

1 TITLE: Temporal (in)stability in English monosyllabic and disyl-
2 labic words: Insights on the effect of voicing on vowel duration

3 AUTHOR: Stefano Coretta

4 AFFILIATION: Linguistics and English Language, University of
5 Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom

6 EMAIL ADDRESS: stefano.coretta@manchester.ac.uk

Abstract

English is one in the wide range of languages in which the duration of vowels is modulated by the voicing of the following consonant: Vowels are shorter when followed by voiceless stops, and longer when followed by voiced stops. The so-called voicing effect has been attributed to a variety of mechanisms. Temporal compensation between the duration of the vowel and the following stop closure is one of these mechanisms. Based on acoustic data from Italian and Polish disyllabic words, the compensatory mechanism has been proposed to be a consequence of the temporal stability of the interval between the consonant releases flanking the vowel. The timing of the VC boundary within this interval determines the respective durations of the vowel and the stop closure. In this paper, it is shown that the duration of the release-to-release interval is not affected by the voicing of the second consonant in English disyllabic words, but that it is in monosyllabic words. It is argued that the stability of the interval can be derived from the isochronous phasing of the vocalic gestures in the VCV sequence of disyllabic words. The absence of the temporal anchor of a second vowel in monosyllabic words, on the other hand, allows the vocalic and the consonant gesture durations to be modified independently. Other aspects of production and perception behind the voicing effect can coexist with a temporal compensation mechanism and cannot be excluded.

Keywords: voicing effect, vowel duration, compensatory adjustment, English

1 Introduction

A well-known cross-linguistic tendency is that vowels have shorter durations when followed by voiceless stops and longer durations when followed by voiced stops. This so-called ‘voicing effect’ has been long documented in a wide range of languages across different linguistic families (Maddieson & Gandour 1976; Beguš 2017). Several hypotheses have been proposed as to the origin of this phenomenon, from articulatory mechanisms to perceptual biases; however, no one particular account has gained universal support.

One such hypothesis, the compensatory temporal adjustment account, states that the voicing effect involves a compensatory mechanism between vowel and consonant closure duration. Vowels are shorter when followed by voiceless stops because the latter have longer closure durations, and, vice versa, vowels are longer before voiced stops because the latter have shorter closure durations. However, the compensatory account fails to clearly identify a speech interval within which compensation is implemented. Both the syllable (Lindblom 1967; Farnetani & Kori 1986) and the word (Slis & Cohen 1969a,b; Lehiste 1970a,b) have been proposed

45 as such intervals, but these have been subsequently criticised on empirical and
46 logical grounds (Chen 1970; Jacewicz et al. 2009; Maddieson & Gandour 1976;
47 Coretta 2019b).

48 In an exploratory study of acoustic durations in Italian and Polish trochaic
49 CVCV words, Coretta (2019b) finds that the duration of the interval between the
50 the release of C1 and the release of C2 (the release-to-release interval) is not af-
51 fected by the voicing status of the second consonant. The duration of the release-
52 to-release interval in words where the second consonant is voiceless (like /pata/) is
53 not significantly different from that in words where the second consonant is voiced
54 (for example, /pada/).¹ The temporal stability of the release-to-release interval is
55 compatible with a compensatory temporal adjustment account of the voicing effect
56 (Lindblom 1967; Slis & Cohen 1969a,b; Lehiste 1970a,b), and it offers a resolution
57 to the drawbacks of previous versions of the account.

58 Given the temporal stability of the release-to-release interval, the timing of the
59 acoustic vowel/consonant (VC) boundary (corresponding to the acoustic vowel
60 offset and the acoustic consonant closure onset) within that interval determines
61 the respective durations of vowel and consonant closure. Since the VC boundary
62 in voiceless stops is timed earlier within the release-to-release relative to voiced
63 stops, the vowel is shorter and closures is longer when the post-vocalic stop is
64 voiceless than when it is voiced. This interpretation agrees with known differences
65 of closure durations in voiceless vs. voiced stops (Lisker 1957; Summers 1987;
66 Davis & Summers 1989; de Jong 1991), namely that voiceless closures are longer
67 than voiced ones. Thus, a possible diachronic pathway to the voicing effect is
68 one in which vowel and closure duration differences emerge from changes in the
69 timing of the VC boundary within the release-to-release interval.

70 Note that the release-to-release interval in itself does not have a special status
71 in the account put forward here. The proposed account of compensatory temporal
72 adjustment can be understood in relation to the *acoustic* duration of vowels, hence
73 the scope of compensation can (but need not) be defined in terms of acoustic inter-
74 vals. As reviewed above, the interval found to be temporally stable across voicing
75 contexts in disyllabic words in Italian and Polish is the release-to-release interval
76 (Coretta 2019b). However, it is desirable to derive the isochrony of this acoustic
77 interval from properties of articulatory coordination. A tentative account of the
78 underlying gestural coordination from which the release-to-release isochrony can
79 be derived is offered here.

¹Note that the release-to-release interval is not equivalent to the VC interval, since the latter does not include the VOT from the consonant preceding the vowel. See Coretta (2019b) for further details.

1.1 A gestural account of the voicing effect

According to Öhman (1966, 1967), the speech stream is composed by a series of continuous vocalic gestures interrupted by gestures of oral constriction (consonants). Fowler (1983) further proposes that the vocalic gestures of a VCV sequence are characterised by a cyclic pattern of production, so that the temporal distance between the two vowels is constant, independent of the nature of the intervening consonant. While the temporal distance of the V-to-V interval is modulated by the number of intervening consonants (Zmarich et al. 2011; Zeroual et al. 2015), the distance can still be expected to be stable within the context of disyllabic words with a single intervocalic consonant that alternates in voicing.

The task-dynamic model (Saltzman et al. 2008) of Articulatory Phonology (Ohala et al. 1986; Browman & Goldstein 1988, 1992), based on the coupled oscillators model (O'Dell & Nieminen 2008), states that any two gestures can be implemented according to two modes. Either they are initiated in synchrony or they are implemented sequentially. These modes of gestural phasing (in-phase and anti-phase) can account for a variety of patterns of articulatory timing. Onset consonants are generally produced in-phase with the following vowel, meaning that the vocalic and consonantal gestures are initiated together. This same mechanism also gives rise to the so-called C-centre effects observed with onsets, by which the acoustic duration of a vowel depends on the number of onset consonants (Browman & Goldstein 1988; Marin & Pouplier 2010; Hermes et al. 2013; Marin & Pouplier 2014). On the other hand, coda consonants tend to be produced anti-phase relative to the preceding vowel and to each other, meaning that they are articulated in sequence. The duration of vowels has been found to be insensitive to the number of coda consonants following them (Marin & Pouplier 2010, 2014).

Further evidence for a vowel-based rhythmic gestural implementation comes from work by Farnetani & Kori (1986) and Celata & Mairano (2014). These studies investigate the relation between vowel duration and syllable structure in Italian. In the first study, it was found that vowels followed by a singleton stop (for example in /la.da/) are longer than vowels followed by a tautosyllabic cluster (/la.dra/). This pattern can easily be derived from a scenario in which the distance between the vowels is the same in the two contexts (/la.da/ and /la.dra/), and the onset consonants follow a C-centre alignment. The onset of the first consonant in the /dr/ cluster is shifted towards the vowel, making the latter shorter. Celata & Mairano (2014) also show that the duration of the consonant/consonant cluster is negatively correlated with the duration of the preceding vowel (although the magnitude of the correlation is low to moderate).

Under the gestural account proposed in this paper, the combined action of the isochrony of the vowel-to-vowel interval and the in-phase alignment of the onset consonant is responsible for the isochrony of the release-to-release interval in

120 CVCV words. The first consonant and vowel are produced in-phase with each
121 other, and these are sequentially followed by the second consonant and vowel,
122 again produced in-phase with each other. Then, the differential duration found in
123 the voicing effect would be a consequence of the different velocity of the closing
124 gesture in voiceless vs. voiced consonants.

125 Summers (1987) shows that the closing gesture of voiceless stops has greater
126 velocity than that of voiced stops. Assuming that the closing gesture of both voice-
127 less and voiced stops is initiated in synchrony with that of the following vowel (as
128 per the in-phase alignment), full oral closure is achieved earlier in voiceless stops
129 relative to its timing in voiced stops. The result is, everything else being equal, a
130 shorter vowel and a longer (full) closure in voiceless stops, and a longer vowel and
131 a shorter closure in voiced stops. Warren & Hay (2006) offer some evidence from
132 lip data, where the lip closing gestures (together with the jaw gesture) accounts for
133 about 80% of the vowel duration difference.

134 A similar chain of reasoning applies to the context of coda consonants. If coda
135 consonants are produced anti-phase with the preceding vowel and the closing ges-
136 ture of the consonant starts at a specific time after the production of the vowel
137 independent of consonant voicing, then the greater velocity of the closing gesture
138 in voiceless stops would result in a shorter vowel and a longer closure relative
139 to voiced stops. However, the articulatory data in de Jong (1991) suggests that
140 the timing of the onset of the stop closing gesture differs depending on voicing in
141 English monosyllabic words. Section 4.3 will discuss a possible solution based
142 on perceptual mechanisms that could reconcile the articulatory predictions put for-
143 ward here and the empirical data.

144 1.2 The voicing effect in English

145 English is one of the most investigated languages in relation to the voicing effect
146 (Meyer 1904; Heffner 1937; House & Fairbanks 1953; Belasco 1953; Peterson &
147 Lehiste 1960; Halle & Stevens 1967; Chen 1970; Klatt 1973; Lisker 1974; Laeuffer
148 1992; Fowler 1992; Hussein 1994; Lampp & Reklis 2004; Warren & Jacks 2005;
149 Durvasula & Luo 2012; Ko 2018). English is also the language in which the voic-
150 ing effect reportedly has the greatest magnitude relative to that of other languages.
151 This special status of English is traditionally attributed to the phonologisation of
152 the voicing effect in this language (Sharf 1964; de Jong 2004). Vowel duration
153 and the vowel-to-consonant duration ratio are considered to be among the most
154 stable cues to consonantal voicing (Peterson & Lehiste 1960; Raphael 1972; Port
155 & Dalby 1982). Kluender et al. (1988) proposed that the difference in vowel du-
156 ration before voiceless vs. voiced stops could have been enhanced and exploited
157 to cue the voicing contrast. This could explain the greater effect of English com-
158 pared for example to the effect in Italian, in which voicing is most robustly cued

by vocal fold vibration during closure (Pape & Jesus 2014).

As it is the case, previous studies on English have reported a difference in vowel duration before voiceless vs. voiced stops which ranges between 20 and 150 ms, while the values for the effect in Italian are lower, between 15 and 25 ms (Magno Caldognetto et al. 1979; Farnetani & Kori 1986; Esposito 2002; Coretta 2019b). A Bayesian meta-analysis of the voicing effect based on estimates obtained from 15 studies suggests that the effect of voicing in English monosyllabic words ranges between 55 and 95 ms at 95% probability, with a meta-analytical mean of 75 ms (see Supplement A). On the other hand, the meta-analytical estimate of the voicing effect for disyllabic words is lower, at about 25 ms (around 50 ms less than in monosyllabic words). This estimate is closer to the effect sizes reported for Italian. Note, however, that the Italian values refer to the effect as observed in disyllabic words.

