roger Vergnaud. 1987. *An essay on stress*. Cambridge, MA: MIT Press.

Hayes, Bruce. 2009a. Faithfulness and componentiality in metrics. In Sharon Inkelas & Kristin Hanson (eds.), *The nature of the word: Essays in honor of Paul Kiparsky*, 113–48. Cambridge, MA: MIT Press.

Hayes, Bruce. 2009b. Textsetting as constraint conflict. In Aroui & Arleo (eds.), 43–62.

Hayes, Bruce & Abigail Kaun. 1996. The role of phonological phrasing in sung and chanted verse. *The Linguistic Review* 13, 243–303.

Hayes, Bruce & Margaret MacEachern. 1998. Quatrain form in English folk verse. *Language* 74, 473–507.

Hayes, Bruce & Russell G. Schuh. 2019. Metrical structure and sung rhythm of the Hausa rajaz. *Language* 95, e253–e299.

Hill, Joseph David. 2008. Syllabification and syllable weight in Ancient Greek songs. Master's thesis, Massachusetts Institute of Technology.

Hirsch, Aron. 2014. What is the domain for weight computation: The syllable or the interval? In John Kingston, Claire Moore-Cantwell, Joe Pater & Robert Staubs (eds.), *Proceedings of the 2013 Meeting on Phonology*. Washington, DC: Linguistic Society of America.

Hyman, Larry. 1985. *A theory of phonological weight*. Dordrecht: Foris.

Jacewicz, Ewa, Robert A. Fox & Joseph Salmons. 2007. Vowel duration in three American English dialects. *American Speech* 82, 367–85.

Jackendoff, Ray & Fred Lerdahl. 2006. The capacity for music: What is it, and what's special about it? *Cognition* 100, 33–72.

Kahn, Daniel. 1976. Syllable-based generalizations in English phonology. PhD dissertation, Massachusetts Institute of Technology.

Katz, Jonah. 2010. Compression effects, perceptual asymmetries, and the grammar of timing. PhD dissertation, Massachusetts Institute of Technology.

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

Keshet, Ezra. 2006. Relatively optimal text-setting. MS, Massachusetts Institute of Technology.

Kiparsky, Paul. 2006. A modular metrics for folk verse. In B. Elan Dresher & Nila Friedberg (eds.), *Formal approaches to poetry: Recent developments in metrics* (Phonology and Phonetics 11), 7–49. Berlin and New York: Mouton de Gruyter.

Kiparsky, Paul. 2020. Stress, meter, and text-setting. In Gussenhoven & Chen (eds.), 657–75.

Ladd, D. Robert & James Kirby. 2020. Tone–melody matching in tone language singing. In Gussenhoven & Chen (eds.), 676–87.

Lerdahl, Fred & Ray Jackendoff. 1983. *A generative theory of tonal music*. Cambridge, MA: MIT Press.

List, George. 1974. The reliability of transcription. *Ethnomusicology* 18, 353–77.

Lunden, Anya. 2017. Syllable weight and duration: A rhyme/intervals comparison. *Proceedings of the Linguistic Society of America* 2, 1–12.

Marcus, Stephen Michael. 1981. Acoustic determinants of perceptual center (P-center) location. *Perception & Psychophysics* 30, 247–56.

McCrary, Kristie Marie. 2004. Reassessing the role of the syllable in Italian phonology: An experimental study of consonant cluster syllabification, definite article allomorphy and segment duration. PhD dissertation, University of California, Los Angeles.

McPherson, Laura. 2018. The talking balafon of the Sambla: Grammatical principles and documentary implications. *Anthropological Linguistics* 60, 255–94.

McPherson, Laura. 2019. Musical adaptation as phonological evidence: Case studies from textsetting, rhyme, and musical surrogates. *Language and Linguistics Compass* e12359.

McPherson, Laura. 2021. Categoricity, variation, and gradience in Sambla balafon segmental encoding. *Frontiers in Communication* 6, 652635.

McPherson, Laura & Kevin M. Ryan. 2018. Tone-tune association in Tommo So (Dogon) folk songs. *Language* 94, 119–56.

Meseguer-Brocal, Gabriel, Alice Cohen-Hadria & Geoffroy Peeters. 2018. DALI: A large dataset of synchronized audio, lyrics and notes, automatically created using teacher–student machine learning paradigm. *19th International Society for Music Information Retrieval Conference*. Paris: International Society for Music Information Retrieval.

Meyer, Julien, Laure Dentel & Frank Seifart. 2012. A methodology for the study of rhythm in drummed forms of languages: Application to Bora Manguaré of Amazon. In *Proceedings of Interspeech* 12*: Annual Conference of the International Speech Communication Association*, 687–90.

Moore-Cantwell, Claire. 2021. Weight and final vowels in the English stress system. *Phonology* 37, 657–95.

Morton, John, Steve Marcus & Clive Frankish. 1976. Perceptual centers (P-centers). *Psychological Review* 83, 405–8.

Nwe, Tin Lay, Minghui Dong, Paul Chan, Xi Wang, Bin Ma & Haizhou Li. 2010. Voice conversion: From spoken vowels to singing vowels. *2010 IEEE International Conference on Multimedia and Expo*, 1421–6.

Oehrle, Richard. 1989. Temporal structures in verse design. In Paul Kiparsky & Gilbert Youmans (eds.), *Rhythm and meter*, 87–119. San Diego, CA: Academic Press.

Olejarczuk, Paul & Vsevolod Kapatsinski. 2018. The metrical parse is guided by gradient phonotactics. *Phonology* 35, 367–405.

Palmer, Caroline & Michael H. Kelly. 1992. Linguistic prosody and musical meter in song. *Journal of Memory and Language* 31, 525–42.

