

1 Temporal (in)stability in English monosyllabic
2 and disyllabic words: Insights on the effect of
3 voicing on vowel duration

4 Stefano Coretta

5 **Abstract**

6 It will come.

7 **1 Introduction**

8 A well-known cross-linguistic tendency is that vowels have shorter durations when
9 followed by voiceless stops and longer durations when followed by voiced stops.
10 This so-called ‘voicing effect’ has been long documented in a wide range of lan-
11 guages across different linguistic families (Maddieson & Gandour 1976; Beguš
12 2017). Several hypotheses have been proposed as to the origin of this phenomenon,
13 from articulatory mechanisms to perceptual biases; however, no one particular ac-
14 count has gained universal support.

15 One such hypothesis, the compensatory temporal adjustment account, states
16 that the voicing effect involves a compensatory mechanism between vowel and
17 consonant closure duration. Vowels are shorter when followed by voiceless stops
18 because the latter have longer closure durations, and, vice versa, vowels are longer
19 before voiced stops because the latter have shorter closure durations. However, the
20 compensatory account fails to clearly identify a speech interval within which com-
21 pensation is implemented. Both the syllable (Lindblom 1967; Farnetani & Kori
22 1986) and the word (Slis & Cohen 1969a,b; Lehiste 1970a,b) have been proposed
23 as such intervals, but these have been subsequently criticised on empirical and
24 logical grounds (Chen 1970; Jacewicz et al. 2009; Maddieson & Gandour 1976;
25 Coretta 2018).

26 In an exploratory study of acoustic durations in Italian and Polish trochaic
27 CVCV words, Coretta (2018) finds that the duration of the interval between the
28 two consonant releases is not affected by the voicing status of the second conso-
29 nant. The duration of the release-to-release interval in words where the second

30 consonant is voiceless (like /pata/) is not significantly different from that in words
31 where the second consonant is voiced (for example, /pada/). The temporal stabil-
32 ity of the release-to-release interval is compatible with a compensatory temporal
33 adjustment account of the voicing effect (Lindblom 1967; Slis & Cohen 1969a,b;
34 Lehiste 1970a,b), and it offers a resolution to the drawbacks of previous versions
35 of the account.

36 Given the temporal stability of the release-to-release interval, the timing of
37 the vowel/consonant (VC) boundary (corresponding to the vowel offset and the
38 consonant closure onset) within that interval will determine the respective dura-
39 tions of vowel and consonant closure. If the VC boundary is timed earlier than
40 50% of the release-to-release, the resulting vowel duration will be shorter than that
41 of the closure duration. Vice versa, a timing of the VC boundary later than 50%
42 of the release-to-release results in a longer vowel and a shorter closure. The out-
43 come is that shorter vowels are followed by longer stops, and longer vowels are
44 followed by shorter stops. This agrees with the known differences of closure du-
45 rations in voiceless vs. voiced stops (Lisker 1957; Van Summers 1987; Davis &
46 Van Summers 1989; de Jong 1991). Thus, a possible diachronic pathway to the
47 voicing effect in disyllabic words is one in which vowel and closure duration dif-
48 ferences emerge from changes in the timing of the VC boundary within the release-
49 to-release interval which affect the voiceless and voiced contexts differently.

50 Note that the release-to-release interval in itself does not have a special status.
51 The proposed account of compensatory temporal adjustment can be understood
52 in relation to the acoustic duration of vowels, hence the scope of compensation
53 can (but need not) be defined in terms of acoustic intervals. The interval found to
54 be temporally stable across voicing contexts in disyllabic words is the release-to-
55 release interval. However, it is desirable to derive the isochrony of this acoustic
56 interval from properties of articulatory coordination. A tentative account of the un-
57 derlying gestural coordination from which the release-to-release isochrony could
58 be derived is offered here.

59 According to Öhman (1966, 1967), the speech stream is composed by a series
60 of continuous vocalic gestures interrupted by gestures of oral constriction (con-
61 sonants). Fowler (1983) further proposes that the vocalic gestures of a VCV se-
62 quence are characterised by a cyclic pattern of production, so that the temporal
63 distance between the two vowels is constant, independent of the nature of the in-
64 tervening consonant. While the temporal distance of the V-to-V interval is modu-
65 lated by the number of intervening consonants (Zmarich et al. 2011; Zeroual et al.
66 2015), the distance can still be expected to be stable within the context of disyllabic
67 words with a single intervocalic consonant that alternates in voicing.

68 The task-dynamic model (Saltzman et al. 2008) of Articulatory Phonology
69 (Ohala et al. 1986; Browman & Goldstein 1988, 1992), based on the coupled os-
70 cillators 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. Relevant to our discussion is that onset consonants are generally produced in-phase with the following vowel, meaning that the vocalic and consonantal gestures are initiated together. This mechanism 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).

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. 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 this scenario, the combined action of the isochrony of the vowel-to-vowel interval and the in-phase alignment of the onset consonant is also responsible for the isochrony of the release-to-release interval in CVCV words. Van Summers (1987) shows that the closing gesture of voiceless stops has greater velocity than that of voiced stops. Assuming that the closing gesture of both voiceless and voiced stops is initiated in synchrony with that of the following vowel (as per the in-phase alignment), full oral closure will be achieved earlier in voiceless than in voiced stops relative to the beginning of the preceding vocalic gesture, while the timing of the consonant release will not be affected, in accordance with the empirical data.

1.1 The voicing effect in English

English is one of the most investigated language in relation to the voicing effect (Meyer 1904; Heffner 1937; House & Fairbanks 1953; Belasco 1953; Peterson & Lehiste 1960; Halle & Stevens 1967; Chen 1970; Klatt 1973; Lisker 1974; Laeuffer 1992; Fowler 1992; Hussein 1994; Lampp & Reklis 2004; Warren & Jacks 2005; Durvasula & Luo 2012; Ko 2018). English is also the language in which the voicing effect has the greatest magnitude relative to that of other languages. This special status of English is traditionally attributed to the phonologisation of the voicing effect in this language (Sharf 1964; de Jong 2004). Vowel duration and

the vowel-to-consonant duration ratio are considered to be among the most stable cues to consonantal voicing (Peterson & Lehiste 1960; Raphael 1972; Port & Dalby 1982). Kluender et al. (1988) proposed that the difference in vowel duration before voiceless vs. voiced stops could have been enhanced and exploited to cue the voicing contrast. This could explain the greater effect of English compared for example to the effect in Italian, in which voicing is most robustly cued by vocal fold vibration during closure (Pape & Jesus 2014).