Nevertheless, it is possible that the alleged differences in magnitude between English and other languages are a product of the different contexts under examination (Laeuffer 1992). Ko (2018), in a more recent investigation of the voicing effect in English monosyllabic words, finds a substantially lower difference in vowel duration (35 ms). The Bayesian meta-analysis (Supplement A) further suggests a potential for publication bias towards bigger effects, which means that the meta-analytical estimate (75 ms) could be an overestimation. Finally, the surveyed studies have a very low number of participants (mean = 3.4, SD = 2.5), which can lead to so-called Type M errors (estimate magnitude errors) and overestimation of the effect (Kirby & Sonderegger 2018; Roettger 2019). In sum, it is generally assumed that the voicing-driven differences in vowel duration are greater in English than in other languages, although the empirical foundation of this conception is not entirely straightforward. Although not the focus of this study, arguments based on differences in effect size will become relevant when discussing the results.

1.3 Research hypotheses

One of the aims of this study is to test whether the same temporal stability observed for the release-to-release interval in Italian and Polish disyllabic words can also be observed in English. English makes a good object of enquiry since much work on the voicing effect has been done on this language, and the phonological structure of the language allows us to directly compare onset and coda stops in disyllabic and monosyllabic words respectively. While the temporal stability of the release-to-release interval is expected in English disyllabic words, monosyllabic words are predicted not to show such stability. As discussed above, an essential component of the release-to-release temporal stability in disyllabic words is the presence of a direct relation between the two vowels in these type of words. Since monosyllabic words don't have a second vowel, there is no direct vowel-to-vowel relation to

198 derive the release-to-release stability from.

199 Furthermore, Jacewicz et al. (2009) report that, in American English, monosyl-
200 labic words are longer when the second consonant is voiced. Based on this find-
201 ing, it is expected that the release-to-release duration should be longer when C2 is
202 voiced. Jacewicz et al. (2009) attribute the difference in monosyllabic word dura-
203 tion to the difference in vowel duration before voiceless vs. voiced stops. Thus, we
204 can expect the magnitude of the difference in release-to-release duration in mono-
205 syllabic words to be close to the difference in vowel duration. This hypothesis also
206 fits with the reported greater effect of voicing on vowel duration in monosyllabic
207 than disyllabic words (with the caveats discussed in Section 1.2).

208 The data in Coretta (2019b) suggests that the intrinsic duration of vowels and
209 consonants can contribute to the duration of the release-to-release interval. In par-
210 ticular, release-to-release intervals containing a high vowel have shorter durations
211 than those with a low vowel. This is not surprising, given the well-known ten-
212 dency of high vowels to be shorter than low vowels (Hertrich & Ackermann 1997;
213 Esposito 2002; Mortensen & Tøndering 2013; Toivonen et al. 2015; Kawahara
214 et al. 2017). As for the consonantal place of articulation, the release-to-release is
215 shorter in Italian and Polish when the second consonant is velar compared to when
216 it is coronal. This could be a consequence of the fact that the closure of velar stops
217 is shorter than that of other stops. For example, Sharf's (1962) data on closure
218 duration in English suggests that the closure of labial stops (60-90 ms) is about
219 10 ms longer than that of velar stops (55-75 ms). It can be expected that release-
220 to-release intervals with a velar stop in English will be about 10 ms shorter than
221 intervals with a labial stop.

222 Another set of objectives concerns the effect of voicing on vowel and closure
223 durations. A conceptual replication of previous studies' effect sizes is sought, with
224 special attention to differences between monosyllabic and disyllabic words. Only
225 a few studies directly compare the effect on vowel durations in different syllabic
226 positions (for example, Sharf (1962) and Klatt (1973)). The reported effects are in
227 the range of 50-55 ms in word-final (closed-syllable) position and 20-25 in word-
228 medial (open-syllable) position. The Bayesian meta-analysis of the voicing effect
229 indicates a mean difference of 50 ms (75 ms in word-final position vs. 25 ms word-
230 medially). As for the duration of stop closure, the data in Sharf (1962) and Luce
231 & Charles-Luce (1985) indicates that voiced closures are 10-13 ms shorter than
232 voiceless closures.

233 While directly investigating the statistical relationship between vowel and clo-
234 sure durations might seem more appropriate for assessing whether a compensation
235 exists between the two, this was not done for three reasons. First, as discussed by
236 Ohala & Lyberg (1976) and Beguš (2017), interpreting the correlations between
237 two temporal intervals that share a boundary is problematic, unless one uses two
238 independent sources for the measurements of vowel and closure durations, such

as multiple annotators (Beguš 2017) or articulatory data from different articulators (de Jong 1991). Neither of these options were available for this study due to time and resources constraints. Second, as discussed in Coretta (2019b), less is known about the interplay between segment durations, speech rate, and the vowel/closure relation, making interpretation of data on correlation difficult in most cases. A third reason for not correlating vowel and closure durations is that the main focus of this study was to assess word-level gestural phasing via acoustic measures. Future work will have to be carried out in which the vowel/closure correlation is investigated using robust and unambiguous methods.

To summarise, the following research questions and respective hypotheses can be formulated:

1. Is the duration of the interval between two consecutive stop releases (the release-to-release interval) in monosyllabic and disyllabic words affected by the voicing of C2 in English?
 - H1a: The duration of the release-to-release interval is not affected by C2 voicing in disyllabic words.
 - H1b: The release-to-release interval is longer in monosyllabic words with a voiced C2 than in monosyllabic words with a voiceless C2.
2. Is the duration of the release-to-release interval affected by (a) the number of syllables of the word, (b) the quality of V1, and (c) the place of C2?
 - H2a: The release-to-release interval is longer in monosyllabic than in disyllabic words.
 - H2b: The duration of the release-to-release interval decreases according to the hierarchy /ɑ:/, /ɜ:/, /i:/.
 - H2c: The release-to-release interval is longer when C2 is labial.
3. What is the estimated difference in the effect of voicing on vowel and stop closure duration between monosyllabic and disyllabic words?
 - H3: The effect of voicing on vowel duration is greater in monosyllabic than in disyllabic words (no specific hypothesis for the effect of number of syllables on the effect of voicing on closure duration).

2 Methods

The following subsections describe the experimental and statistical methods of this study. The research design and data analyses were pre-registered on the Open Science Framework prior to data collection.² The research compendium of this paper

²The pre-registration can be found at this temporary link for peer-review: https://osf.io/hwr94/?view_only=d994915422144efaae4a5915237cb386.

273 with data (Coretta 2019a) and analysis scripts is also available on the Open Science
274 Framework.³ Choices on experimental design and analysis were made within the
275 Bayesian framework of statistical inference (see Section 2.1 and Section 2.7 for
276 details).

277 **2.1 Sample size and stopping rule**

278 Sample size and a stopping rule were decided prior to data collection with a
279 Bayesian method of sample determination based on the Region Of Practical
280 Equivalence (ROPE, Kruschke 2015; Vasishth et al. 2018b). A ‘no-effect’ region
281 of values around 0 is first identified. This null region (the ROPE) can be thought
282 of as a Bayesian 95% credible interval of a distribution, the values within which
283 can be interpreted as a negligible or null effect. For this study, a ROPE between
284 -10 and +10 ms has been chosen. The width of 20 ms is based on the estimates
285 of the just noticeable difference in Huggins (1972) and Nooteboom & Doodeman
286 (1980). Differences in release-to-release durations below 10 ms (either positive
287 or negative) will be interpreted as compatible with a null effect.

288 Once a ROPE width is set, the goal is to collect data during sequential testing
289 until the width of the 95% credible interval (CI) of the tested effect is equal to
290 or less than the ROPE width (in this study, 20 ms). In other words, the objective
291 is to reach estimate precision, rather than significance (as in frequentist null hy-
292 pothesis testing). Inference can then be made based on the credible interval of the
293 sought effect. When the precision goal is reached (the CI width is equal or lower
294 than the ROPE width), three possible scenarios can arise: (1) the CI of the effect
295 completely overlaps with the ROPE around 0, in which case the data supports a
296 practically equivalent null effect; (2) the CI of the effect completely lies outside
297 the ROPE, which indicates that the data support the effect to be within that CI; (3)
298 the CI partially overlaps with the ROPE, in which case no decision can be made on
299 whether the data support one hypothesis over the other, although it still possible
300 to infer the sign of the effect (if the CI partially overlaps with the right side of the
301 ROPE without including 0, there is evidence for a positive effect, while if the CI
302 overlaps with the left side of the ROPE without including 0, there is evidence for
303 a negative effect).

304 An initial minimum of 20 participants was chosen for sequential testing. Due
305 to resource and time constraints specific to this particular study, a second condition
306 had to be included in the stopping rule such that data collection would have to
307 stop on 5 April 2019, independent of the ROPE condition.

³The analysis code can be found at this temporary link for peer-review: https://osf.io/32fst/?view_only=2dc237b60f4c4c77b6ec10300b9e528e. A public link will be generated in case of acceptance.

Table 1: Test $C_1\acute{V}_1C_2(VC)$ words.

teep	teepus	teek	teekus
teeb	teebus	teeg	teegus
terp	terpus	terk	terkus
terb	terbus	terg	tergus
tarp	tarpus	tark	tarkus
tarb	tarbus	targ	targus

2.2 Participants

The participants of this study were 15 native speakers of British English, who were born and raised in the Greater Manchester area. The speakers were all undergraduate students at the University of Manchester with no reported hearing or speaking disorders, and with normal or corrected to normal vision. The participants signed a written consent form and received £5 for participation.

2.3 Equipment

Audio recordings were obtained in a sound-attenuated room in the Phonetics Laboratory of the University of Manchester, with a Zoom H4n Pro recorder and a RØDE Lavalier microphone, at a sample rate of 44100 Hz (16-bit, downsampled to 22050 Hz for analysis). The Lavalier microphone was clipped on the participants clothes, about 20 cm from the mouth, displaced a few centimetres to one side.

2.4 Materials

The test words were $C_1\acute{V}_1C_2(VC)$ words, where $C_1 = /t/$, $V_1 = /i:, 3:, \alpha:/$, $C_2 = /p, b, k, g/$, and $(VC) = /əs/$. $/əs/$ was chosen for its lower parsability as a native suffix, in order to prevent morphological complexity in disyllabic words. This structure specification generates 24 test words, shown in Table 1. All of these are nonce words, with the exception of *turk* and *tarp*, and of *teek* via the homophone *teak*. Building stimuli from a structure template rather than from the lexicon ensures greater experimental and statistical control. Moreover, the use of nonce words removes or reduces confounds from some usage variables, like for example lexical frequency.⁴ Each word was embedded in the following frame sentences: *I'll say X this Thursday, You'll say X this Monday, She'll say X this Sunday, We'll say X this*

⁴The three real words in the materials have low lexical frequency (Zipf 1-7 log-frequency: *tarp* 2.23, *teak* 2.76, and *turk* 2.91) according to the SUBTLEX-UK corpus (Van Heuven et al. 2014).

332 *Friday, They'll say X this Tuesday*. Each word + frame combination was included
333 once in the stimuli list, so that each speaker read a total of 120 sentence stimuli
334 (24 words \times 5 frames). A total of 1800 observations were recorded (120 stimuli \times
335 15 speakers).