Patel, Aniruddh D. 2008. *Music, language and the brain*. Oxford: Oxford University Press.

Patel, Aniruddh D., Anders Löfqvist & Walter Naito. 1999. The acoustics and kinematics of regularly timed speech: A database and method for the study of the P-center problem. In *Proceedings of the XIVth International Congress of Phonetic Sciences*, 405–8.

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

Pitt, Mark, Laura Dilley, Keith Johnson, Scott Kiesling, William Raymond, Elizabeth Hume & Eric Fosler-Lussier. 2007. *Buckeye Corpus of Conversational Speech* (2nd release). Columbus, OH: Department of Psychology, Ohio State University. www.buckeyecorpus.osu.edu

Port, Robert. 2007. The problem of speech patterns in time. In M. Gareth Gaskell (ed.), *The Oxford handbook of psycholinguistics*, 503–14. Oxford: Oxford University Press.

Proto, Teresa. 2015. Prosody, melody and rhythm in vocal music: The problem of textsetting in a linguistic perspective. *Linguistics in the Netherlands* 32, 116–29.

Proto, Teresa & François Dell. 2013. The structure of metrical patterns in tunes and in literary verse: Evidence from discrepancies between musical and linguistic rhythm in Italian songs. *Probus* 25, 105–38.

Ross, Jaan & Ilse Lehiste. 2011. *The temporal structure of Estonian runic songs*. Berlin: Mouton de Gruyter.

Ryan, Kevin M. 2011a. Gradient syllable weight and weight universals in quantitative metrics. *Phonology* 28, 413–54.

Ryan, Kevin M. 2011b. Gradient weight in phonology. PhD dissertation, University of California, Los Angeles.

Ryan, Kevin M. 2014. Onsets contribute to syllable weight: Statistical evidence from stress and meter. *Language* 90, 309–41.

Ryan, Kevin M. 2016. Phonological weight. *Language and Linguistics Compass* 10, 720–33.

Ryan, Kevin M. 2019. *Prosodic weight: Categories and continua*. Oxford: Oxford University Press.

San, Nay & Myfany Turpin. 2021. Text-setting in Kaytetye. In *Proceedings of the 2020 Annual Meeting on Phonology* 9, 1–9.

van Santen, J. P. H. 1992. Contextual effects on vowel duration. *Speech Communication* 11, 513–46.

Schellenberg, Murray. 2012. Does language determine music in tone languages? *Ethnomusicology* 56, 266–78.

Schuh, Russell G. 2001. The metrics of a Bole song style, Kona. MS, University of California, Los Angeles.

Schuh, Russell G. 2011. Quantitative metrics in Chadic and other Afroasiatic languages. *Brill's Annual of Afroasiatic Languages and Linguistics* 3, 202–35.

Seifart, Frank, Julien Meyer, Sven Grawunder & Laure Dentel. 2018. Reducing language to rhythm: Amazonian Bora drummed language exploits speech rhythm for long-distance communication. *Royal Society Open Science* 5, 170354.

Selkirk, Elisabeth O. 1982. The syllable. In Harry van der Hulst & Norval Smith (eds.), *The structure of phonological representations*, part II, 337–84. Dordrecht: Foris.

Simpson, Andrew J. R., Gerard Roma & Mark D. Plumbley. 2015. Deep karaoke: Extracting vocals from musical mixtures using a convolutional deep neural network. In E. Vincent, A. Yeredor, Z. Koldovský & P. Tichavský (eds.), *Latent variable analysis and signal separation*, 429–36. Cham: Springer.

Starr, Rebecca L. & Stephanie S. Shih. 2017. The syllable as a prosodic unit in Japanese lexical strata: Evidence from text-setting. *Glossa* 2, 1–34.

Steriade, Donca. 2008. Resyllabification in the quantitative meters of Ancient Greek: Evidence for an Interval Theory of Weight. MS, Massachusetts Institute of Technology.

Steriade, Donca. 2012. Invervals vs. syllables as units of linguistic rhythm. Handouts, EALING, Paris.

Steriade, Donca. 2019. CiV lengthening and the weight of CV. In Margit Bowler, Philip T. Duncan, Travis Major & Harold Torrence (eds.), *Schuhschrift: Papers in honor of Russell Schuh*, 161–77. eScholarship Publishing.

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press

Šturm, Pavel & Jan Volín. 2016. P-centres in natural disyllabic Czech words in a large-scale speech–metronome synchronization experiment. *Journal of Phonetics* 55, 38–52.

Tan, Ivan, Ethan Lustig & David Temperley. 2019. Anticipatory syncopation in rock: A corpus study. *Music Perception* 36, 353–70.

Temperley, Nicholas & David Temperley. 2013. Stress-meter alignment in French vocal music. *Journal of the Acoustical Society of America* 134, 520–7.

Umeda, Noriko. 1975. Vowel duration in American English. *Journal of the Acoustical Society of America* 58, 434–45.

Villepastour, Amanda. 2014. Talking tones and singing speech among the Yorùbá of Southwest Nigeria. In Gerda Lechleitner & Christian Liebl (eds.), *Jahrbuch des Phonogrammarchivs der Österreichischen Akademie der Wissenschaften*, vol. 4, 29–46. Göttingen: Cuvillier.

Villing, Rudi C. 2010. Hearing the moment: Measures and models of the perceptual centre. PhD dissertation, National University of Ireland Maynooth.

Wee, Lian Hee. 2007. Unraveling the relation between Mandarin tones and musical melody. *Journal of Chinese Linguistics* 35, 128–43.

West, Martin L. 1992. *Ancient Greek music*. Oxford: Clarendon Press.

https://doi.org/10.1017/S1360674322000156 Published online by Cambridge University Press