Indeed, previous studies on English report 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 (Caldognetto et al. 1979; Farnetani & Kori 1986; Esposito 2002; Coretta 2018). A Bayesian meta-analysis of the voicing effect (see Supplement A) returned a 95% credible interval for the effect of voicing in English monosyllabic words between 55 and 95 ms, with a meta-analytical mean of 75 ms. In other words, we can be 95% confident that the effect is between 55-95 ms. 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 also that the Italian values refer to the effect as observed in disyllabic words.

However, it is possible that the alleged differences in magnitude between English and other languages are a product of the different contexts under examination (Laeufer 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 (see Supplement A) further suggests a potential for publication bias, 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.

1.2 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. 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-

149 to-release temporal stability in disyllabic words is the presence of a direct relation
150 between the two vowels in these words. Since monosyllabic words don't have a
151 second vowel, there is no direct vowel-to-vowel relation to derive the release-to-
152 release stability from.

153 Furthermore, Jacewicz et al. (2009) report that, in American English, monosyl-
154 labic words are longer when the second consonant is voiced. Based on this find-
155 ing, it is expected that the release-to-release duration should be longer when C2 is
156 voiced. Jacewicz et al. (2009) attribute the difference in monosyllabic word dura-
157 tion to the difference in vowel duration before voiceless vs. voiced stops. Thus, we
158 can expect the magnitude of the difference in release-to-release duration in mono-
159 syllabic words to be close to the difference in vowel duration. This hypothesis also
160 fits with the reported greater effect of voicing on vowel duration in monosyllabic
161 than disyllabic words.

162 The data in Coretta (2018) suggests that the intrinsic duration of vowels and
163 consonants can contribute to the duration of the release-to-release interval. In par-
164 ticular, release-to-release intervals containing a high vowel have shorter durations
165 than those with a low vowel. This is not surprising, given the well-known ten-
166 dency of high vowels to be shorter than low vowels (Hertrich & Ackermann 1997;
167 Esposito 2002; Mortensen & Tøndering 2013; Toivonen et al. 2015; Kawahara
168 et al. 2017). As for the consonantal place of articulation, the release-to-release is
169 shorter in Italian and Polish when the second consonant is velar compared to when
170 it is coronal. This could be a consequence of the fact that the closure of velar stops
171 is shorter than that of other stops. For example, Sharf's (1962) data on closure
172 duration in English suggests that the closure of labial stops (60-90 ms) is about
173 10 ms longer than that of velar stops (55-75 ms). It can be expected that release-
174 to-release intervals with a velar stop in English will be about 10 ms shorter than
175 intervals with a labial stop.

176 Another set of objectives concerns the effect of voicing on vowel and closure
177 durations. A conceptual replication of previous studies' effect sizes is sought, with
178 special attention to differences between monosyllabic and disyllabic words. Only a
179 few studies directly compare the effect in different syllabic positions (for example,
180 Sharf (1962) and Klatt (1973)). The reported effects are in the range of 50-55 ms
181 in word-final (closed-syllable) position and 20-25 in word-medial (open-syllable)
182 position. The Bayesian meta-analysis of the voicing effect indicates a mean dif-
183 ference of 50 ms (75 ms in word-final position vs. 25 ms word-medially).

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

- 186 1. Is the duration of the interval between two consecutive stop releases (the
187 release-to-release interval) in monosyllabic and disyllabic words affected
188 by the voicing of C2 in English?

- 189 • H1a: The duration of the release-to-release interval is not affected by C2
190 voicing in disyllabic words.
- 191 • H1b: The release-to-release interval is longer in monosyllabic words with a
192 voiced C2 than in monosyllabic words with a voiceless C2.
- 193 2. Is the duration of the release-to-release interval affected by (a) the number
194 of syllables of the word, (b) the quality of V1, and (c) the place of C2?
- 195 • H2a: The release-to-release interval is longer in monosyllabic than in disyll-
196 labic words.
- 197 • H2b: The duration of the release-to-release interval decreases according to
198 the hierarchy /ɑ:/, /ɜ:/, /i:/.
- 199 • H2c: The release-to-release interval is longer when C2 is labial.
- 200 3. What is the estimated difference in the effect of voicing on vowel and stop
201 closure duration between monosyllabic and disyllabic words?
- 202 • H3: The effect of voicing on vowel duration is greater in monosyllabic than
203 in disyllabic words (no specific hypothesis in relation to closure duration).

204 **2 Methods**

205 The following subsections describe the experimental and statistical methods of
206 this study. The research design and data analyses were pre-registered on the Open
207 Science Framework prior to data collection (https://osf.io/hwr94/?view_only=d994915422144efaae4a5915237cb386). The research compendium
208 of this paper with data and analysis scripts is also available on the Open Science
209 Framework. Choices on experimental design and analysis were made within the
210 Bayesian framework of statistical inference (see Section 2.1 and Section 2.7 for
211 details).
212

213 **2.1 Sample size and stopping rule**

214 Sample size and a stopping rule were decided prior to data collection with a
215 Bayesian method of sample determination based on the Region Of Practical
216 Equivalence (ROPE, Kruschke 2015; Vasishth et al. 2018b). A ‘no-effect’ region
217 of values around 0 is first identified. This null region (the ROPE) can be thought
218 of as a Bayesian 95% credible interval of a distribution, the values within which
219 can be interpreted as a negligible or null effect. For this study, a ROPE between
220 -10 and +10 ms has been chosen. The width of 20 ms is based on the estimates
221 of the just noticeable difference in Huggins (1972) and Nooteboom & Doodeman

222 (1980). Differences in release-to-release durations below 10 ms (either positive
223 or negative) will be interpreted as compatible with a null effect.