336 2.5 Procedure

337 The experimental procedure was first explained to the participants prior to record-
338 ing. The participants also familiarised themselves with the materials by reading
339 them aloud. They were instructed not to insert pauses anywhere within the sen-
340 tence stimuli and to keep a similar intonation contour for the total duration of the
341 experiment. They were also given the chance to take any number of breaks at any
342 point during recording. Misreadings or speech errors were corrected by asking the
343 participant to repeat the stimulus. The reading task took around 6 to 10 minutes,
344 while the total experiment session lasted about 25 minutes. Data collection started
345 on 19 February 2019 and ended on 5 April 2019.

346 2.6 Data processing and measurements

347 A forced-aligned transcription was obtained with the SPeech Phonetisation Align-
348 ment and Syllabification software (SPPAS, Bigi 2015). The automatic annotation
349 was corrected by the author according to the principles of phonetic segmentation
350 detailed in Machač & Skarnitzl (2009). A custom Praat script was written to au-
351 tomatically detect the burst onset of the consonants in the test words, using the
352 algorithm in Ananthapadmanabha et al. (2014). The output was checked and man-
353 ually corrected by the author when necessary.

354 The following measures were obtained via a custom Praat script:

- 355 • Duration of the release-to-release interval: from the release of C1 to the
356 release of C2.
- 357 • V1 duration: from appearance to disappearance of higher formant structure
358 in the spectrogram in correspondence of V1 (Machač & Skarnitzl 2009).
- 359 • C2 closure duration: from disappearance of higher formant structure in the
360 V1C2 sequence to the release of C2 (Machač & Skarnitzl 2009).
- 361 • Speech rate: calculated as the number of syllables per second (number of
362 syllables in the sentence divided by the sentence duration in seconds, Plug
363 & Smith 2018).

2.7 Statistical analysis

The choice of Bayesian over frequentist statistics stems from a recent discussion of the problems associated with the reliance of p -values in statistical inference (Wagenmakers 2007; Munafò et al. 2017; Kirby & Sonderegger 2018; Roettger 2019). Bayesian statistics also offers a straightforward framework for investigating the absence of differences across conditions (a ‘null effect’) based on the ROPE (Section 2.1), as it is in part the case in this study. Another favourable aspect of Bayesian methods is that more focus is given to the distributions of the enquired effects, rather than on point estimates (which are less informative when matters of statistical power are taken into consideration, see a discussion of Type S-M errors in Kirby & Sonderegger 2018) and an arbitrary significance cut-off point. Furthermore, Bayesian inference is centred around an incremental procedure of reallocation of credibility between natural states and on evidence based on observed data (Kruschke 2015), rather than on a series of hypothetical experimental replications (Wagenmakers 2007).⁵ For an introduction to Bayesian statistics in phonetics, see Vasishth et al. (2018a), and Nicenboim et al. (2018), while for a general introduction see Etz et al. (2018), McElreath (2015), Kruschke (2015), and references therein. While a thorough discussion of Bayesian methods would be beyond the scope of this paper, it is relevant to provide the less familiar reader with the basic tools for interpreting analyses and results.

Particular weight will be given to the estimated distributions of the sought effects in presenting the results of this study. The estimated distribution of an effect (or parameter) is the posterior distribution of that effect (or parameter). The posterior distribution is an approximation of the parameter distribution, and it takes into account the specified prior for that parameter, i.e. the theoretical probability of the parameter as known or derived by the researcher. The inclusion of priors in the analysis is at the heart of Bayesian modelling, which relies on prior knowledge for the estimation of parameter values. For each relevant term in the models, the 95% credible intervals (CI) should be taken as a summary of the posterior distribution, and inference should be based on the posterior rather than on the point estimate (the posterior mean, represented here with $\bar{\theta}$). A 95% CI can be interpreted as the 95% probability that a parameter lies within that interval range. For example, if the 95% CI is between 10 and 30 ms, there is a 95% probability that the true parameter value is between 10 and 30 ms, with extreme values being less likely than values in the centre of the interval.

In each model, priors are specified for each of the parameters to be estimated. The priors are in the form of particular distributions, like the Gaussian (normal) or

⁵I am not advocating here against p -values in absolute terms. On the contrary, p -values are still useful in that they provide us with a practical solution in situations that involve, for example, decision-making.

the Cauchy distribution. A prior defines the prior knowledge of where the parameter might lie within a range of values. For example, a prior as a normal distribution with mean 200 ms and standard deviation 50 indicates the researcher's belief that the parameter lies between 100 and 300 ms with 95% probability (i.e., the mean minus twice the standard deviation, and the mean plus twice the standard deviation).

Statistical analysis was performed in R v3.5.3 (R Core Team 2019). Bayesian regression models were fit with brms (Bürkner 2017, 2018). Each model was run with four MCMC chains and 2000 iterations per chain, of which 1000 for warm-up. A Gaussian (normal) distribution was used in all the models as the response distribution. All factors were coded using treatment contrasts (the first level in this list was set as the reference level): number of syllables (disyllabic, monosyllabic), vowel (/ɑ:/, /ɜ:/, /i:/), C2 voicing (voiceless, voiced), C2 place of articulation (velar, labial). Speech rate was centred when included in the models so that the intercept could be interpreted as the intercept at mean speech rate. A seed (1234) was set in all models to ensure reproducibility of the output. The priors used in the models reported here will be discussed along with the results in the following sections.

A concern could be raised that the priors might have greater influence on the posterior distributions than the observed data, which would be undesired. A sensitivity analysis based on posterior z-scores and shrinkage (Betancourt 2018) indicates that the posterior distributions discussed in this study are not just based on the priors, but are also informed by the data.

3 Results

This section reports the results of the Bayesian models, grouped by outcome variable (release-to-release, vowel duration, closure duration). A description of the model structure and priors is given for each model, followed by the presentation of the posterior distributions of the relevant terms. Each model is assigned a number (1 to 5), and the text refers to these.

Model convergence was reached in all the reported models ($\hat{R} = 1$) and no major divergences in the MCMC chains were observed. The posterior predictive check plots indicate that the observed distributions are slightly positively skewed so that a log-normal distribution would have been more appropriate. Previous work has shown that speech-units duration does follow, as a general trend, a log-normal distribution (Rosen 2005; Ratnikova 2017), and the practice of transforming duration data to the logarithmic scale is not uncommon (Gahl & Baayen 2019). However, the deviations from a Gaussian distribution here were minimal, and an informal comparison of one of the models fitted with a log-normal distribution led to virtually identical results.

Table 2: Summary of the Bayesian regression fitted to release-to-release duration (model 1, see Section 3.1)

Predictor	Mean	SD	Q2.5	Q97.5	CI width
Intercept	263.71	9.64	244.17	283.00	38.84
Voicing = voiced	-4.43	10.03	-23.86	15.45	39.30
Num. syll. = monosyllabic	17.34	9.76	-1.58	36.53	38.11
Speech rate (cntr.)	-36.10	2.06	-40.14	-32.13	8.01
voiced \times monosyll.	16.53	12.72	-8.41	41.41	49.83

3.1 Release-to-release duration

A Bayesian regression was fit to model the duration of the release-to-release interval (model 1). The following terms were included as fixed effects: C2 voicing (voiceless, voiced), number of syllables (disyllabic, monosyllabic), centred speech rate, an interaction between C2 voicing and number of syllables. A by-speaker and by-word random intercept, and a by-speaker random coefficient for C2 voicing were entered as random effects. The following priors were used.⁶ Two weakly informative priors based on the results from Coretta (2019b) were chosen for the intercept and the effect of C2 voicing. The former prior is a normal distribution with mean 200 ms and SD = 50, while the latter a normal distribution with mean 0 ms and SD = 25. A weakly informative prior as a normal distribution with mean 50 ms and SD = 25 was specified for the effect of number of syllables. The prior is based on differences in vowel duration between mono- vs. disyllabic words, which range between 30 and 100 ms (Sharf 1962; Klatt 1973). The same prior was used for the interaction between C2 voicing and number of syllables, based on the reported differences in voicing effect in mono- vs. disyllabic words (Sharf 1962; Klatt 1973). The prior for the effect of centred speech rate is a normal distribution with mean -25 ms and SD = 10, and is based on results from Coretta (2019b). For the random effects, a half Cauchy distribution (location = 0, scale = 25) was used for the standard deviation and the residual standard deviation, and a LKJ(2) distribution for the correlation among the random terms.

Table 2 gives the posterior mean, posterior standard deviation, 2.5 and 97.5 quantiles (lower and upper bounds of the 95% credible interval), and the credible interval's width of the fixed effects of model 1. The precision goal (CI width \leq 20 ms, based on the ROPE) was reached only for centred speech rate (CI width = 8.14 ms). The posterior distribution of the estimated effect of C2 voicing on the release-to-release duration has a 95% credible interval (95% CI) between -23.86

⁶A weakly informative prior is a prior that intentionally provides weaker information on the effect than what is known about an effect (Gelman 2006), while conversely a strongly informative prior is a prior that conveys greater information.

Table 3: Summary of the Bayesian regression fitted to release-to-release duration (model 2, see Section 3.1)

Predictor	Mean	SD	Q2.5	Q97.5	CI width
Intercept	289.05	8.14	273.01	305.09	32.08
Vowel = /ɜ:/	-8.58	6.90	-21.90	4.87	26.78
Vowel = /i:/	-36.94	6.96	-50.10	-22.26	27.84
C2 place = labial	2.46	5.68	-9.15	13.28	22.44
Speech rate (cntr.)	-37.48	2.05	-41.51	-33.37	8.14

and 15.45 ms (the mean is -4.43 ms, SD = 10.03). The 95% CI of the estimated interaction between C2 voicing and number of syllables tends towards positive values, between -8.41 and 41.41 ms ($\bar{\theta}$ = 16.53 ms, SD = 12.72). The difference in duration of the release-to-release interval between monosyllabic and disyllabic words is more clearly positive, between -1.58 and 36.53 ms (95% CI, $\bar{\theta}$ = 17.34, SD = 9.76). Speech rate has a strong negative effect on the release-to-release duration with 95% CI = [-40.14, -32.13].

A second Bayesian regression (model 2) was fitted with the release-to-release duration as the outcome variable to test the effects of vowel and C2 place of articulation, which were entered as terms in the model without interactions. Centred speech rate was also included. The random effects structure was the same as with the first model. The relevant priors from the first model were kept. For the effects of vowel (/ɜ:/, /i:/) and place of articulation (labial), the very weakly informative prior used is a normal distribution with mean = 0 ms and SD = 30. This prior was based on duration differences depending on vowel height (Heffner 1937; House & Fairbanks 1953; Hertrich & Ackermann 1997) and labial place of articulation (Sharf 1962), which both range between 10 and 30 ms.

