Compensatory aspects of the effect of voicing on vowel duration in English

Stefano Coretta

Abstract

It will come.

1 Introduction

Consonants and vowels are known to exert a reciprocal influence in a variety of ways. One such way is the well-established tendency for vowels to have shorter durations when followed by voiceless stops and longer durations when followed by voiced stops. This so-called 'voicing effect' has been long recognised 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 unequivocal consensus.

In an exploratory study of acoustic durations in Italian and Polish, Coretta (2018) proposes that the diachronic pathway to the voicing effect derives from a compensatory mechanism between vowel and consonant closure duration. Coretta finds that the duration of the interval between two consecutive consonant releases is not affected by the voicing status of the second consonants. In particular, disyllabic words of the form CV́CV (with lexical stress on the first syllable) were investigated. The consonant following the stressed vowel was either voiceless (for example, /pata/) or voiced (/pada/). The results of that study suggest that the duration of the release-to-release interval in words like /pata/ is not significantly different from that in words like /pada/.

Coretta (2018) argues that the temporal stability of the release-to-release interval is compatible with a compensatory temporal adjustment account of the voicing effect (Lindblom 1967; Slis & Cohen 1969a,b; Lehiste 1970a,b). This account states that vowels are shorter when followed by voiceless stops because the latter have longer closure durations, and vice versa with voiced stops, which have shorter closure durations. Coretta (2018) also reviews some of the shortcomings of the compensatory account and concludes that the release-to-release stability offers a resolution. In particular, previous versions of the account fail 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 as such intervals, but these have been subsequently criticised on empirical grounds (Chen 1970; Jacewicz et al. 2009; Maddieson & Gandour 1976).

Given the temporal stability of the release-to-release interval and the differential closure duration in voiceless vs. voiced stops (Lisker 1957; Van Summers 1987; Davis & Van Summers 1989; de Jong 1991), it follows that the timing of the consonant closure onset within that interval will decide on the respective durations of vowel and consonant closure. Coretta (2018) thus proposes that the source of the voicing effect can be seen in the temporal stability of the release-to-release interval in relation to the voicing of the second consonant and in the effect of voicing on closure duration. Naturally, other properties of speech production and perception can further contribute to the emergence and enhancement of the voicing effect, and these are not mutually exclusive with a compensatory account.

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; Laeufer 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 the magnitude of the effect in 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 opposition. This could explain the greater effect of English compared for example with 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. This means that, based on the surveyed data, the effect lies within that interval at a probability of 95%. In other words, we can have 95% confidence 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 values reported for Italian. Note also that the Italian values pertain to the effect of 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 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 effect (Kirby & Sonderegger 2018; Roettger 2019). In sum, it is generally assumed that the voicing-driven difference in vowel duration is greater in English than in other languages, although it should be stressed that the empirical foundation of this conception is not entirely straightforward.

1.1 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. Jacewicz et al. (2009) report that, in American English, monosyllabic words are longer when the second consonant is voiced. Based on this finding, it is expected that the duration of the release-to-release interval will differ in monosyllabic words depending on C2 voicing. More specifically, the release-to-release duration should be longer when C2 is voiced. Jacewicz et al. (2009) attribute the difference in monosyllabic word duration to the difference in vowel duration before voiceless vs. voiced stops. Thus, we can expect the magnitude of the difference in release-to-release duration in monosyllabic words to be close to the difference in vowel duration. This hypothesis also fits with the reported greater effect of voicing on vowel duration in monosyllabic than disyllabic words.

The data in Coretta (2018) suggest that the intrinsic duration of vowels and consonants can contribute to the duration of the release-to-release interval. In particular, release-to-release intervals containing a high vowel have shorter durations than those with a low vowel in Coretta (2018). This is not surprising, given the well known cross-linguistic tendency of high vowels to be shorter than low

vowels (Hertrich & Ackermann 1997; Esposito 2002; Mortensen & Tøndering 2013; Toivonen et al. 2015; Kawahara et al. 2017). As for the consonant place of articulation, the release-to-release interval is shorter in Italian and Polish when the second consonant is velar compared to when it is coronal. This could be a consequence of the fact that the closure of velar stops is shorter than that of other stops. For example, Sharf's (1962) data on closure duration in English suggests that the closure of labial stops (60-90 ms) is about 10 ms longer than that of velar stops (55-75 ms). It can be expected that release-to-release intervals with a velar stop in English will be about 10 ms shorter than intervals with a labial stop.

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

The following research questions and respective hypotheses were 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 shorter when C2 is velar.

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.

2 Methods

The following subsections describe the experimental and statistical methods. The research design and data analyses of this study were pre-registered at the Open Science Framework prior to data collection (https://osf.io/hwr94/?view_only=d994915422144efaae4a5915237cb386). The research compendium of this paper containing data and analysis scripts is also available on the Open Science Framework.