224 Once a ROPE width is set, the goal is to collect data during sequential testing
225 until the width of the 95% credible interval (CI) of the tested effect is equal to
226 or less than the ROPE width (in this study, 20 ms). In other words, the objective
227 is to reach estimate precision, rather than significance (as in frequentist null hy-
228 pothesis testing). Inference can then be made based on the credible interval of the
229 sought effect. When the precision goal is reached (the CI width is equal or lower
230 than the ROPE width), three possible scenarios can arise: (1) the CI of the effect
231 completely overlaps with the ROPE around 0, in which case the data supports a
232 practically equivalent null effect; (2) the CI of the effect completely lies outside
233 the ROPE, which indicates that the data support the effect to be within that CI; (3)
234 the CI partially overlaps with the ROPE, in which case no decision can be made on
235 whether the data support one hypothesis over the other, although it still possible
236 to infer the sign of the effect (if the CI partially overlaps with the right side of the
237 ROPE without including 0, there is evidence for a positive effect, while if the CI
238 overlaps with the left side of the ROPE without including 0, there is evidence for
239 a negative effect).

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

244 2.2 Participants

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

250 2.3 Equipment

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

Table 1: Test $C_1\hat{V}_1C_2$ (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

2.4 Materials

The test words were $C_1\hat{V}_1C_2$ (VC) words, where $C_1 = /t/, V_1 = /i:, ɜ:, ɑ:/, C_2 = /p, b, k, g/,$ 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.¹ 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 Friday, They'll say X this Tuesday*. Each word + frame combination was included once in the stimuli list, so that each speaker read a total of 120 sentence stimuli (24 words \times 5 frames). A total of 1800 observations were recorded (120 stimuli \times 15 speakers).

2.5 Procedure

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

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

2.6 Data processing and measurements

A forced-aligned transcription was obtained with the SPeech Phonetisation Alignment and Syllabification software (SPPAS, Bigi 2015). The automatic annotation was corrected by the author according to the principles of phonetic segmentation detailed in Machač & Skarnitzl (2009). A custom Praat script was written to automatically detect the burst onset of the consonants in the test words, using the algorithm in Ananthapadmanabha et al. (2014). The output was checked and manually corrected by the author when necessary.

The following measures were obtained via a custom Praat script:

- Duration of the release-to-release interval: from the release of C1 to the release of C2.
- V1 duration: from appearance to disappearance of higher formant structure in the spectrogram in correspondence of V1 (Machač & Skarnitzl 2009).
- C2 closure duration: from disappearance of higher formant structure in the V1C2 sequence to the release of C2 (Machač & Skarnitzl 2009).
- Speech rate: calculated as the number of syllables per second (number of syllables in the sentence divided by the sentence duration in seconds, Plug & 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).² For an introduction to Bayesian statistics in phonetics, see

²I 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.

315 Vasishth et al. (2018a), and Nicenboim et al. (2018), while for a general intro-
316 duction see Etz et al. (2018), McElreath (2015), Kruschke (2015), and references
317 therein. While a thorough discussion of Bayesian methods would be beyond the
318 scope of this paper, it is relevant to provide the less familiar reader with the basic
319 tools for interpreting analyses and results.

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

335 In each model, priors are specified for each of the parameters to be estimated.
336 The priors are in the form of particular distributions, like the Gaussian (normal) or
337 the Cauchy distribution. A prior defines the prior knowledge of where the param-
338 eter might lie within a range of values. For example, a prior as a normal distribution
339 with mean 200 ms and standard deviation 50 indicates the researcher's belief that
340 the parameter lies between 100 and 300 ms with 95% probability (i.e., the mean
341 minus twice the standard deviation, and the mean plus twice the standard devia-
342 tion).

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

354 A concern could be raised that the priors might have greater influence on the
355 posterior distributions than the observed data. A sensitivity analysis based on pos-

356 terior z-scores and shrinkage (Betancourt 2018) indicates that the models discussed
357 in this study are highly informed by the observed data and don't heavily rely on
358 prior specifications.

359 **3 Results**

360 This section reports the results of the Bayesian models, grouped by outcome vari-
361 able (release-to-release, vowel duration, closure duration). A description of the
362 model structure and priors is given for each model, followed by the presentation
363 of the posterior distributions of the relevant terms. The data and R code used for
364 analysis are available as part of the paper's research compendium (Coretta 2019a).
365 Each model is assigned a number (1 to 5), and the text refers to these.

366 Model convergence was reached in all the reported models ($\hat{R} = 1$) and no
367 major divergences in the MCMC chains were observed. The posterior predictive
368 check plots indicate that the observed distributions are slightly positively skewed
369 so that a log-normal distribution would have been more appropriate. Previous
370 work has shown that speech-units duration does follow, as a general trend, a log-
371 normal distribution (Rosen 2005; Ratnikova 2017). However, the deviations from
372 a Gaussian distribution are minimal, and an informal comparison of one of the
373 models fitted with a log-normal distribution led to virtually identical results.

374 **3.1 Release-to-release duration**

375 A Bayesian regression was fit to model the duration of the release-to-release in-
376 terval (model 1). The following terms were included as fixed effects: C2 voicing
377 (voiceless, voiced), number of syllables (disyllabic, monosyllabic), centred speech
378 rate, an interaction between C2 voicing and number of syllables. A by-speaker
379 and by-word random intercept, and a by-speaker random coefficient for C2 voic-
380 ing were entered as random effects. The following priors were used. Two weakly
381 informative priors based on the results from Coretta (2018) were chosen for the
382 intercept and the effect of C2 voicing. The former prior is a normal distribution
383 with mean 200 ms and SD = 50, while the latter a normal distribution with mean 0
384 ms and SD = 25. A weakly informative prior as a normal distribution with mean
385 50 ms and SD = 25 was specified for the effect of number of syllables. The prior
386 is based on differences in vowel duration between mono- vs. disyllabic words,
387 which range between 30 and 100 ms (Sharf 1962; Klatt 1973). The same prior
388 was used for the interaction between C2 voicing and number of syllables, based
389 on the reported differences in voicing effect in mono- vs. disyllabic words (Sharf
390 1962; Klatt 1973). The prior for the effect of centred speech rate is a normal distri-
391 bution with mean -25 ms and SD = 10, and is based on results from Coretta (2018).

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

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 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

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

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 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]).