The summary of the fixed effects of model 2 are given in Table 3. As with model 1, only the CI width of speech rate reached the intended precision. The posterior distribution of the effect of the vowel /ɜ:/ shows that this vowel tends to a negative effect, with a 95% CI between -21.90 and 4.87 ms ($\bar{\theta}$ = -8.58 ms, SD = 6.9). The vowel /i:/ has a more robust negative effect on release-to-release duration, with a 95% CI between -50.10 and -22.26 ($\bar{\theta}$ = -36.94 ms, SD = 6.96). Less clear is the effect of C2 place of articulation (velar vs. labial stop): The mean of the posterior is 2.46 ms (SD = 5.68), and the 95% CI is [-9.15, 13.28].

The credible intervals of the effects in the models reported above have widths which are greater than the chosen ROPE width of 20 ms. The wide credible intervals indicate that the estimated posterior distributions of the effects have a somewhat high degree of uncertainty with them. This uncertainty is potentially due to not controlling for vowel and number of syllables in the first and second model respectively. An exploratory model (model 3) was thus fitted to the data, in which

Table 4: Summary of the Bayesian regression fitted to release-to-release duration and predictors from model 1 and 2 (model 3, see Section 3.1)

Predictor	Mean	SD	Q2.5	Q97.5	CI width
Intercept	280.81	6.99	266.72	294.37	27.66
Voicing = voiced	-2.43	4.06	-10.45	5.65	16.10
Num. syll. = monosyllabic	16.03	3.32	9.17	22.48	13.31
Vowel = /ɜ:/	-10.05	2.95	-15.92	-4.24	11.68
Vowel = /i:/	-39.03	2.99	-45.03	-32.76	12.27
C2 place = labial	2.46	2.39	-2.29	7.28	9.57
Speech rate (cntr.)	-36.10	1.99	-39.96	-32.24	7.72
voiced × monosyll.	11.67	4.71	2.65	20.98	18.33

all the terms from the two models above were included. The same priors of the two separate models were used in the combined model.

Including all the relevant terms in the model (C2 voicing and place, vowel, number of syllables in interaction with C2 voicing) reduces the width of the credible intervals substantially. Figure 1 shows the posterior distributions of the model terms with a variety of credible intervals. The posterior distribution of the C2 voicing effect on release-to-release duration is tighter than that of model 1 (95% CI = [-10.45, 5.65]) while the mean (-2.43 ms, SD = 4.06) is virtually unchanged (-4.43 ms, only a 2 ms difference). The estimated effect of syllable number is robustly positive (95% CI = [9.17, 22.48]), with a mean (16.03 ms, SD = 3.32) similar to that in model 1. The posterior distribution of the interaction between number of syllables and C2 voicing (95% CI = [2.65, 20.98]) suggests a positive and medium-sized interaction effect ($\bar{\theta}$ = 11.67 ms, SD = 4.71). This result indicates that the duration of the release-to-release is greater in monosyllabic words with voiced C2 than in monosyllabic words with voiceless C2. The effects of vowel and place of articulation have similar means as in model 2, but the credible intervals are smaller. The release-to-release is on average 10.05 ms (SD = 2.95, 95% CI = [-15.92, -4.24]) shorter if the vowel is /ɜ:/ and 39.3 ms (SD = 2.99, 95% CI = [-45.03, -32.76]) shorter if the vowel is /i:/. C2 place of articulation (labial) has a negligible positive mean effect (2.6 ms, SD = 2.39, 95% CI = [-2.29, 7.28]).

3.2 Vowel duration

A Bayesian regression model was fitted to test vowel duration (model 4). The following terms were entered: C2 voicing (voiceless, voiced), vowel (/a:/, /ɜ:/, /i:/), number of syllables (disyllabic, monosyllabic), centred speech rate, all possible interactions between C2 voicing, vowel, and number of syllables. The same random structure as in the previous models was used (a by-speaker and by-word random intercept, and a by-speaker random coefficient for C2 voicing).

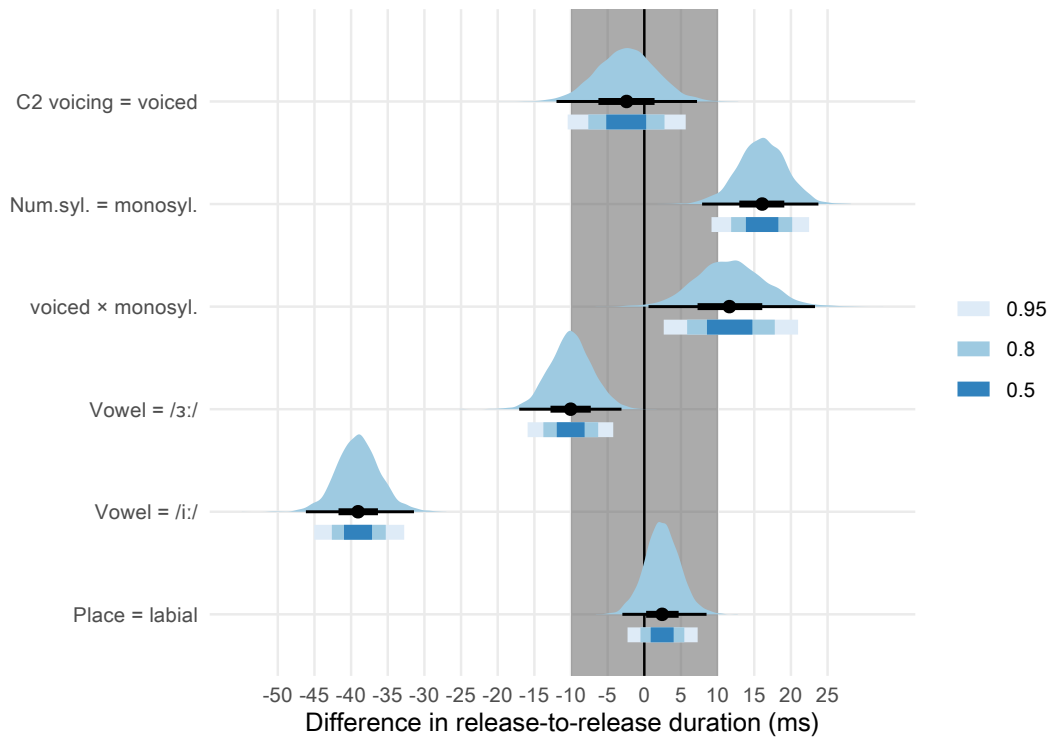


Figure 1: Posterior distributions and Bayesian credible intervals of the effects on release-to-release duration (model 3). For each effect, the thick blue-coloured bars indicate (from darker to lighter) the 50%, 80%, and 95% CI. The black point with bars are the posterior median (the point), the 98% (thin bar) and 66% (thicker bar) CI. The shaded grey area around 0 is the ROPE.

Table 5: Summary of the Bayesian regression fitted to vowel duration (model 4, see Section 3.2)

Predictor	Mean	SD	Q2.5	Q97.5	CI width
Intercept	124.91	5.96	112.94	136.77	23.83
Voicing = voiced	13.65	5.16	3.73	24.09	20.36
Vowel = /ɜ:/	-9.03	5.13	-19.08	1.63	20.71
Vowel = /i:/	-36.77	5.00	-46.42	-26.67	19.74
Num. syll. = monosyllabic	14.91	5.07	5.15	25.14	19.99
Speech rate (cntr.)	-18.03	1.48	-20.93	-15.29	5.63
voiced × /ɜ:/	0.24	6.83	-13.70	13.94	27.64
voiced × /i:/	6.73	6.59	-6.54	19.26	25.80
voiced × monosyll.	4.03	6.70	-8.98	17.69	26.67
/ɜ:/ × monosyll.	0.53	7.07	-13.57	14.57	28.15
/i:/ × monosyll.	-16.07	6.93	-30.03	-2.68	27.35
voiced × /ɜ:/ × monosyll.	-2.94	9.46	-21.37	15.77	37.14
voiced × /i:/ × monosyll.	14.46	9.18	-3.59	31.99	35.58

For the prior of the intercept of vowel duration, a normal distribution with mean 145 ms and standard deviation 30 was used (Heffner 1937; House & Fairbanks 1953; Peterson & Lehiste 1960; Sharf 1962; Chen 1970; Klatt 1973; Davis & Summers 1989; Laeuffer 1992; Ko 2018). A normal distribution with mean 50 ms and standard deviation 20 was used as the prior for the effect of voicing on vowel duration (based on the above studies). A normal prior with mean 50 and standard deviation 25 was chosen instead for the effect of number of syllables and the interaction C2 voicing/number of syllables. For the effects of vowel, vowel/number of syllables interaction, and the three-way interaction vowel/number of syllables/C2 voicing, the prior was a normal distribution with mean 0 and standard deviation 30, based on differences reported in the studies above. A slightly more informative prior was used for the interaction between C2 voicing and vowel (mean = 0, SD = 20). The same priors as in the previous models were included for the random effects.

Table 5 reports the summary of model 4, while Figure 2 shows the posterior distributions and credible intervals. The precision target was reached in the non-interacting predictors (permitting a few milliseconds above 20), with the exception of the intercept. All the interactions terms have CI widths above 25 ms. The 95% CI of the posterior distribution of the duration of /ɑ:/ is included in the range 112.94–136.77 ms ($\bar{\theta}$ = 124.91 ms, SD = 5.96). The vowel /ɜ:/ is 9.03 ms shorter (SD = 5.16) with CI = [-19.08, 1.63], while /i:/ is 36.77 ms shorter (SD = 5, 95% CI = [-46.42, -26.67]). C2 voicing has a small but robust positive effect on vowel duration in disyllabic words. The posterior distribution of the effect of voicing on /ɑ:/ has mean 13.65 ms (SD = 5.16) and 95% CI = [3.73, 24.09]. The posterior of the interaction of voicing with vowel when the vowel is /ɜ:/ is quite spread out around 0, with the 95% CI between -13.70 and 13.94 ms. This indicates that /ɑ:/

and /ɜ:/ are similar in their behaviour of voicing-driven durational differences. On the other hand, the effect of voicing is on average 6.73 ms greater (SD = 6.59, 95% CI = [-6.54, 19.26]) when the vowel is /i:/.

The magnitude of the voicing effect in disyllabic vs. monosyllabic words is modulated by the identity of the vowel. The posterior distribution for the interaction C2 voicing/number of syllables when the vowel is /ɑ:/ has mean 4.03 ms (SD = 6.7) and 95% CI [-8.98, 17.69]. This distribution indicates the possibility for a very small increase of the effect from disyllabic to monosyllabic words with /ɑ:/. The three-way interaction C2 voicing/vowel/number of syllables suggests that the effect of voicing in monosyllabic words with /ɜ:/ is very similar to that of monosyllabic /ɑ:/-words ($\bar{\theta} = -2.94$, SD = 9.46, 95% CI = [-21.37, 15.77]). On the other hand, the effect increases by 14.46 ms (SD = 9.18, CI = [-3.59, 31.99]) in monosyllabic words with /i:/ relative to disyllabic /i:/-words. Note that the credible intervals of these interaction effect are quite large, so that a wide range of values are probable at 95% confidence.

3.3 Consonant closure duration

To test various effects on C2 closure duration, model 5 was fit with closure duration as the outcome variable and the following predictors: C2 voicing (voiceless, voiced), C2 place of articulation (velar, labial), number of syllables (disyllabic, monosyllabic), all interactions between these predictor terms, and centred speech rate. The random effects were again a by-speaker and a by-word random intercept, and a by-speaker random coefficient for C2 voicing.