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

2.1 Sample size and stopping rule

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

Once a ROPE width is set, the goal is to collect data for sequential testing until the width of the 95% credible interval of the tested effect is equal to or less than the ROPE width (in this study, 20 ms). In other words, the objective is about the precision of the estimates, rather than whether a difference can be detected or not (like in standard frequentist power analyses). An initial minimum of 20 participants was chosen. Due to resource and time constraints specific to this particular study, a second condition had to be included in the stopping rule such that data collection would be have to stop on 5 April 2019, independent of the the ROPE condition.

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 = /iː, ɜː, ɑː/, C_2 = /p, b, k, g/, and (VC) = /əs/. This structure specification generates 24 test words, shown in Table 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 × 5 frames). A total of 1800 observations were recorded (120 stimuli × 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.

2.6 Data processing and measurements

The audio recordings were downsampled to 22050 Hz for analysis. 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:

- 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

Statistical analysis was performed in R v3.5.2 (R Core Team 2018). 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).[1] 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 choice of Bayesian over frequentist statistics stems from the misplaced reliance on *p*-values in the context of commonly low-powered studies (Munafò et al. 2017; Roettger 2019). 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

[1] Note that the order of the levels in the vowel and place factors are reversed compared to that in the pre-registration. This was done to match the vowel height order in Coretta (2018), from low to high, and to keep a back-to-front order for place for expositional purposes. Changing the order of the levels of course does not affect the results.

paper, it is relevant to provide the less familiar reader with the basic tools for interpreting analyses and results.

More weight will be given here on the estimated distributions of the sought effects, rather than on point estimates (as in frequentist regression models). 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 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). A possible concern is that the priors might have too much of an influence on the results. A sensitivity analysis based on posterior *z*-scores and shrinkage (Betancourt 2018) indicates that the models discussed in this study are highly informed by the observed data and don't heavily rely on prior specifications.

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. The full R code used for analysis is available as part of the paper's research compendium. 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). However, the deviations from a Gaussian distribution are minimal, and an informal comparison of one of the models fitted with a log-normal distribution led to virtually identical results.

3.1 Release-to-release duration

A Bayesian regression was fit to model the duration of the release-to-release interval. 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. Two weakly informative priors based on the results from Coretta (2018) were chosen for the intercept and the effect of C2 voicing. The former

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 × monosyll.	16.53	12.72	-8.41	41.41	49.83

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

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 priors as a normal distribution with mean 50 ms and SD = 25 was specified for the effect of number of syllable. 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 (2018). 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 ≤ 20 ms, based on the ROPE) was reached only for the 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 is mostly positive, 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 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 Sharf (1962), which range between 10 and 30 ms.

The summary of the fixed effects of model 2 are given in Table 3. As with model 1, the CI width of speech rate only reached the intended precision. The posterior distribution of the effect of the vowel /ɜ:/ shows that this vowel tends to a somewhat negative effect, with a 95% CI between −21.90 and 4.87

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

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 large 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 a variety of credible intervals for the model terms. 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 now more 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 coefficient ($\bar{\theta}$ = 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, 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 vs. 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; Laeufer 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