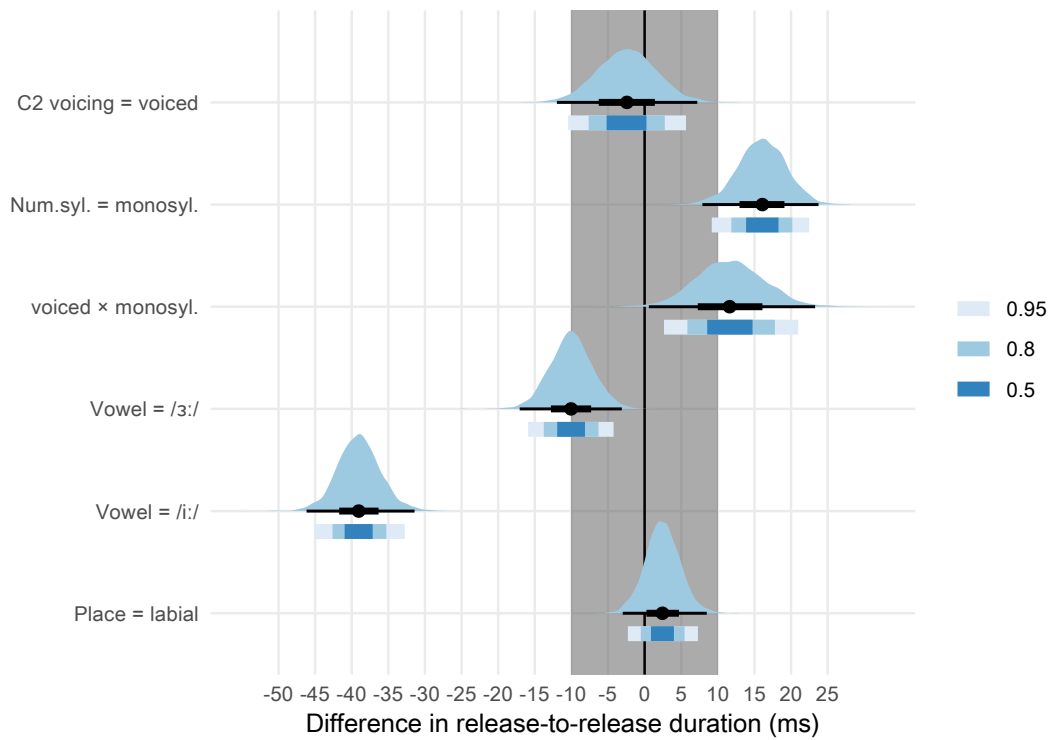


Figure 1: Posterior distributions and Bayesian credible intervals of the effects on release-to-release duration (model 3). For each effect, the thick blue-coloured bars indicate (from darker to lighter) the 50%, 80%, and 95% CI. The black point with bars are the posterior median (the point), the 98%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

Table 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

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 (/ɑ:/, /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).

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 & Van 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 ($\theta = 124.91$ 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

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

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

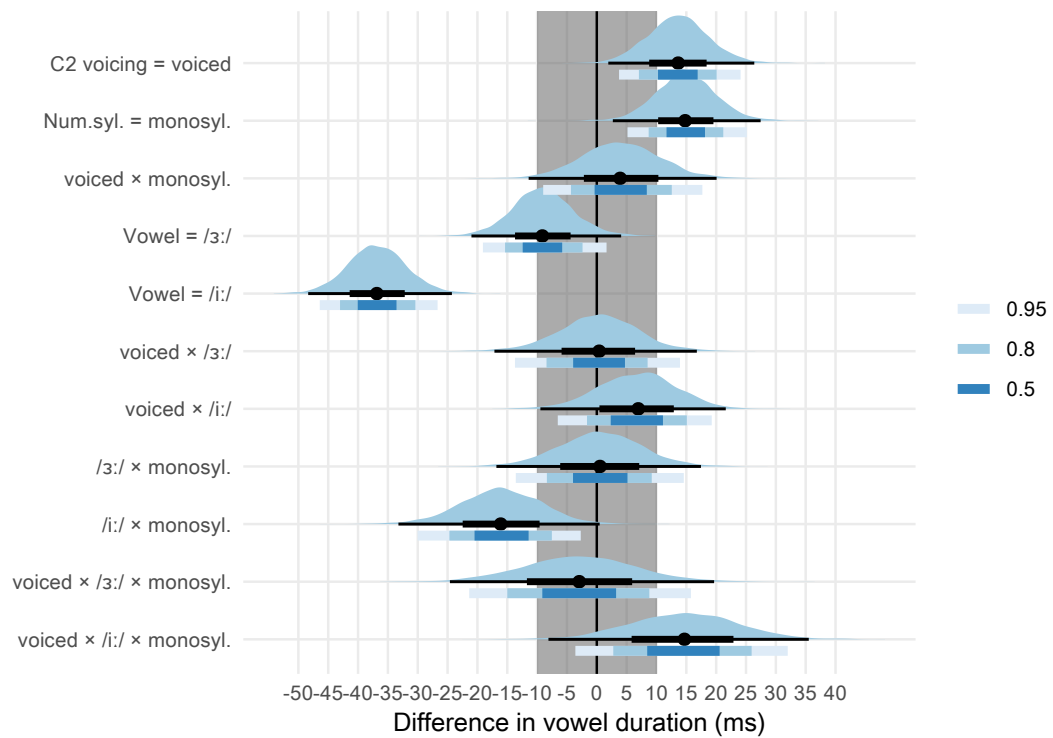


Figure 2: Posterior distributions and Bayesian credible intervals of the effects on vowel duration (model 4). For each effect, the thick blue-coloured bars indicate (from darker to lighter) the 50%, 80%, and 95% CI. The black point with bars are the posterior median (the point), the 98% (thin bar) and 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 \times labial	1.37	3.94	-6.79	8.93	15.72
voiced \times monosyll.	1.82	4.06	-6.08	9.70	15.78
labial \times monosyll.	-0.74	4.02	-8.95	6.88	15.83
voiced \times labial \times monosyll.	6.41	5.66	-4.72	17.45	22.17

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) is suggested by the 95% CI from -0.03 to 10.76 ms (θ = 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.