As priors, a normal distribution with mean 90 ms (SD = 20) was used for the intercept, based on Sharf (1962) and Luce & Charles-Luce (1985). The means reported in these studies also indicate that the closure of the stop in monosyllabic words is 10-30 ms shorter when the stop is voiced. A normal distribution with mean -20 ms (SD = 10) was chosen as the prior of the effect of C2 voicing on closure duration. The same studies indicate that labial stops have a closure which is 10-20 ms longer than the closure of velar stops. For the effect of C2 place, a normal distribution with mean 15 ms (SD = 10) was used.

The summary of model 5 is shown in Table 6. See Figure 3 for the posteriors and credible intervals of the effects. The 96% CI width of all the terms, with the exception of the three-way interaction (voicing/place/number of syllables), is below 20 ms (the precision goal has been reached). The posterior distribution of the intercept for closure duration (corresponding to the duration of voiceless velar stops in disyllabic words) has mean 74.75 ms (SD = 2.86) and 95% CI = [69.07, 80.59]. The effect of C2 voicing on closure duration is certainly negative, between -26.77 and -14.74 ms (95% CI). The posterior mean of this effect is -20.79 ms (SD = 3.06). A very small positive effect of place of articulation (labial)

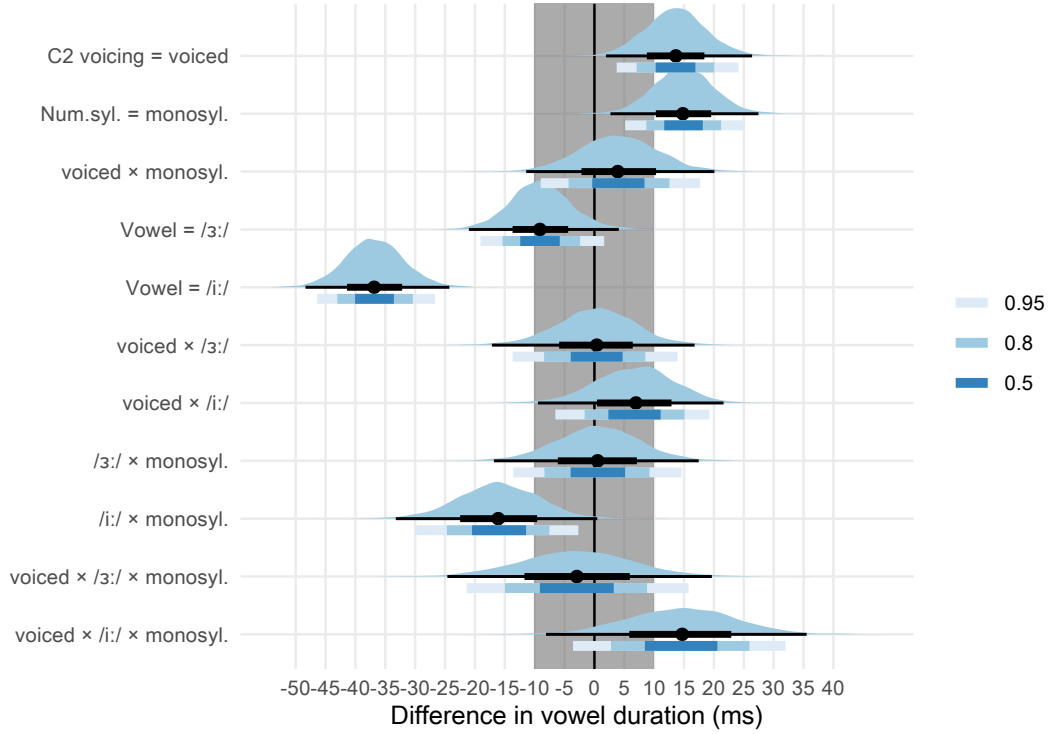


Figure 2: Posterior distributions and Bayesian credible intervals of the effects on vowel duration (model 4). For each effect, the thick blue-coloured bars indicate (from darker to lighter) the 50%, 80%, and 95% CI. The black point with bars are the posterior median (the point), the 98% (thin bar) and 66% (thicker bar) CI. The shaded grey area around 0 is the ROPE.

Table 6: Summary of the Bayesian regression fitted to closure duration (model 5, see Section 3.3)

Predictor	Mean	SD	Q2.5	Q97.5	CI width
Intercept	74.75	2.86	69.07	80.59	11.52
Voicing = voiced	-20.79	3.06	-26.77	-14.74	12.03
C2 place = labial	5.19	2.77	-0.03	10.76	10.79
Num. syll. = monosyllabic	2.98	2.90	-2.80	8.77	11.58
Speech rate (cntr.)	-9.21	1.26	-11.71	-6.74	4.97
voiced × labial	1.37	3.94	-6.79	8.93	15.72
voiced × monosyll.	1.82	4.06	-6.08	9.70	15.78
labial × monosyll.	-0.74	4.02	-8.95	6.88	15.83
voiced × labial × monosyll.	6.41	5.66	-4.72	17.45	22.17

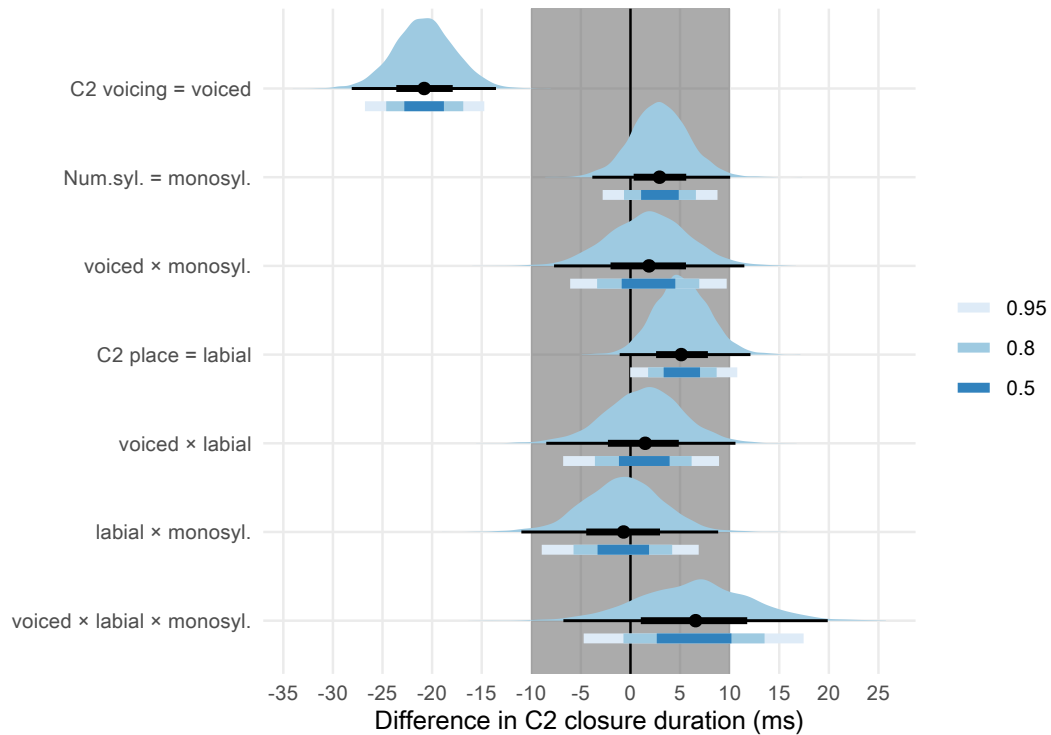


Figure 3: Posterior distributions and Bayesian credible intervals of the effects on closure duration (model 4). For each effect, the thick blue-coloured bars indicate (from darker to lighter) the 50%, 80%, and 95% CI. The black point with bars are the posterior median (the point), the 98% (thin bar) and 66% (thicker bar) CI. The shaded grey area around 0 is the ROPE.

is suggested by the 95% CI from -0.03 to 10.76 ms ($\bar{\theta} = 5.19$ ms, $SD = 2.77$). A possibly even smaller effect of number of syllables or no effect at all can be inferred from the posterior distribution which has mean 2.98 ms and $SD 2.9$ (95% CI = $[-2.8, 8.77]$). Note that the 95% CIs of the posterior distributions of all the effects, with the exception for the effect of voicing, are within the ROPE around 0.

4 Discussion

This study set out to build on the results discussed in Coretta (2019b) by investigating durational properties of the release-to-release interval in English monosyllabic and disyllabic words. It was expected that the release-to-release interval would not be affected by C2 voicing in disyllabic words but it would in monosyllabic words.

Moreover, a conceptual replication of studies on the effect of consonant voicing on vowel and closure durations was sought, with a focus on comparing the effect in mono- vs. disyllabic words. This section discusses in turn the results in relation to the release-to-release interval duration (Section 4.1) and to vowel and closure durations (Section 4.2) by comparing them with the hypotheses of this study. Section 4.3 synthesises and links these findings back to the articulatory grounding of the temporal properties of the release-to-release interval in mono- and disyllabic words (Section 1). Limitations and future work are also discussed.

4.1 Release-to-release interval

The first question (see Section 1.3) asked whether the voicing of C2 in disyllabic and monosyllabic words in English influences the duration of the release-to-release interval. Coretta (2019b) showed that the release-to-release interval duration is not affected by C2 voicing in disyllabic words of Italian and Polish. The hypotheses were that, in English, the interval is not affected in disyllabic words, like in Italian and Polish, but that it is in monosyllabic words. In sum, the results of this study indicate that the release-to-release duration of disyllabic words in English is relatively stable independent of whether C2 is voiceless (like in /tɑ:pəs/) or voiced C2 (/tɑ:bəs/). On the other hand, the release-to-release in monosyllabic words is longer if C2 is voiced (like in /tɑ:b/ vs. /tɑ:p/).

Two pre-registered Bayesian regression models were fitted to the release-to-release duration (model 1-2). The established ROPE target has not been achieved (see Section 2.1). An exploratory model (model 3) including all predictors from model 1 and 2 resulted in higher estimate precision (CI widths below 20 ms). The results of model 3 suggest a negligible effect of C2 voicing on the interval duration in disyllabic words (hypothesis 1a), with a 95% probability that the true effect is between -10 and +5 ms. At lower levels of probability, the posterior distribution indicates an effect between -6 and 1 ms (60% probability). If the voicing of C2 is conditioning the duration of the release-to-release interval, this effect is very small.

The possible small effect of C2 voicing in disyllabic words could be related to an annotation bias which affects the identification of stop releases. English voiceless stops are generally followed by aspiration, and the glottal friction that makes up aspiration could mask the burst of the release. If the release of the post-vocalic voiceless stops is annotated later than the actual release (by mistaking peaks in the aspiration noise for the release burst), this could lead to longer release-to-release durations when C2 is voiceless compared to when it is voiced. Such annotation bias could explain the quite small negative effect of voicing on the interval duration, and why it is in the opposite direction of the one predicted for monosyllabic words (i.e. *longer* release-to-release when C2 is voiced).