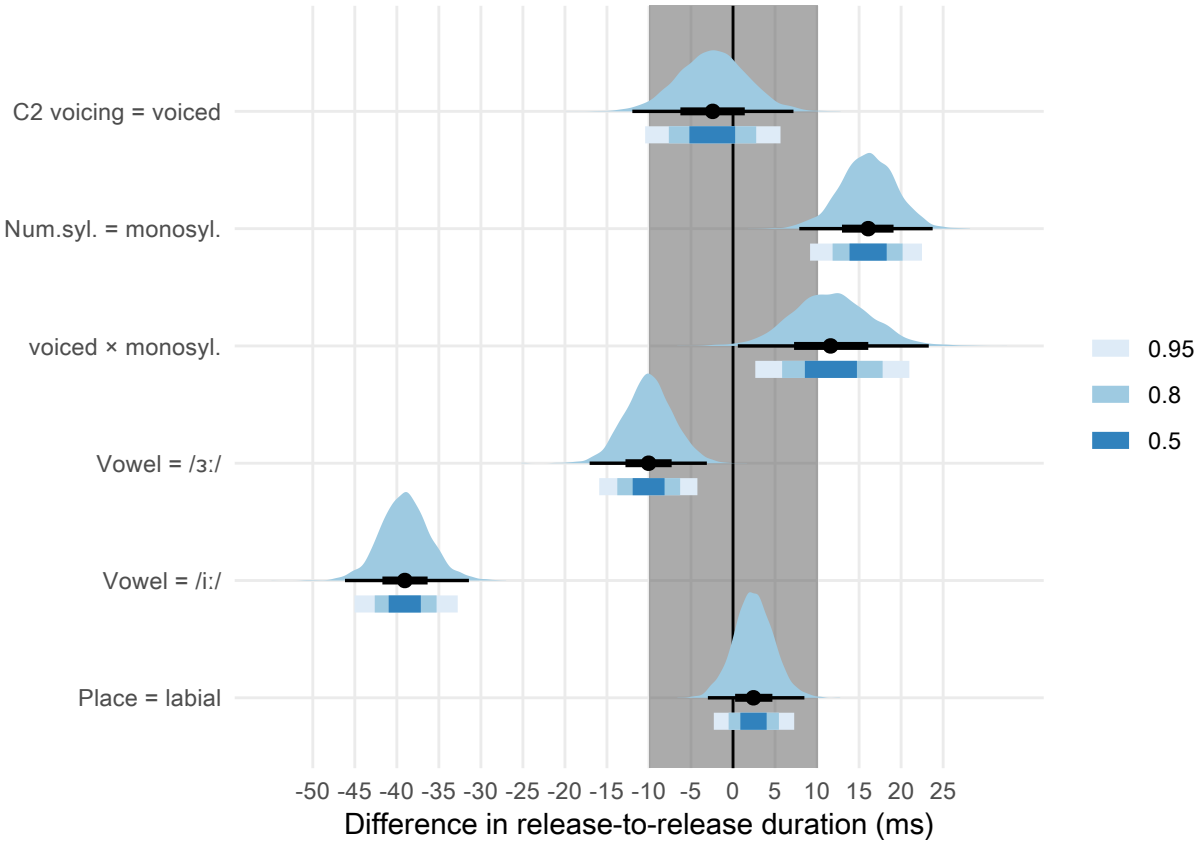

Figure 1: Bayesian credible intervals of the effects on release-to-release duration (model 3). For each effect, the thick blue-coloured bars indicate (from darker to lighter) the 50%, 80%, and 95% CI. The black point with bars are the posterior median (the point), the 98% (thin bar) and 66% (thicker 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	124.95	5.89	113.62	137.05	23.43
Voicing = voiced	13.48	5.30	3.46	24.36	20.90
Vowel = /ɜː/	-9.17	5.27	-19.06	1.46	20.52
Vowel = /iː/	-36.81	5.29	-46.93	-26.51	20.42
Num. syll. = monosyllabic	14.84	5.05	4.89	25.05	20.16
Speech rate (cntr.)	-18.01	1.49	-20.85	-15.13	5.71
voiced × /ɜː/	0.37	6.89	-13.77	13.36	27.13
voiced × /iː/	6.76	6.96	-7.13	20.57	27.70
voiced × monosyll.	4.12	6.69	-9.61	17.69	27.30
/ɜː/ × monosyll.	0.55	7.21	-13.76	14.18	27.94
/iː/ × monosyll.	-15.93	7.12	-29.78	-1.79	27.99
voiced × /ɜː/ × monosyll.	-2.89	9.53	-21.19	16.43	37.62
voiced × /iː/ × monosyll.	14.34	9.64	-4.47	33.47	37.94

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 exeption of the intercept. All the interactions terms have Ci widths above 25 ms. The 95% CI of the posterior distribution of the duration of /ɑː/ is included in the range 113.15–137.06 ms ($\bar{\theta}$ = 125.16 ms, SD = 6.02). The vowel /ɜː/ is 9.19 ms shorter (SD = 5.29) with CI = [−19.24, 1.49], while /iː/ is 36.97 ms shorter (SD = 5.14, 95% CI = [−46.99, −26.76]). C2 voicing has a small but robust positive effect on vowel duration in disyllabic words. The posterior distribution of the effect of voicing on /ɑː/ has mean 13.47 ms (SD = 5.14) and 95% CI = [3.80, 24.36]. The posterior of the interaction of voicing with vowel when the vowel is /ɜː/ is quite spread out, with the 95% CI between −13.77 and 13.40 ms. This indicates that /ɑː/ and /ɜː/ are similar in their behaviour of voicing-driven durational differences. On the other hand, the effect of voicing is on average 6.75 ms greater (SD = 6.76, 95% CI = [−7.45, 19.55]) when the vowel is /iː/.

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 /ɑː/ has mean 4.07 ms (SD = 6.6) and 95% CI [−8.94, 17.30]. This distribution indicates the possibility for a very small increase of the effect from disyllabic to monosyllabic words with /ɑː/. The three-way interaction C2 voicing/vowel/number of syllables suggests that the effect of voicing in monosyllabic words with /ɜː/ is very similar to that of monosyllabic /ɑː/-words ($\bar{\theta}$ = −2.66, SD = 9,44, 95% CI = [−21.57, 16.23]). On the other hand, the effect increases by 14.5 ms (SD = 9.43, CI = [−4.27, 33.41]) in monosyllabic words with /iː/ relative to disyllabic /iː/-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,