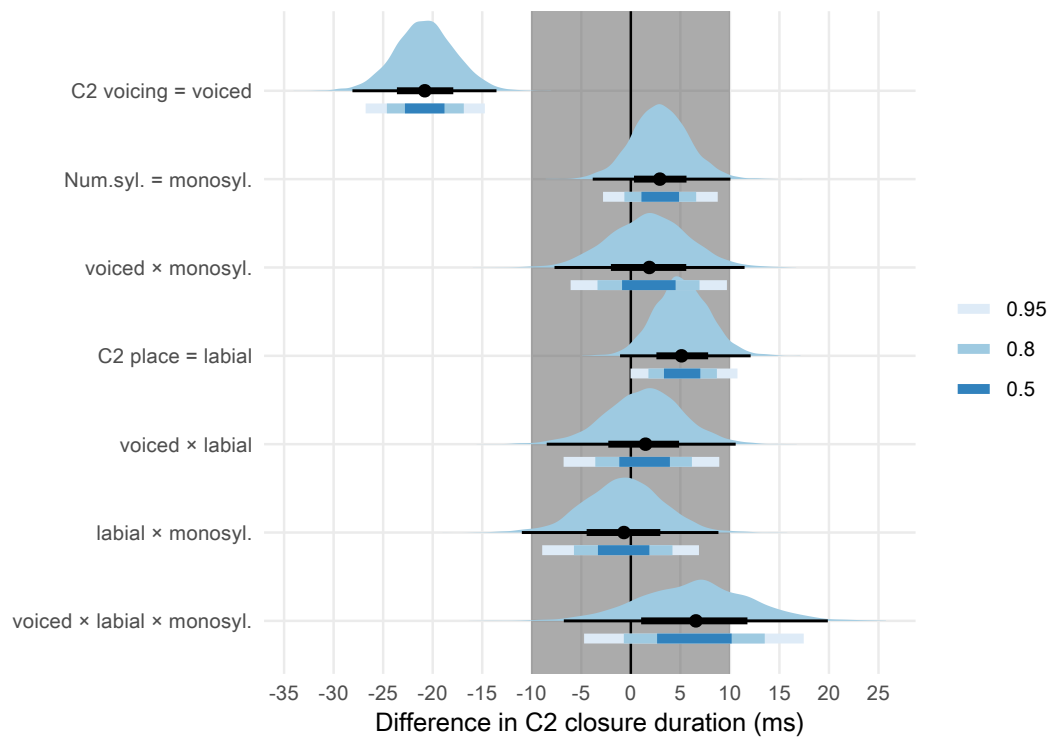


Figure 3: Posterior distributions and Bayesian credible intervals of the effects on closure duration (model 4). For each effect, the thick blue-coloured bars indicate (from darker to lighter) the 50%, 80%, and 95% CI. The black point with bars are the posterior median (the point), the 98% (thin bar)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

4 Discussion

This study set out to build on the results discussed in Coretta (2018) 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 1.2) asked whether the voicing of C2 in disyllabic and monosyllabic words in English influences the duration of the release-to-release interval. Coretta (2018) 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/).

A Bayesian regression model was fitted to the release-to-release duration (model 3). The results 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

568 bias could explain the quite small negative effect of voicing on the interval dura-
569 tion, and why it is in the opposite direction of the one predicted for monosyllabic
570 words (i.e. *longer* release-to-release when C2 is voiced).

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

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

596 4.2 Vowel and closure duration

597 Question 3 addressed the effect of voicing on vowel and closure duration, and the
598 possible differences between disyllabic and monosyllabic words. The effect of
599 voicing on vowel duration found in this study was estimated to lie between 4 and
600 25 ms. This range of values is very similar to that reported in Coretta (2018) for
601 Italian and Polish disyllabic words (the 95% confidence interval for the effect in
602 these languages is [8, 25]), monosyllabic words were not tested). When compared
603 to the values in previous studies that investigated disyllabic words (Sharf 1962;
604 Klatt 1973; Davis & Van Summers 1989), the effect size found in this study tends
605 towards smaller values. However, note that the posterior distribution of the effect
606 in the current study is entirely contained in the meta-analytical posterior distribu-

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

It was expected that the voicing effect on vowels would be stronger in monosyllabic than in disyllabic words (hypothesis 3). The credible intervals of the posterior distributions from model 4, which are larger than the ROPE, make interpretation less straightforward. At 80% probability, the difference in voicing effect between mono- and disyllabic words is between -5 and $+12.5$ ms. The distribution is skewed towards the positive side, and this is compatible with results from previous studies, although the CI includes 0. The magnitude, however, is considerably lower than what previously reported. More data is needed to reach a sensible estimate precision and reduce uncertainty.

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

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

As a general trend, the differences in vowel and closure duration found in this study are smaller than those known from the literature, and considerably so in the case of vowels. A possible reason for this discrepancy could be found in problems

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 (2018) 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.³ 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 2019b).

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.

³Note that isochrony here is intended as pertaining the context of voiceless vs. voiced stops only.

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

5 Conclusion

This paper set out to investigate temporal properties of the so-called ‘voicing effect’, by which vowels are shorter when followed by voiceless stops and longer when followed by voiced stops. Coretta (2018) proposes that the voicing effect emerges via a mechanism of relative timing of the VC boundary within a temporally stable interval. Such interval was argued to be the interval between two consecutive releases, as evidenced by acoustic data from Italian and Polish disyllabic words. The temporal stability of the release-to-release in relation to consonantal voicing is thought to derive from two properties of gestural phasing, namely the isochrony of the distance between the vowels in a VCV sequence, and in-phase alignment of onset consonants and the following vowel. On the other hand, the lack of an articulatory anchor (a second vowel) in monosyllabic words would allow the release-to-release duration to be affected by C2 voicing and differ in the monosyllabic context.

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

722 reported for monosyllabic words. The data in this study, and the cumulative evi-
723 dence from previous studies as evinced by a Bayesian meta-analysis, however, do
724 not equivocally provide support for a difference in the effect between mono- and
725 disyllabic words, and future work is necessary to shed light on the matter.

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

737 References

- 738 Ananthapadmanabha, T. V., A. P. Prathosh & A. G. Ramakrishnan. 2014. Detec-
739 tion of the closure-burst transitions of stops and affricates in continuous speech
740 using the plosion index. *The Journal of the Acoustical Society of America* 135(1).
741 460–471. doi:10.1121/1.4836055.
- 742 Beguš, Gašper. 2017. Effects of ejective stops on preceding vowel duration. *The*
743 *Journal of the Acoustical Society of America* 142(4). 2168–2184. doi:10.1121/
744 1.5007728.
- 745 Belasco, Simon. 1953. The influence of force of articulation of consonants on
746 vowel duration. *The Journal of the Acoustical Society of America* 25(5). 1015–
747 1016.
- 748 Betancourt, Michael. 2018. Calibrating model-based inferences and decisions.
749 arXiv preprint arXiv:1803.08393.
- 750 Bigi, Brigitte. 2015. SPPAS - Multi-lingual approaches to the automatic annotation
751 of speech. *The Phonetician* 111–112. 54–69.
- 752 Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*.
753 Cambridge University Press.
- 754 Blevins, Juliette. 2006. A theoretical synopsis of Evolutionary Phonology. *Theo-*
755 *retical linguistics* 32(2). 117–166.