On the other hand, the release-to-release interval in monosyllabic words is longer when C2 is voiced (for example, /tɑ:b/) vs. when it is voiceless (/tɑ:p/). The interaction term between number of syllables in the word and C2 voicing is positive, between +2.5 and +21 ms (at 95% probability), which means that the effect of C2 voicing increases by 2.5 to 21 ms in monosyllabic words relative to the effect in disyllabic words. This result is compatible with hypothesis 1b that the release-to-release interval is longer in monosyllabic words with a voiced C2 than in monosyllabic words with a voiceless C2. As discussed in Section 1, the absence of release-to-release isochrony in monosyllabic words is possibly due to the absence of a second vowel which would constitute the left articulatory anchor for vowel isochrony, which in turn is argued to be the necessary element for the release-to-release temporal stability.

The second question posed at the beginning of the paper was about other effects on the release-to-release duration. As expected by hypothesis 2a, the release-to-release is longer in monosyllabic than in disyllabic words. At 95% probability, the effect of number of syllables (from di- to monosyllabic) is between 9 and 22.5 ms. As for hypothesis 2b, the results are more robust for /i:/ than for /ɜ:/. When the vowel is /i:/, the release-to-release interval is 33 to 45 ms shorter compared with an interval with /ɑ:/. The posterior distribution of the effect when the vowel is /ɜ:/ substantially overlaps with the ROPE, although it tends towards the negative side. If there is an effect with this vowel compared to /ɑ:/, it is negative and possibly around -10 ms. Finally, hypothesis 2c is not unequivocally corroborated. The posterior distribution of the effect of C2 place of articulation (labial) has very high precision (9.5 ms) and it is between 0 and 5 ms (at somewhat less than 80% probability). However, it lies within the ROPE and it is very close to 0.

4.2 Vowel and closure duration

Question 3 addressed the effect of voicing on vowel and closure duration, and the possible differences between disyllabic and monosyllabic words. The effect of voicing on vowel duration found in this study was estimated to lie between 4 and 25 ms. This range of values is very similar to that reported in Coretta (2019b) for Italian and Polish disyllabic words (the 95% confidence interval for the effect in these languages is [8, 25]), monosyllabic words were not tested). When compared to the values in previous studies that investigated disyllabic words (Sharf 1962; Klatt 1973; Davis & Summers 1989), the effect size found in this study tends towards smaller values. However, note that the posterior distribution of the effect in the current study is entirely contained in the meta-analytical posterior distribution of the effect in the other studies, which roughly ranges between -15 and +65 ms (see Supplement A). Thus, we can assume that the deviation of this study from previous ones is not substantial. As for the effect of number of syllables on vowel

678 duration, a similar effect to that of voicing was found, whereby vowel durations
679 increase by 5 to 25 ms in monosyllabic words compared to disyllabic words. This
680 relation corresponds to what has previously been reported in the literature. Finally,
681 given that the 95% CIs of the effects of voicing and number of syllables overlap
682 with the right side of the ROPE without including 0, the data supports positive
683 effects, but inference on their magnitude should be carefully wighted.

684 It was expected that the voicing effect on vowels would be stronger in mono-
685 syllabic than in disyllabic words (hypothesis 3). The credible intervals of the pos-
686 terior distributions from model 4, which are larger than the ROPE, make interpre-
687 tation less straightforward. At 80% probability, the difference in voicing effect
688 between mono- and disyllabic words is between -5 and $+12.5$ ms. The distribu-
689 tion is skewed towards the positive side, and this is compatible with results from
690 previous studies, although the CI includes 0. The magnitude, however, is consider-
691 ably lower than what previously reported. More data is needed to reach a sensible
692 estimate precision and reduce uncertainty.

693 The three-way interaction between C2 voicing, vowel, and number of sylla-
694 bles reveals that the effect in monosyllabic words with the vowel /ɜ:/ is similar to
695 that with /ɑ:/. On the other hand, the effect is larger if the vowel is /i:/. Model
696 4 estimates an effect increase of about 14.5 ms ($[-4.27, 33.41]$). Note that the
697 credible interval is very wide (38 ms) and it spans over both negative and positive
698 values, although tends more towards the latter. Moreover, the vowel /i:/ followed
699 by a voiceless stop has, according to the model, the same duration in monosyllabic
700 and disyllabic words. While it is not clear why the vowel should have the same
701 duration in these contexts, this pattern suggest a possible process of /i:/ shortening
702 in monosyllabic words. More research is warranted in relation to the observed
703 patterns.

704 Turning now to consonants, it was expected for voiced closures to be longer
705 than voiceless closures, but there was no specific hypothesis concerning the effect
706 of number of syllables on the effect of voicing on closure durations. C2 voicing
707 has a robust negative effect on closure duration, so that voiced closures are 14.6-
708 26.8 ms shorter than voiceless closures. The effects of number of syllables, place,
709 and interactions all have credible intervals that are narrower than 20 ms (the ROPE
710 width) but they lie entirely within the ROPE around 0. If these variables do have
711 an effect on closure duration, the present analysis suggests that the means of these
712 effects are between 0 and 5 ms. These values are smaller than what the results in
713 Sharf (1962), which indicate a difference of 15 ms between velar and labial closure
714 durations.

715 As a general trend, the differences in vowel and closure duration found in this
716 study are smaller than those known from the literature, and considerably so in the
717 case of vowels. A possible reason for this discrepancy could be found in problems
718 arising from Type M errors (as briefly discussed in Section 1), and in differences

of speech rate, as evidenced by comparing average segment durations. While the model's intercept of vowel duration in this study is approximately 125 ms (SD = 5.89), the mean vowel duration in the studies surveyed in the meta-analysis (Supplement A) is 150 ms (SD = 36). These longer durations may indicate lower speech rates in older studies and so the effect of voicing may have been greater there than at higher speech rates, assuming a linear increase of the effect. However, the ratio between vowel duration and the effect of voicing differs (a third in this study vs. half in previous work). Ko's findings 2018 support the idea that the voicing effect (and the vowel-to-consonant ratio) are not stable across speaking rates, with the consequence that differences are enhanced at decreased speaking rates. More studies like Ko (2018) are needed to settle the issue of the diverging results.

4.3 General discussion

Coretta (2019b) proposes that the voicing-related adjustments in the relative timing of the closure onset within an isochronous speech interval (acoustically identified as the release-to-release interval) is the diachronic precursor of the cross-linguistically widespread effect of voicing on vowel duration.⁷ Given that the duration of the release-to-release interval in Italian, Polish, and English disyllabic words is not affected by the voicing of the post-vocalic consonant, the relative durations of vowel and closure are thought to depend on the timing of the VC boundary within that interval. A later VC boundary implies a longer vowel and a shorter closure, while, vice versa, an earlier boundary produces a shorter vowel and a longer closure. Behind the differential timing of the VC boundary within the release-to-release interval, several other accounts can be envisaged, like accounts relating to laryngeal and supraglottal adjustments (Halle & Stevens 1967; Beguš 2017; Coretta 2019c).

The absence of temporal stability in monosyllabic words needs to be reconciled with the presence of the voicing effect in this context. A possible solution to the incongruence could be sought in diachrony (Blevins 2004, 2006), by speculating that the release-to-release interval was temporally stable even in monosyllabic words in earlier historical stages, via two possible scenarios. When a monosyllabic word historically derives from a disyllabic word, it could be further conjectured that the monosyllabic word has simply inherited the isochrony of the release-to-release interval and, with it, the voicing effect from its disyllabic predecessor. Alternatively, the emergence of the voicing effect in monosyllabic words could just be a direct consequence of mechanisms of VC boundary timing, as mentioned above.

Independent on the pathway to the voicing effect, subsequent perceptual biases,

⁷Note that isochrony here is intended as pertaining the context of voiceless vs. voiced stops only.

755 like the ones proposed by the perceptual accounts by Javkin (1976) and Kluender
756 et al. (1988), can further contribute to the enhancement of the effect of voicing, for
757 example as a means to enhance the perceptual difference of voiceless vs. voiced
758 stops (Lisker 1974, 1986; Stevens & Keyser 1989). In the case of disyllabic words,
759 movements of the VC boundary within the isochronous interval will logically af-
760 fect both vowel duration and closure duration. On the other hand, the absence
761 of a temporal articulatory anchor in monosyllabic words would allow articulatory
762 stretching or compression to operate independently on the vocalic and the conso-
763 nantal gestures. The articulatory studies in Raphael (1972) and de Jong (1991) do
764 suggest that the vocalic gesture is executed for a prolonged time when the follow-
765 ing consonant is voiced. While differences in the magnitude of the voicing effect
766 should be replicated in future studies, the potentially greater effect of voicing in
767 monosyllabic words could be ascribed to unconstrained mechanisms affecting the
768 VC boundary (articulatory and/or perceptual).

769 5 Conclusion

770 This paper set out to investigate temporal properties of the so-called ‘voicing ef-
771 fect’, by which vowels are shorter when followed by voiceless stops and longer
772 when followed by voiced stops. Coretta (2019b) proposes that the voicing effect
773 emerges via a mechanism of relative timing of the VC boundary within a tempo-
774 rally stable interval. Such interval was argued to be the interval between two con-
775 secutive releases, as evidenced by acoustic data from Italian and Polish disyllabic
776 words. The temporal stability of the release-to-release in relation to consonantal
777 voicing is thought to derive from two properties of gestural phasing, namely the
778 isochrony of the distance between the vowels in a VCV sequence, and in-phase
779 alignment of onset consonants and the following vowel. On the other hand, the
780 lack of an articulatory anchor (a second vowel) in monosyllabic words would al-
781 low the release-to-release duration to be affected by C2 voicing and differ in the
782 monosyllabic context.

783 This study adds to the current status of knowledge on temporal aspects of the
784 voicing effect by showing that the release-to-release interval is not affected by C2
785 voicing in English disyllabic words, as in Italian and Polish, and that, instead, it
786 is longer in monosyllabic words when C2 is voiced. While the timing of the VC
787 boundary within the release-to-release in disyllabic words affects both vowel and
788 closure durations in a logically dependent way, vowel and closure durations can be
789 modulated more independently in monosyllabic words. The less constrained op-
790 eration of production and perceptual mechanisms affecting the timing of the VC
791 boundary was argued to be the reason for the seemingly greater effect of voicing
792 reported for monosyllabic words. The data in this study, and the cumulative evi-

dence from previous studies as evinced by a Bayesian meta-analysis, however, do not equivocally provide support for a difference in the effect between mono- and disyllabic words, and future work is necessary to shed light on the matter.

To conclude, the results of this study suggest some directions of research. Future studies should further investigate the articulatory temporal patterns of vocalic and consonantal gestures in disyllabic words. In particular, a complete assessment of the isochrony (or lack thereof) of consecutive vocalic gestures should include a variety of oppositions, involving voicing, place of articulation, number of consonants, syllabic affiliation, and prosodic contexts. Moreover, work is needed to shed light on the timing of the consonant closing gesture relative to the articulatory gesture of the preceding vowel in voiceless vs. voiced stops. Finally, the scenario of emergence of the voicing effect offered here should be examined in relation to other consonantal effects on vowel duration, like other laryngeal effects and effects of manner of articulation.