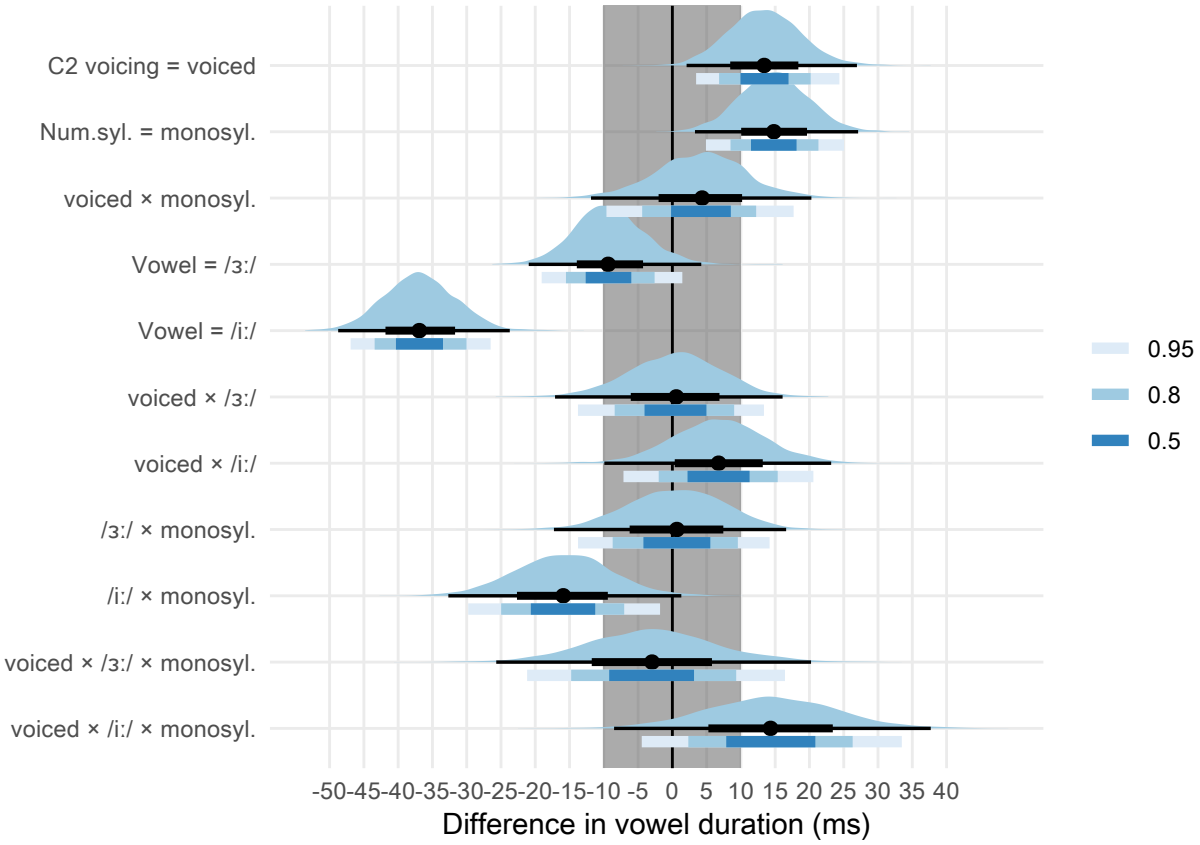

Figure 2: Bayesian credible intervals of the effects on vowel duration (model 4). For each effect, the thick blue-coloured bars indicate (from darker to lighter) the 50%, 80%, and 95% CI. The black point with bars are the posterior median (the point), the 98% (thin bar) and 66% (thicker bar) CI. The 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	74.58	2.96	68.70	80.28	11.58
Voicing = voiced	-20.84	3.04	-26.68	-14.80	11.88
C2 place = labial	5.18	2.84	-0.16	10.85	11.00
Num. syll. = monosyllabic	2.95	2.88	-2.64	8.70	11.34
Speech rate (cntr.)	-9.17	1.28	-11.70	-6.70	5.00
voiced × labial	1.47	3.95	-6.18	9.06	15.23
voiced × monosyll.	1.96	4.04	-6.09	10.01	16.10
labial × monosyll.	-0.73	3.96	-8.45	7.06	15.51
voiced × labial × monosyll.	6.31	5.74	-5.10	17.78	22.88

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 Sharf (1962) and Luce & Charles-Luce (1985) 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. 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 74.62 ms (SD = 2.85) and 95% CI = [69.17, 80.42]. The effect of C2 voicing on closure duration is certainly negative, between −26.76 and −14.57 ms (95% CI). The posterior mean of this effect is −20.83 ms (SD = 3.09). A very small positive effect of place of articulation (labial) is suggested by the 95% CI from −0.30 to 10.63 ms ($\bar{\theta}$ = 5.1 ms, SD = 2.73). A possibly even smaller effect of number of syllables or no effect at all can be inferred from the posterior distribution which has mean 2.81 ms and SD 2.84 (95% CI = [−2.70, 8.19]). See Figure 3 for the posteriors and credible intervals of the effects on closure duration.