- 756 Browman, Catherine P. & Louis Goldstein. 1988. Some notes on syllable structure
757 in articulatory phonology. *Phonetica* 45(2-4). 140–155.
- 758 Browman, Catherine P. & Louis Goldstein. 1992. Articulatory phonology: An
759 overview. *Phonetica* 49. 155–180.
- 760 Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel mod-
761 els using Stan. *Journal of Statistical Software* 80(1). 1–28. doi:10.18637/jss.v
762 080.i01.
- 763 Bürkner, Paul-Christian. 2018. Advanced Bayesian multilevel modeling with the
764 r package brms. *The R Journal* 10(1). 395–411. doi:10.32614/RJ-2018-017.
- 765 Caldognetto, Emanuela Magno, Franco Ferrero, Kyriaki Vagges & Maria Bagno.
766 1979. Indici acustici e indici percettivi nel riconoscimento dei suoni linguis-
767 tici (con applicazione alle consonanti occlusive dell’italiano). *Acta Phoniatica*
768 *Latina* 2. 219–246.
- 769 Celata, Chiara & Paolo Mairano. 2014. On the timing of V-to-V intervals in Italian:
770 a review, and some new hypotheses. *Revista de Filología Románica* 31. 37. doi:
771 10.5209/rev_RFRM.2014.v31.n1.51022.
- 772 Chen, Matthew. 1970. Vowel length variation as a function of the voicing of the
773 consonant environment. *Phonetica* 22(3). 129–159.
- 774 Coretta, Stefano. 2018. An exploratory study of voicing-related differences in
775 vowel duration as compensatory temporal adjustment in Italian and Polish. Sub-
776 mitted.
- 777 Coretta, Stefano. 2019a. Compensatory aspects of the effect of voicing on vowel
778 duration in English [Research compendium]. Open Science Framework. doi:
779 10.17605/OSF.IO/M4RZY.
- 780 Coretta, Stefano. 2019b. Longer vowel duration correlates with greater tongue
781 root displacement: Acoustic and articulatory data from Italian and Polish.
782 Manuscript.
- 783 Davis, Stuart & W. Van Summers. 1989. Vowel length and closure duration in
784 word-medial VC sequences. *Journal of Phonetics* 17. 339–353.
- 785 Durvasula, Karthik & Qian Luo. 2012. Voicing, aspiration, and vowel duration in
786 Hindi. *Proceedings of Meetings on Acoustics* 18. 1–10. doi:10.1121/1.4895027.
- 787 Esposito, Anna. 2002. On vowel height and consonantal voicing effects: Data
788 from Italian. *Phonetica* 59(4). 197–231. doi:10.1159/000068347.

- 789 Etz, Alexander, Quentin F. Gronau, Fabian Dablander, Peter A. Edelsbrunner &
790 Beth Baribault. 2018. How to become a Bayesian in eight easy steps: An
791 annotated reading list. *Psychonomic Bulletin & Review* 25(1). 219–234. doi:
792 10.3758/s13423-017-1317-5.
- 793 Farnetani, Edda & Shiro Kori. 1986. Effects of syllable and word structure on
794 segmental durations in spoken Italian. *Speech communication* 5(1). 17–34. doi:
795 10.1016/0167-6393(86)90027-0.
- 796 Fowler, Carol A. 1983. Converging sources of evidence on spoken and perceived
797 rhythms of speech: Cyclic production of vowels in monosyllabic stress feet.
798 *Journal of Experimental Psychology: General* 112(3). 386. doi:10.1037/0096-
799 3445.112.3.386.
- 800 Fowler, Carol A. 1992. Vowel duration and closure duration in voiced and un-
801 voiced stops: There are no contrast effects here. *Journal of Phonetics* 20(1).
802 143–165.
- 803 Halle, Morris & Kenneth Stevens. 1967. Mechanism of glottal vibration for vowels
804 and consonants. *The Journal of the Acoustical Society of America* 41(6). 1613–
805 1613. doi:10.1121/1.2143736.
- 806 Heffner, R.-M.S. 1937. Notes on the length of vowels. *American Speech* 12. 128–
807 134. doi:10.2307/452621.
- 808 Hermes, Anne, Doris Mücke & Martine Grice. 2013. Gestural coordination of
809 Italian word-initial clusters: the case of ‘impure s’. *Phonology* 30(01). 1–25.
- 810 Hertrich, Ingo & Hermann Ackermann. 1997. Articulatory control of phonological
811 vowel length contrasts: Kinematic analysis of labial gestures. *The Journal of*
812 *the Acoustical Society of America* 102(1). 523–536. doi:10.1121/1.419725.
- 813 House, Arthur S. & Grant Fairbanks. 1953. The influence of consonant environ-
814 ment upon the secondary acoustical characteristics of vowels. *The Journal of*
815 *the Acoustical Society of America* 25(1). 105–113. doi:10.1121/1.1906982.
- 816 Huggins, A. William F. 1972. Just noticeable differences for segment duration
817 in natural speech. *The Journal of the Acoustical Society of America* 51(4B).
818 1270–1278. doi:10.1121/1.1912971.
- 819 Hussein, Lutfi. 1994. *Voicing-dependent vowel duration in Standard Arabic and its*
820 *acquisition by adult American students*: The Ohio State University dissertation.