References

- Ananthapadmanabha, Tirupattur V., Aragulla Prasad Prathosh & Angarai Ganesan Ramakrishnan. 2014. Detection of the closure-burst transitions of stops and affricates in continuous speech using the plosion index. *The Journal of the Acoustical Society of America* 135(1). 460–471. doi:10.1121/1.4836055.
- Beguš, Gašper. 2017. Effects of ejective stops on preceding vowel duration. *The Journal of the Acoustical Society of America* 142(4). 2168–2184. doi:10.1121/1.5007728.
- Belasco, Simon. 1953. The influence of force of articulation of consonants on vowel duration. *The Journal of the Acoustical Society of America* 25(5). 1015–1016.
- Betancourt, Michael. 2018. Calibrating model-based inferences and decisions. arXiv preprint arXiv:1803.08393.
- Bigi, Brigitte. 2015. SPPAS - Multi-lingual approaches to the automatic annotation of speech. *The Phonetician* 111–112. 54–69.
- Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.
- Blevins, Juliette. 2006. A theoretical synopsis of Evolutionary Phonology. *Theoretical linguistics* 32(2). 117–166. doi:10.1515/TL.2006.009.

- 826 Browman, Catherine P. & Louis Goldstein. 1988. Some notes on syllable struc-
827 ture in articulatory phonology. *Phonetica* 45(2-4). 140–155. doi:10.1159/
828 000261823.
- 829 Browman, Catherine P. & Louis Goldstein. 1992. Articulatory phonology: An
830 overview. *Phonetica* 49. 155–180. doi:10.1159/000261913.
- 831 Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel mod-
832 els using Stan. *Journal of Statistical Software* 80(1). 1–28. doi:10.18637/jss.
833 v080.i01.
- 834 Bürkner, Paul-Christian. 2018. Advanced Bayesian multilevel modeling with the
835 R package brms. *The R Journal* 10(1). 395–411. doi:10.32614/RJ-2018-017.
- 836 Celata, Chiara & Paolo Mairano. 2014. On the timing of V-to-V intervals in Italian:
837 a review, and some new hypotheses. *Revista de Filología Románica* 31. 37. doi:
838 10.5209/rev_RFRM.2014.v31.n1.51022.
- 839 Chen, Matthew. 1970. Vowel length variation as a function of the voicing of the
840 consonant environment. *Phonetica* 22(3). 129–159. doi:10.1159/000259312.
- 841 Coretta, Stefano. 2019a. Compensatory aspects of the effect of voicing on vowel
842 duration in English [Data]. Open Science Framework [https://osf.io/
843 ep8wb/](https://osf.io/ep8wb/).
- 844 Coretta, Stefano. 2019b. An exploratory study of voicing-related differences in
845 vowel duration as compensatory temporal adjustment in Italian and Polish. Ac-
846 cepted for publication at Glossa.
- 847 Coretta, Stefano. 2019c. Longer vowel duration correlates with greater tongue
848 root displacement: Acoustic and articulatory data from Italian and Polish. OSF
849 pre-print. doi:10.31219/osf.io/zrqyx.
- 850 Davis, Stuart & W. Van Summers. 1989. Vowel length and closure duration in
851 word-medial VC sequences. *Journal of Phonetics* 17. 339–353. doi:10.1121/1.
852 2026892.
- 853 Durvasula, Karthik & Qian Luo. 2012. Voicing, aspiration, and vowel duration in
854 Hindi. *Proceedings of Meetings on Acoustics* 18. 1–10. doi:10.1121/1.4895027.
- 855 Esposito, Anna. 2002. On vowel height and consonantal voicing effects: Data
856 from Italian. *Phonetica* 59(4). 197–231. doi:10.1159/000068347.

- 857 Etz, Alexander, Quentin F. Gronau, Fabian Dablander, Peter A. Edelsbrunner &
858 Beth Baribault. 2018. How to become a Bayesian in eight easy steps: An
859 annotated reading list. *Psychonomic Bulletin & Review* 25(1). 219–234. doi:
860 10.3758/s13423-017-1317-5.
- 861 Farnetani, Edda & Shiro Kori. 1986. Effects of syllable and word structure on
862 segmental durations in spoken Italian. *Speech Communication* 5(1). 17–34. doi:
863 10.1016/0167-6393(86)90027-0.
- 864 Fowler, Carol A. 1983. Converging sources of evidence on spoken and per-
865 ceived rhythms of speech: Cyclic production of vowels in monosyllabic stress
866 feet. *Journal of Experimental Psychology: General* 112(3). 386. doi:10.1037/
867 0096-3445.112.3.386.
- 868 Fowler, Carol A. 1992. Vowel duration and closure duration in voiced and un-
869 voiced stops: There are no contrast effects here. *Journal of Phonetics* 20(1).
870 143–165.
- 871 Gahl, Susanne & R. Harald Baayen. 2019. Twenty-eight years of vowels: Tracking
872 phonetic variation through young to middle age adulthood. *Journal of Phonetics*
873 74. 42–54. doi:10.1016/j.wocn.2019.02.001.
- 874 Gelman, Andrew. 2006. Prior distributions for variance parameters in hierarchical
875 models. *Bayesian analysis* 1(3). 515–534. doi:10.1214/06-BA117A.
- 876 Halle, Morris & Kenneth Noble Stevens. 1967. Mechanism of glottal vibration
877 for vowels and consonants. *The Journal of the Acoustical Society of America*
878 41(6). 1613–1613. doi:10.1121/1.2143736.
- 879 Heffner, R.-M.S. 1937. Notes on the length of vowels. *American Speech* 12. 128–
880 134. doi:10.2307/452621.
- 881 Hermes, Anne, Doris Mücke & Martine Grice. 2013. Gestural coordination of
882 italian word-initial clusters: the case of ‘impure s’. *Phonology* 30(01). 1–25.
883 doi:10.1017/S095267571300002X.
- 884 Hertrich, Ingo & Hermann Ackermann. 1997. Articulatory control of phonological
885 vowel length contrasts: Kinematic analysis of labial gestures. *The Journal of*
886 *the Acoustical Society of America* 102(1). 523–536. doi:10.1121/1.419725.
- 887 House, Arthur S. & Grant Fairbanks. 1953. The influence of consonant environ-
888 ment upon the secondary acoustical characteristics of vowels. *The Journal of*
889 *the Acoustical Society of America* 25(1). 105–113. doi:10.1121/1.1906982.

- 890 Huggins, A. William F. 1972. Just noticeable differences for segment duration
891 in natural speech. *The Journal of the Acoustical Society of America* 51(4B).
892 1270–1278. doi:10.1121/1.1912971.
- 893 Hussein, Lutfi. 1994. *Voicing-dependent vowel duration in Standard Arabic and*
894 *its acquisition by adult American students*: Columbus, OH: The Ohio State Uni-
895 versity dissertation.
- 896 Jacewicz, Ewa, Robert Allen Fox & Samantha Lyle. 2009. Variation in stop
897 consonant voicing in two regional varieties of American English. *Jour-*
898 *nal of the International Phonetic Association* 39(3). 313–334. doi:10.1017/
899 S0025100309990156.
- 900 Javkin, Hector R. 1976. The perceptual basis of vowel duration differences associ-
901 ated with the voiced/voiceless distinction. *Report of the Phonology Laboratory,*
902 *UC Berkeley* 1. 78–92.
- 903 de Jong, Kenneth. 1991. An articulatory study of consonant-induced vowel dura-
904 tion changes in English. *Phonetica* 48(1). 1–17. doi:10.1121/1.2028316.
- 905 de Jong, Kenneth. 2004. Stress, lexical focus, and segmental focus in English:
906 Patterns of variation in vowel duration. *Journal of Phonetics* 32(4). 493–516.
907 doi:10.1016/j.wocn.2004.05.002.
- 908 Kawahara, Shigeto, Donna Erickson & Atsuo Suemitsu. 2017. The phonetics of
909 jaw displacement in Japanese vowels. *Acoustical Science and Technology* 38(2).
910 99–107. doi:10.1250/ast.38.99.
- 911 Kirby, James & Morgan Sonderegger. 2018. Mixed-effects design analysis for
912 experimental phonetics. *Journal of Phonetics* 70. 70–85. doi:10.1016/j.wocn.
913 2018.05.005.
- 914 Klatt, Dennis H. 1973. Interaction between two factors that influence vowel du-
915 ration. *The Journal of the Acoustical Society of America* 54(4). 1102–1104.
916 doi:10.1121/1.1914322.
- 917 Kluender, Keith R., Randy L. Diehl & Beverly A. Wright. 1988. Vowel-length
918 differences before voiced and voiceless consonants: An auditory explanation.
919 *Journal of Phonetics* 16. 153–169.
- 920 Ko, Eon-Suk. 2018. Asymmetric effects of speaking rate on the vowel/consonant
921 ratio conditioned by coda voicing in English. *Phonetics and Speech Sciences*
922 10(2). 45–50. doi:10.13064/KSSS.2018.10.2.045.

- 923 Kruschke, John. 2015. *Doing Bayesian data analysis: A tutorial with R, JAGS,*
924 *and Stan (2nd edition)*. Amsterdam, The Netherlands: Academic Press.
- 925 Laeuffer, Christiane. 1992. Patterns of voicing-conditioned vowel duration in
926 French and English. *Journal of Phonetics* 20(4). 411–440.
- 927 Lampp, Claire & Heidi Reklis. 2004. Effects of coda voicing and aspiration on
928 Hindi vowels. *The Journal of the Acoustical Society of America* 115(5). 2540–
929 2540. doi:10.1121/1.4783577.
- 930 Lehiste, Ilse. 1970a. Temporal organization of higher-level linguistic units. *The*
931 *Journal of the Acoustical Society of America* 48(1A). 111. doi:10.1121/1.
932 1974906.
- 933 Lehiste, Ilse. 1970b. Temporal organization of spoken language. In *OSU Work-*
934 *ing Papers in Linguistics*, vol. 4, 96–114. [https://linguistics.osu.edu/](https://linguistics.osu.edu/sites/linguistics.osu.edu/files/workingpapers/osu_wpl_04.pdf)
935 [sites/linguistics.osu.edu/files/workingpapers/osu_wpl_04.pdf](https://linguistics.osu.edu/files/workingpapers/osu_wpl_04.pdf).
- 936 Lindblom, Björn. 1967. Vowel duration and a model of lip mandible coordi-
937 nation. *Speech Transmission Laboratory Quarterly Progress Status Report*
938 4. 1–29. [http://www.speech.kth.se/prod/publications/files/qpsr/](http://www.speech.kth.se/prod/publications/files/qpsr/1967/1967_8_4_001-029.pdf)
939 [1967/1967_8_4_001-029.pdf](http://www.speech.kth.se/prod/publications/files/qpsr/1967/1967_8_4_001-029.pdf).
- 940 Lisker, Leigh. 1957. Closure duration and the intervocalic voiced-voiceless dis-
941 tinction in English. *Language* 33(1). 42–49. doi:10.2307/410949.
- 942 Lisker, Leigh. 1974. On “explaining” vowel duration variation. In *Proceedings of*
943 *the Linguistic Society of America*, 225–232.
- 944 Lisker, Leigh. 1986. “Voicing” in English: a catalogue of acoustic features sig-
945 naling /b/ versus /p/ in trochees. *Language and Speech* 29(1). 3–11. doi:
946 10.1177/002383098602900102.
- 947 Luce, Paul A. & Jan Charles-Luce. 1985. Contextual effects on vowel dura-
948 tion, closure duration, and the consonant/vowel ratio in speech production.
949 *The Journal of the Acoustical Society of America* 78(6). 1949–1957. doi:
950 10.1121/1.392651.
- 951 Machač, Pavel & Radek Skarnitzl. 2009. *Principles of phonetic segmentation*.
952 Praha: Epocha.
- 953 Maddieson, Ian & Jack Gandour. 1976. Vowel length before aspirated con-
954 sonants. In *UCLA Working papers in Phonetics*, vol. 31, 46–52. [https:](https://escholarship.org/uc/item/31f5j8m7)
955 [/escholarship.org/uc/item/31f5j8m7](https://escholarship.org/uc/item/31f5j8m7).