4 Discussion

This study set out to investigate whether the results from Coretta (2018) could be replicated for English and extended to other contexts. It was expected that the release-to-release interval would not be affected by C2 voicing in disyllabic words but it is 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 synthesised these findings and proposes a diachronic pathway to the emergence of the voicing-driven durational differential (the voicing effect) based on a mechanism of temporal adjustments of gestural phasing. Finally, an articulatory grounding of the temporal properties of the release-to-release interval in mono- and disyllabic words is offered, with a discussion of limitations and future work.

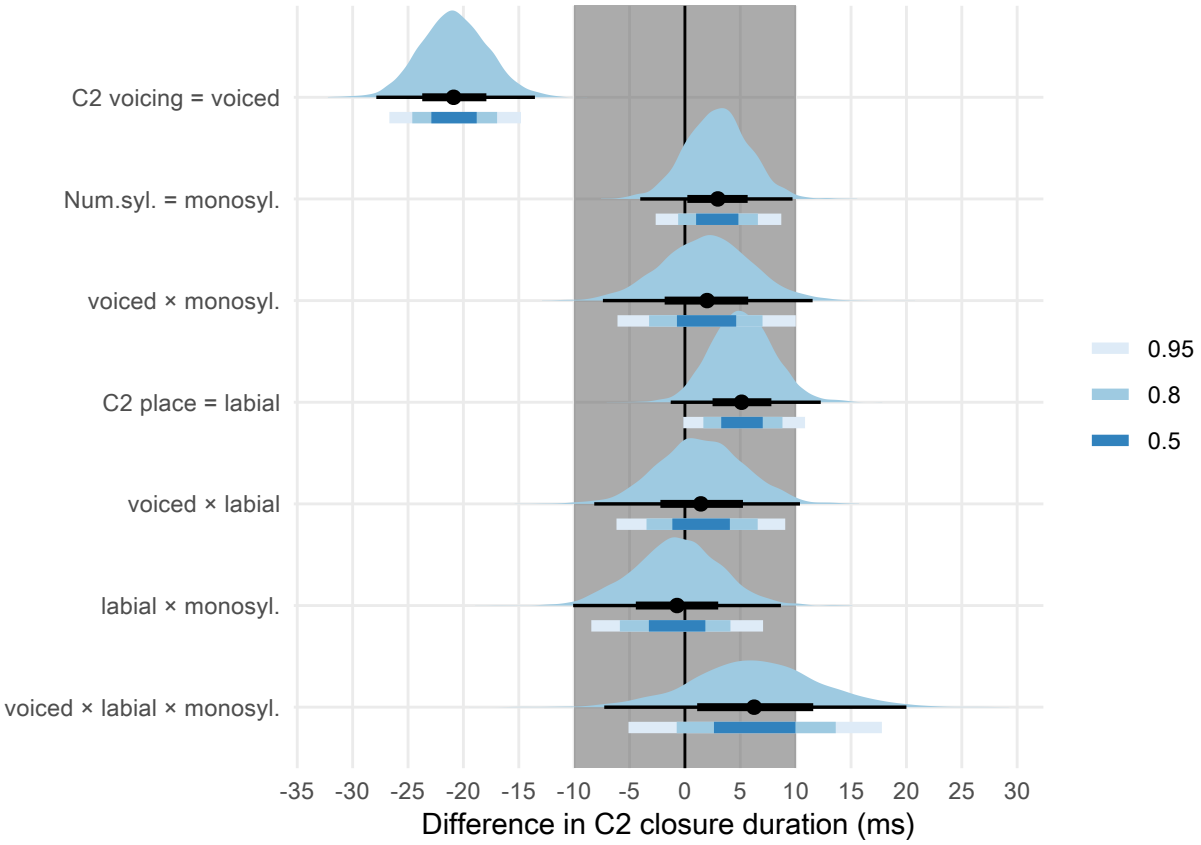

Figure 3: Bayesian credible intervals of the effects on closure duration (model 4). For each effect, the thick blue-coloured bars indicate (from darker to lighter) the 50%, 80%, and 95% CI. The black point with bars are the posterior median (the point), the 98% (thin bar) and 66% (thicker bar) CI. The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

4.1 Release-to-release interval

The first question (see Section 1.1) asked whether the voicing of C2 in disyllabic and monosyllabic words in English influence 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. The results of this study indicate that the release-to-release duration of disyllabic words in English is very similar in words where C2 is voiceless (like *tarpus*) and words with a voiced C2 (*tarbus*).