- 821 Jacewicz, Ewa, Robert Allen Fox & Samantha Lyle. 2009. Variation in stop con-
822 sonant voicing in two regional varieties of American English. *Journal of the*
823 *International Phonetic Association* 39(3). 313–334. doi:10.1017/S002510030
824 9990156.
- 825 Javkin, Hector R. 1976. The perceptual basis of vowel duration differences associ-
826 ated with the voiced/voiceless distinction. *Report of the Phonology Laboratory,*
827 *UC Berkeley* 1. 78–92.
- 828 de Jong, Kenneth. 1991. An articulatory study of consonant-induced vowel dura-
829 tion changes in English. *Phonetica* 48(1). 1–17. doi:10.1121/1.2028316.
- 830 de Jong, Kenneth. 2004. Stress, lexical focus, and segmental focus in English:
831 patterns of variation in vowel duration. *Journal of Phonetics* 32(4). 493–516.
832 doi:10.1016/j.wocn.2004.05.002.
- 833 Kawahara, Shigeto, Donna Erickson & Atsuo Suemitsu. 2017. The phonetics of
834 jaw displacement in Japanese vowels. *Acoustical Science and Technology* 38(2).
835 99–107. doi:10.1250/ast.38.99.
- 836 Kirby, James & Morgan Sonderegger. 2018. Mixed-effects design analysis for
837 experimental phonetics. *Journal of Phonetics* 70. 70–85. doi:10.1016/j.wocn.2
838 018.05.005.
- 839 Klatt, Dennis H. 1973. Interaction between two factors that influence vowel du-
840 ration. *The Journal of the Acoustical Society of America* 54(4). 1102–1104.
841 doi:10.1121/1.1914322.
- 842 Kluender, Keith R., Randy L. Diehl & Beverly A. Wright. 1988. Vowel-length
843 differences before voiced and voiceless consonants: An auditory explanation.
844 *Journal of Phonetics* 16. 153–169.
- 845 Ko, Eon-Suk. 2018. Asymmetric effects of speaking rate on the vowel/consonant
846 ratio conditioned by coda voicing in English. *Phonetics and Speech Sciences*
847 10(2). 45–50. doi:10.13064/KSSS.2018.10.2.045.
- 848 Kruschke, John. 2015. *Doing Bayesian data analysis: A tutorial with R, JAGS,*
849 *and Stan (2nd edition)*. Amsterdam, The Netherlands: Academic Press.
- 850 Laeuffer, Christiane. 1992. Patterns of voicing-conditioned vowel duration in
851 French and English. *Journal of Phonetics* 20(4). 411–440.
- 852 Lampp, Claire & Heidi Reklis. 2004. Effects of coda voicing and aspiration on
853 Hindi vowels. *The Journal of the Acoustical Society of America* 115(5). 2540–
854 2540. doi:10.1121/1.4783577.

- 855 Lehiste, Ilse. 1970a. Temporal organization of higher-level linguistic units. *The*
856 *Journal of the Acoustical Society of America* 48(1A). 111–111. doi:10.1121/1.
857 1974906.
- 858 Lehiste, Ilse. 1970b. Temporal organization of spoken language. In *Working pa-*
859 *pers in linguistics*, vol. 4, 96–114. doi:10.1121/1.1974906.
- 860 Lindblom, Björn. 1967. Vowel duration and a model of lip mandible coordination.
861 *Speech Transmission Laboratory Quarterly Progress Status Report* 4. 1–29.
- 862 Lisker, Leigh. 1957. Closure duration and the intervocalic voiced-voiceless dis-
863 tinction in English. *Language* 33(1). 42–49. doi:10.2307/410949.
- 864 Lisker, Leigh. 1974. On “explaining” vowel duration variation. In *Proceedings of*
865 *the Linguistic Society of America*, 225–232.
- 866 Lisker, Leigh. 1986. “Voicing” in English: a catalogue of acoustic features sig-
867 naling /b/ versus /p/ in trochees. *Language and Speech* 29(1). 3–11. doi:
868 10.1177/002383098602900102.
- 869 Luce, Paul A & Jan Charles-Luce. 1985. Contextual effects on vowel duration, clo-
870 sure duration, and the consonant/vowel ratio in speech production. *The Journal*
871 *of the Acoustical Society of America* 78(6). 1949–1957. doi:10.1121/1.392651.
- 872 Machač, Pavel & Radek Skarnitzl. 2009. *Principles of phonetic segmentation*.
873 Epoque.
- 874 Maddieson, Ian & Jack Gandour. 1976. Vowel length before aspirated consonants.
875 In *UCLA Working papers in Phonetics*, vol. 31, 46–52.
- 876 Marin, Stefania & Marianne Pouplier. 2010. Temporal organization of complex
877 onsets and codas in American English: Testing the predictions of a gestural
878 coupling model. *Motor Control* 14(3). 380–407. doi:10.1123/mcj.14.3.380.
- 879 Marin, Stefania & Marianne Pouplier. 2014. Articulatory synergies in the temporal
880 organization of liquid clusters in Romanian. *Journal of Phonetics* 42. 24–36.
881 doi:10.1016/j.wocn.2013.11.001.
- 882 McElreath, Richard. 2015. *Statistical rethinking: A bayesian course with examples*
883 *in R and Stan*. CRC Press.
- 884 Meyer, Ernst Alfred. 1904. Zur vokaldauer im deutschen. In *Nordiska studier*
885 *tillegnade A. Noreen*, 347–356. K.W. Appelbergs Boktryckeri: Uppsala.

- 886 Mortensen, Johannes & John Tøndering. 2013. The effect of vowel height on
887 Voice Onset Time in stop consonants in CV sequences in spontaneous Danish.
888 In *Proceedings of Fonetik 2013*, Linköping, Sweden: Linköping University.
- 889 Munafò, Marcus R., Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button,
890 Christopher D. Chambers, Nathalie Percie Du Sert, Uri Simonsohn, Eric-Jan
891 Wagenmakers, Jennifer J. Ware & John P. A. Ioannidis. 2017. A manifesto for
892 reproducible science. *Nature Human Behaviour* 1(1). 0021. doi:10.1038/s415
893 62-016-0021.
- 894 Nicenboim, Bruno, Timo B. Roettger & Shravan Vasishth. 2018. Using meta-
895 analysis for evidence synthesis: The case of incomplete neutralization in ger-
896 man. *Journal of Phonetics* 70. 39–55. doi:10.1016/j.wocn.2018.06.001.
- 897 Nooteboom, Sieb G. & Gert J. N. Doodeman. 1980. Production and perception
898 of vowel length in spoken sentences. *The Journal of the Acoustical Society of*
899 *America* 67(1). 276–287. doi:10.1121/1.383737.
- 900 O'Dell, Michael L. & Tommi Nieminen. 2008. Coupled oscillator model for
901 speech timing: Overview and examples. In *Nordic prosody: Proceedings of*
902 *the Xth conference*, 179–190.
- 903 Ohala, John J, Catherine P Browman & Louis M Goldstein. 1986. Towards an
904 articulatory phonology. *Phonology* 3. 219–252.
- 905 Öhman, Sven E. G. 1966. Coarticulation in VCV utterances: Spectrographic mea-
906 surements. *The Journal of the Acoustical Society of America* 39(1). 151–168.
907 doi:10.1121/1.1909864.
- 908 Öhman, Sven E. G. 1967. Numerical model of coarticulation. *The Journal of the*
909 *Acoustical Society of America* 41(2). 310–320. doi:10.1121/1.1910340.
- 910 Pape, Daniel & Luis MT Jesus. 2014. Production and perception of velar stop
911 (de)voicing in European Portuguese and Italian. *EURASIP Journal on Audio,*
912 *Speech, and Music Processing* 2014(1). 6.
- 913 Peterson, Gordon E. & Ilse Lehiste. 1960. Duration of syllable nuclei in English.
914 *The Journal of the Acoustical Society of America* 32(6). 693–703. doi:10.1121/
915 1.1908183.
- 916 Plug, Leendert & Rachel Smith. 2018. Segments, syllables and speech tempo per-
917 ception. In *Proceedings of the 9th international conference on speech prosody*
918 *2018*, 279–283. doi:10.21437/SpeechProsody.2018-57.