- 956 Magno Caldognetto, Emanuela, Franco Ferrero, Kyriaki Vaggas & Maria Bagno.
957 1979. Indici acustici e indici percettivi nel riconoscimento dei suoni linguis-
958 tici (con applicazione alle consonanti occlusive dell'italiano). *Acta Phoniatica*
959 *Latina* 2. 219–246.
- 960 Marin, Stefania & Marianne Pouplier. 2010. Temporal organization of complex
961 onsets and codas in American English: Testing the predictions of a gestural
962 coupling model. *Motor Control* 14(3). 380–407. doi:10.1123/mcj.14.3.380.
- 963 Marin, Stefania & Marianne Pouplier. 2014. Articulatory synergies in the temporal
964 organization of liquid clusters in Romanian. *Journal of Phonetics* 42. 24–36.
965 doi:10.1016/j.wocn.2013.11.001.
- 966 McElreath, Richard. 2015. *Statistical rethinking: A bayesian course with examples*
967 *in R and Stan*. Boca Raton, FL: CRC Press.
- 968 Meyer, Ernst Alfred. 1904. Zur Vokaldauer im Deutschen. In *Nordiska studier*
969 *tillegnade A. Noreen*, 347–356. Uppsala: K.W. Appelbergs Boktryckeri.
- 970 Mortensen, Johannes & John Tøndering. 2013. The effect of vowel height on
971 Voice Onset Time in stop consonants in CV sequences in spontaneous Danish.
972 In *Proceedings of Fonetik 2013*, Linköping, Sweden: Linköping University.
- 973 Munafò, Marcus R., Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button,
974 Christopher D. Chambers, Nathalie Percie Du Sert, Uri Simonsohn, Eric-Jan
975 Wagenmakers, Jennifer J. Ware & John P. A. Ioannidis. 2017. A manifesto
976 for reproducible science. *Nature Human Behaviour* 1(1). 0021. doi:10.1038/
977 s41562-016-0021.
- 978 Nicenboim, Bruno, Timo B. Roettger & Shravan Vasishth. 2018. Using meta-
979 analysis for evidence synthesis: The case of incomplete neutralization in Ger-
980 man. *Journal of Phonetics* 70. 39–55. doi:10.1016/j.wocn.2018.06.001.
- 981 Nooteboom, Sieb G. & Gert J. N. Doodeman. 1980. Production and perception
982 of vowel length in spoken sentences. *The Journal of the Acoustical Society of*
983 *America* 67(1). 276–287. doi:10.1121/1.383737.
- 984 O'Dell, Michael L. & Tommi Nieminen. 2008. Coupled oscillator model for
985 speech timing: Overview and examples. In *Nordic prosody: Proceedings of*
986 *the Xth conference*, 179–190.
- 987 Ohala, John J., Catherine P. Browman & Louis M. Goldstein. 1986. To-
988 wards an articulatory phonology. *Phonology* 3. 219–252. doi:10.1017/
989 S0952675700000658.

- 990 Ohala, John J & Bertil Lyberg. 1976. Comments on “temporal interactions within a
991 phrase and sentence context”. *The Journal of the Acoustical Society of America*
992 59(4). 990–992. doi:10.1121/1.380926.
- 993 Öhman, Sven E. G. 1966. Coarticulation in VCV utterances: Spectrographic mea-
994 surements. *The Journal of the Acoustical Society of America* 39(1). 151–168.
995 doi:10.1121/1.1909864.
- 996 Öhman, Sven E. G. 1967. Numerical model of coarticulation. *The Journal of the*
997 *Acoustical Society of America* 41(2). 310–320. doi:10.1121/1.1910340.
- 998 Pape, Daniel & Luis M. T. Jesus. 2014. Production and perception of velar stop
999 (de)voicing in European Portuguese and Italian. *EURASIP Journal on Audio,*
1000 *Speech, and Music Processing* 2014(1). 6. [http://asmp.eurasipjournals.](http://asmp.eurasipjournals.com/content/2014/1/6)
1001 [com/content/2014/1/6](http://asmp.eurasipjournals.com/content/2014/1/6).
- 1002 Peterson, Gordon E. & Ilse Lehiste. 1960. Duration of syllable nuclei in English.
1003 *The Journal of the Acoustical Society of America* 32(6). 693–703. doi:10.1121/
1004 1.1908183.
- 1005 Plug, Leendert & Rachel Smith. 2018. Segments, syllables and speech tempo per-
1006 ception. In *Proceedings of the 9th International Conference on Speech Prosody*
1007 *2018*, 279–283. doi:10.21437/SpeechProsody.2018-57.
- 1008 Port, Robert F. & Jonathan Dalby. 1982. Consonant/vowel ratio as a cue for voicing
1009 in English. *Perception & Psychophysics* 32(2). 141–152.
- 1010 R Core Team. 2019. R: A language and environment for statistical computing.
1011 <https://www.R-project.org/>.
- 1012 Raphael, Lawrence J. 1972. Preceding vowel duration as a cue to the perception of
1013 the voicing characteristic of word final consonants in American English. *The*
1014 *Journal of the Acoustical Society of America* 51(4B). 1296–1303. doi:10.1121/
1015 1.1912974.
- 1016 Ratnikova, E. I. 2017. Towards a log-normal model of phonation units lengths
1017 distribution in the oral utterances. *International Research Journal* 3(57). 46–49.
1018 doi:10.23670/IRJ.2017.57.103.
- 1019 Roettger, Timo B. 2019. Researcher degrees of freedom in phonetic sciences. *Lab-*
1020 *oratory Phonology: Journal of the Association for Laboratory Phonology* 10(1).
1021 1–27. doi:10.5334/labphon.147.

- 1022 Rosen, Kristin M. 2005. Analysis of speech segment duration with the lognormal
1023 distribution: A basis for unification and comparison. *Journal of Phonetics* 33(4).
1024 411–426. doi:10.1016/j.wocn.2005.02.001.
- 1025 Saltzman, Elliot, Hosung Nam, Jelena Krivokapic & Louis Goldstein. 2008. A
1026 task-dynamic toolkit for modeling the effects of prosodic structure on articula-
1027 tion. In *Proceedings of the 4th International Conference on Speech Prosody*,
1028 175–184.
- 1029 Sharf, Donald J. 1962. Duration of post-stress intervocalic stops and preceding
1030 vowels. *Language and Speech* 5(1). 26–30.
- 1031 Sharf, Donald J. 1964. Vowel duration in whispered and in normal speech. *Lan-*
1032 *guage and Speech* 7(2). 89–97.
- 1033 Slis, Iman Hans & Antonie Cohen. 1969a. On the complex regulating the
1034 voiced-voiceless distinction I. *Language and Speech* 12(2). 80–102. doi:
1035 10.1177/002383096901200202.
- 1036 Slis, Iman Hans & Antonie Cohen. 1969b. On the complex regulating the voiced-
1037 voiceless distinction II. *Language and Speech* 12(3). 137–155. doi:10.1177/
1038 002383096901200301.
- 1039 Stevens, Kenneth N. & Samuel Jay Keyser. 1989. Primary features and their
1040 enhancement in consonants. *Language* 81–106. [https://www.jstor.org/
1041 stable/414843](https://www.jstor.org/stable/414843).
- 1042 Summers, W. Van. 1987. Effects of stress and final-consonant voicing on vowel
1043 production: Articulatory and acoustic analyses. *The Journal of the Acoustical*
1044 *Society of America* 82(3). 847–863. doi:10.1121/1.395284.
- 1045 Toivonen, Ida, Lev Blumenfeld, Andrea Gormley, Leah Hoiting, John Logan,
1046 Nalini Ramlakhan & Adam Stone. 2015. Vowel height and duration. In Ulrike
1047 Steindl, Thomas Borer, Huilin Fang, Alfredo García Pardo, Peter Guekguezian,
1048 Brian Hsu, Charlie O’Hara & Iris Chuoying Ouyang (eds.), *Proceedings of the*
1049 *32nd West Coast Conference on Formal Linguistics*, vol. 32, 64–71. Somerville,
1050 MA: Cascadilla Proceedings Project.
- 1051 Van Heuven, W. J. B., P. Mander, E. Keuleers & M. Brysbaert. 2014. Subtlex-UK:
1052 A new and improved word frequency database for British English. *Quarterly*
1053 *Journal of Experimental Psychology* 67. 1176–1190.
- 1054 Vasishth, Shravan, M. Beckman, B. Nicenboim, Fangfang Li & Eun Jong Kong.
1055 2018a. Bayesian data analysis in the phonetic sciences: A tutorial introduction.
1056 *Journal of Phonetics* 71. 147–161. doi:10.1016/j.wocn.2018.07.008.

- 1057 Vasishth, Shravan, Daniela Mertzen, Lena A. Jäger & Andrew Gelman. 2018b.
1058 The statistical significance filter leads to overoptimistic expectations of repli-
1059 cability. *Journal of Memory and Language* 103. 151–175. doi:10.1016/j.jml.
1060 2018.07.004.
- 1061 Wagenmakers, Eric-Jan. 2007. A practical solution to the pervasive problems
1062 of *p* values. *Psychonomic Bulletin & Review* 14(5). 779–804. doi:10.3758/
1063 BF03194105.
- 1064 Warren, Paul & Jen Hay. 2006. Using sound change to explore the mental lexicon.
1065 In M. Claire Fletcher-Flinn & G. M. Haberman (eds.), *Cognition and language:
1066 Perspectives from New Zealand*, chap. 8, 105–126. Brisbane, QLD, AUS: Aus-
1067 tralian Academic Press.
- 1068 Warren, Willis & Adam Jacks. 2005. Lip and jaw closing gesture durations in
1069 syllable final voiced and voiceless stops. *The Journal of the Acoustical Society
1070 of America* 117(4). 2618–2618. doi:10.1121/1.4778168.
- 1071 Zeroual, Chakir, Philip Hoole, Adamantios I. Gafos & John H. Esling. 2015. Ges-
1072 tural coordination differences between intervocalic simple and geminate plo-
1073 sives in Moroccan Arabic: An EMA investigation. In *Proceedings of ICPHS*,
1074 1–5.
- 1075 Zmarich, Claudio, Barbara Gili Fivela, Pascal Perrier, Christophe Savariaux &
1076 Graziano Tisato. 2011. Speech timing organization for the phonological length
1077 contrast in Italian consonants. In *Twelfth annual conference of the international
1078 speech communication association*, 401–404.