A Bayesian regression model was fitted to the release-to-release duration (model 3). The results of model 3 suggest a null 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 of the effect indicates an effect between −5 and 0 ms (50% probability). If the voicing of C2 is conditioning the duration of the release-to-release interval, this effect is very small and probably negative (around −2.5 ms).

The possible small negative 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, *tarb*) vs. when it is voiceless (*tarp*). 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.

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 interval 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 /iː/ than for /ɜː/. When the vowel is /iː/, the release-to-release interval is 33 to 45 ms shorter compared with an interval with /ɑː/. The posterior distribution of the effect when the vowel is /ɜː/ substantially overlaps with the ROPE, although it tends towards the negative side. If there is an effect with this vowel compared to /ɑː/, 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 (2018) 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 & Van 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 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.

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. The magnitude, however, is considerably lower than what previously reported. Based on what the posterior distribution indicates, and with the caveat that more data is needed to reach a sensible estimate precision and reduce uncertainty, we can argue for a mean effect increase in monosyllabic words by a factor of 1.3.

The three-way interaction between C2 voicing, vowel, and number of syllables reveals that the effect in monosyllabic words with the vowel /ɜː/ is similar to that with /ɑː/. On the other hand, the effect is larger if the vowel is /iː/. Model 4 estimates an effect increase of about 14.5 ms ([−4.27, 33.41]). Note that, although the credible interval is very wide (38 ms), it overlaps less extensively with the ROPE around 0, thus suggesting somewhat more clearly a positive effect. However, the vowel /iː/ followed by a voiceless stop is also about 16 ms shorter in monosyllabic words than the same vowel in disyllabic words. While it is not clear why the vowel is shorter in that context, it is possible that the estimated simultaneous increase (by the voicing effect) and decrease (by vowel identity) of vowel duration in that context by the same amount (16 ms) is the product of statistical modelling, making it difficult to draw any certain conclusion. Research on the duration of English tense vowels and on a possible process of /iː/ shortening is needed to shed light on 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 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.[2] Given that the duration of the release-to-release interval in Italian and Polish disyllabic words is not affected by the voicing of the post-vocalic consonant, the relative durations of vowel and closure depend on when the closure is achieved within that interval. A later closure onset implies a longer vowel and a shorter closure, while, vice versa, an earlier closure onset produces a shorter vowel and a longer closure. Vowels are shorter when followed by a voiceless stops because the closure of voiceless stop is achieved earlier than that of voiced stops, and vice versa.

The present study showed that also the release-to-release interval of English disyllabic words with either a voiceless or a voiced stop is isochronous. The voicing status of the second consonant of the interval does not affect its duration. The same pathway of temporal compensation proposed in the context of the Italian and Polish data can be envisaged for English, and possibly more generally in a cross-linguistic point of view. While future research will have to focus on investigating the source of the isochrony of the interval, an articulatory mechanism of gestural phasing is proposed here as the basis of the acoustic patterns.

4.3.1 Release-to-release isochorony as a consequence of gestural phasing

As already argued in Coretta (2018), the release-to-release interval in itself is not special. The compensatory temporal adjustment account can be understood in relation to the acoustic duration of vowels, hence the scope of compensation can (but need not) be defined in terms of acoustic intervals. The interval found to be temporally stable across voicing context is the release-to-release interval. However, it is desirable to derive the isochrony of this acoustic interval from properties of articulatory coordination. While a full theory of gestural phasing is beyond the scope of this study, here I offer a tentative account of the underlying gestural coordination from which the release-to-release isochrony can be derived.

Öhman (1966, 1967) proposed that the the speech stream is composed by a series of continuous vocalic gestures interrupted by gestures of oral constriction (consonants). Fowler (1983) further argued 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, and it is not affected by the number of intervening consonants.

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), extends the cyclicity argument by assuming that any two gestures can be implemented according to two modes: Either in synchrony or 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 consonant 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 the vowel depends on the number of onset consonants (Browman & Goldstein 1988; Marin & Pouplier 2010; Hermes et al. 2013; Marin & Pouplier 2014).

[2] Note that isochrony here is intended as pertaining the context of voiceless vs. voices stops only. It is not implied that the interval must be isochronous in all phonological contexts.

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.ta/) are longer than vowels followed by a tautosyllabic cluster (/la.dra/). This pattern can easily be derived from the C-centre alignment of the cluster /dr/, and if we assume that the distance between the vowels is the same in the two contexts (/la.ta/ and /la.dra/). 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. For a discussion on how speech rate could mask statistical correlations see Beguš (2017) and Coretta (2018).

Turning now to the voicing contrast, Van Summers (1987) and de Jong (1991) show 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 C-centre effect), full oral closure will be achieved earlier in voiceless than voiced stops relative to the beginning of the vocalic gesture. Now, if the latter is at a stable temporal distance from the gesture of the preceding vowel, it follows that the acoustic duration of the preceding vowel will be longer because of the later closure of voiced stops, while the time of the release is the same in voiceless and voiced stops. The articulatory apparatus just discussed could be implemented, for example, according to the mechanisms of the selection-coordination theory proposed in Tilsen (2013, 2016) (the reader is referred to Tilsen's work for the details).