- 919 Port, Robert F & Jonathan Dalby. 1982. Consonant/vowel ratio as a cue for voicing
920 in English. *Perception & Psychophysics* 32(2). 141–152.
- 921 R Core Team. 2019. R: A language and environment for statistical computing.
922 <https://www.R-project.org/>.
- 923 Raphael, Lawrence J. 1972. Preceding vowel duration as a cue to the perception of
924 the voicing characteristic of word final consonants in American English. *The*
925 *Journal of the Acoustical Society of America* 51(4B). 1296–1303. doi:10.1121/
926 1.1912974.
- 927 Ratnikova, E. I. 2017. Towards a log-normal model of phonation units lengths
928 distribution in the oral utterances. *International Research Journal* 3(57). 46–49.
929 doi:10.23670/IRJ.2017.57.103.
- 930 Roettger, Timo B. 2019. Researcher degrees of freedom in phonetic sciences. *Lab-*
931 *oratory Phonology: Journal of the Association for Laboratory Phonology* 10(1).
932 1–27. doi:10.5334/labphon.147.
- 933 Rosen, Kristin M. 2005. Analysis of speech segment duration with the lognormal
934 distribution: A basis for unification and comparison. *Journal of Phonetics* 33(4).
935 411–426. doi:10.1016/j.wocn.2005.02.001.
- 936 Saltzman, Elliot, Hosung Nam, Jelena Krivokapic & Louis Goldstein. 2008. A
937 task-dynamic toolkit for modeling the effects of prosodic structure on articu-
938 lation. In *Proceedings of the 4th international conference on speech prosody*
939 *(speech prosody 2008), campinas, brazil*, 175–184.
- 940 Sharf, Donald J. 1962. Duration of post-stress intervocalic stops and preceding
941 vowels. *Language and speech* 5(1). 26–30.
- 942 Sharf, Donald J. 1964. Vowel duration in whispered and in normal speech. *Lan-*
943 *guage and speech* 7(2). 89–97.
- 944 Slis, Iman H. & Antonie Cohen. 1969a. On the complex regulating the voiced-
945 voiceless distinction II. *Language and speech* 12(3). 137–155. doi:10.1177/00
946 2383096901200301.
- 947 Slis, Iman Hans & Antonie Cohen. 1969b. On the complex regulating the voiced-
948 voiceless distinction I. *Language and speech* 12(2). 80–102. doi:10.1177/0023
949 83096901200202.
- 950 Stevens, Kenneth N. & Samuel Jay Keyser. 1989. Primary features and their en-
951 hancement in consonants. *Language* 81–106.

- 952 Toivonen, Ida, Lev Blumenfeld, Andrea Gormley, Leah Hoiting, John Logan,
953 Nalini Ramlakhan & Adam Stone. 2015. Vowel height and duration. In Ulrike
954 Steindl, Thomas Borer, Huilin Fang, Alfredo García Pardo, Peter Guekguezian,
955 Brian Hsu, Charlie O'Hara & Iris Chuoying Ouyang (eds.), *Proceedings of the*
956 *32nd west coast conference on formal linguistics*, vol. 32, 64–71. Somerville,
957 MA: Cascadilla Proceedings Project.
- 958 Van Heuven, W. J. B., P. Mandera, E. Keuleers & M. Brysbaert. 2014. Subtlex-UK:
959 A new and improved word frequency database for british english. *Quarterly*
960 *Journal of Experimental Psychology* 67. 1176–1190.
- 961 Van Summers, W. 1987. Effects of stress and final-consonant voicing on vowel
962 production: Articulatory and acoustic analyses. *The Journal of the Acoustical*
963 *Society of America* 82(3). 847–863. doi:10.1121/1.395284.
- 964 Vasishth, Shravan, M. Beckman, B. Nicenboim, Fangfang Li & Eun Jong Kong.
965 2018a. Bayesian data analysis in the phonetic sciences: A tutorial introduction.
966 *Journal of Phonetics* 71. 147–161. doi:10.1016/j.wocn.2018.07.008.
- 967 Vasishth, Shravan, Daniela Mertzen, Lena A. Jäger & Andrew Gelman. 2018b.
968 The statistical significance filter leads to overoptimistic expectations of replica-
969 bility. *Journal of Memory and Language* 103. 151–175. doi:10.1016/j.jml.20
970 18.07.004.
- 971 Wagenmakers, Eric-Jan. 2007. A practical solution to the pervasive problems of
972 *p* values. *Psychonomic bulletin & review* 14(5). 779–804. doi:10.3758/BF03
973 194105.
- 974 Warren, Willis & Adam Jacks. 2005. Lip and jaw closing gesture durations in
975 syllable final voiced and voiceless stops. *The Journal of the Acoustical Society*
976 *of America* 117(4). 2618–2618. doi:10.1121/1.4778168.
- 977 Zeroual, Chakir, Philip Hoole, Adamantios I. Gafos & John H. Esling. 2015. Ges-
978 tural coordination differences between intervocalic simple and geminate plo-
979 sives in Moroccan Arabic: An EMA investigation. In *Proceedings of ICPHS*,
980 1–5.
- 981 Zmarich, Claudio, Barbara Gili Fivela, Pascal Perrier, Christophe Savariaux &
982 Graziano Tisato. 2011. Speech timing organization for the phonological length
983 contrast in Italian consonants. In *Twelfth annual conference of the international*
984 *speech communication association*, 401–404.