4.3.2 Disyllabic vs. monosyllabic words

A complication for the diachronic pathway of the voicing effect discussed here arises from the pattern observed with English monosyllabic words. In this context, the release-to-release interval is not isochronous and is instead affected by C2 voicing. Words with a voiced C2 have longer release-to-release intervals (by about 5-17.5 ms at 80% probability). The compensatory mechanism set forth in this paper is based on the premise that the interval within which compensation happens must be isochronous. Since English monosyllabic words don't show voicing-wise isochrony, we either have to abandon a compensatory account or find a mechanism that can explain why the release-to-release interval in monosyllabic words is characterised by both absence of isochrony and compensation of segment durations.

A possible solution to the dilemma could be cautiously put forward from principles of Evolutionary Phonology (Blevins 2004, 2006). Most of present-day English monosyllabic words are diachronically related to disyllabic words in Old English and Proto-Germanic which, through vowel reduction and loss, became monosyllabic in English (and other Germanic languages). It is possible that a mechanism of compensation was in action at the diachronic stage in which present-day monosyllabic words were disyllabic. It can be speculated that a voicing effect should be reconstructed for Proto-Germanic itself. This hypothesis is consistent with a general principle of historical linguistics by which if a feature is present in the majority of daughter languages (which is the case for the voicing effect), it is sensible to reconstruct such feature for the ancestor proto-language. Furthermore, given that in almost all the investigated Indo-European languages the duration of vowels is modulated by the voicing of a following stop, we could even extend this principle to argue that the voicing effect could be reconstructed even from Proto-Indo-European.

Once the second vowel in disyllabic CVCV words is deleted via diachronic change, the second consonant loses its anchor vowel (V2) and shifts its affiliation to the preceding vowel instead. In other words, it becomes a coda consonant. Coda consonants are produced anti-phase with the preceding vocalic nucleus and among themselves, meaning that each gesture is executed in sequence. At this

point, disruption of V-to-V isochrony is no longer an issue, and certain perceptual biases (as proposed by perceptual accounts of the voicing effect like Javkin 1976 and Kluender et al. 1988) are able to come into action, by stretching or compressing the vocalic gesture. The duration of the vowel of monosyllabic words is then free to be modulated in order to enhance the perceptual difference of voiceless vs. voiced stops (Lisker 1974, 1986; Stevens & Keyser 1989).

5 Conclusion

TBA.

References

Ananthapadmanabha, T. V., A. P. Prathosh & A. G. 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.

Browman, Catherine P. & Louis Goldstein. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica* 45(2-4). 140–155.

Browman, Catherine P. & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49. 155–180.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1). 1–28. doi:10.18637/jss.v080.i01.

Bürkner, Paul-Christian. 2018. Advanced Bayesian multilevel modeling with the r package brms. *The R Journal* 10(1). 395–411. doi:10.32614/RJ-2018-017.

Caldognetto, Emanuela Magno, Franco Ferrero, Kyriaki Vagges & Maria Bagno. 1979. Indici acustici e indici percettivi nel riconoscimento dei suoni linguistici (con applicazione alle consonanti occlusive dell'italiano). *Acta Phoniatrica Latina* 2. 219–246.

Celata, Chiara & Paolo Mairano. 2014. On the timing of V-to-V intervals in Italian: a review, and some new hypotheses. *Revista de Filología Románica* 31. 37. doi:10.5209/rev_RFRM.2014.v31.n1.51022.

Chen, Matthew. 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22(3). 129–159.

Coretta, Stefano. 2018. An exploratory study of voicing-related differences in vowel duration as compensatory temporal adjustment in Italian and Polish. Submitted.

Davis, Stuart & W. Van Summers. 1989. Vowel length and closure duration in word-medial VC sequences. *Journal of Phonetics* 17. 339–353.

Durvasula, Karthik & Qian Luo. 2012. Voicing, aspiration, and vowel duration in Hindi. *Proceedings of Meetings on Acoustics* 18. 1–10. doi:10.1121/1.4895027.

Esposito, Anna. 2002. On vowel height and consonantal voicing effects: Data from Italian. *Phonetica* 59(4). 197–231. doi:10.1159/000068347.

Etz, Alexander, Quentin F. Gronau, Fabian Dablander, Peter A. Edelsbrunner & Beth Baribault. 2018. How to become a Bayesian in eight easy steps: An annotated reading list. *Psychonomic Bulletin & Review* 25(1). 219–234. doi:10.3758/s13423-017-1317-5.

Farnetani, Edda & Shiro Kori. 1986. Effects of syllable and word structure on segmental durations in spoken Italian. *Speech communication* 5(1). 17–34. doi:10.1016/0167-6393(86)90027-0.

Fowler, Carol A. 1983. Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General* 112(3). 386. doi:10.1037/0096-3445.112.3.386.

Fowler, Carol A. 1992. Vowel duration and closure duration in voiced and unvoiced stops: There are no contrast effects here. *Journal of Phonetics* 20(1). 143–165.

Halle, Morris & Kenneth Stevens. 1967. Mechanism of glottal vibration for vowels and consonants. *The Journal of the Acoustical Society of America* 41(6). 1613–1613. doi:10.1121/1.2143736.

Heffner, R.-M.S. 1937. Notes on the length of vowels. *American Speech* 12. 128–134. doi:10.2307/452621.

Hermes, Anne, Doris Mücke & Martine Grice. 2013. Gestural coordination of italian word-initial clusters: the case of 'impure s'. *Phonology* 30(01). 1–25.

Hertrich, Ingo & Hermann Ackermann. 1997. Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures. *The Journal of the Acoustical Society of America* 102(1). 523–536. doi:10.1121/1.419725.

House, Arthur S. & Grant Fairbanks. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America* 25(1). 105–113. doi:10.1121/1.1906982.

Huggins, A. William F. 1972. Just noticeable differences for segment duration in natural speech. *The Journal of the Acoustical Society of America* 51(4B). 1270–1278. doi:10.1121/1.1912971.

Hussein, Lutfi. 1994. *Voicing-dependent vowel duration in Standard Arabic and its acquisition by adult American students*: The Ohio State University dissertation.

Jacewicz, Ewa, Robert Allen Fox & Samantha Lyle. 2009. Variation in stop consonant voicing in two regional varieties of American English. *Journal of the International Phonetic Association* 39(3). 313–334. doi:10.1017/S0025100309990156.

Javkin, Hector R. 1976. The perceptual basis of vowel duration differences associated with the voiced/voiceless distinction. *Report of the Phonology Laboratory, UC Berkeley* 1. 78–92.

de Jong, Kenneth. 1991. An articulatory study of consonant-induced vowel duration changes in English. *Phonetica* 48(1). 1–17. doi:10.1121/1.2028316.

de Jong, Kenneth. 2004. Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. *Journal of Phonetics* 32(4). 493–516. doi:10.1016/j.wocn.2004.05.002.

Kawahara, Shigeto, Donna Erickson & Atsuo Suemitsu. 2017. The phonetics of jaw displacement in Japanese vowels. *Acoustical Science and Technology* 38(2). 99–107. doi:10.1250/ast.38.99.

Kirby, James & Morgan Sonderegger. 2018. Mixed-effects design analysis for experimental phonetics. *Journal of Phonetics* 70. 70–85. doi:10.1016/j.wocn.2018.05.005.

Klatt, Dennis H. 1973. Interaction between two factors that influence vowel duration. *The Journal of the Acoustical Society of America* 54(4). 1102–1104. doi:10.1121/1.1914322.

Kluender, Keith R., Randy L. Diehl & Beverly A. Wright. 1988. Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics* 16. 153–169.

Ko, Eon-Suk. 2018. Asymmetric effects of speaking rate on the vowel/consonant ratio conditioned by coda voicing in English. *Phonetics and Speech Sciences* 10(2). 45–50. doi:10.13064/KSSS.2018.10.2.045.

Kruschke, John. 2015. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan (2nd edition)*. Amsterdam, The Netherlands: Academic Press.

Laeufer, Christiane. 1992. Patterns of voicing-conditioned vowel duration in French and English. *Journal of Phonetics* 20(4). 411–440.

Lampp, Claire & Heidi Reklis. 2004. Effects of coda voicing and aspiration on Hindi vowels. *The Journal of the Acoustical Society of America* 115(5). 2540–2540. doi:10.1121/1.4783577.

Lehiste, Ilse. 1970a. Temporal organization of higher-level linguistic units. *The Journal of the Acoustical Society of America* 48(1A). 111–111. doi:10.1121/1.1974906.

Lehiste, Ilse. 1970b. Temporal organization of spoken language. In *Working papers in linguistics*, vol. 4, 96–114. doi:10.1121/1.1974906.

Lindblom, Björn. 1967. Vowel duration and a model of lip mandible coordination. *Speech Transmission Laboratory Quarterly Progress Status Report* 4. 1–29.

Lisker, Leigh. 1957. Closure duration and the intervocalic voiced-voiceless distinction in English. *Language* 33(1). 42–49. doi:10.2307/410949.

Lisker, Leigh. 1974. On “explaining” vowel duration variation. In *Proceedings of the Linguistic Society of America*, 225–232.

Lisker, Leigh. 1986. “Voicing” in English: a catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29(1). 3–11. doi:10.1177/002383098602900102.

Luce, Paul A & Jan Charles-Luce. 1985. Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *The Journal of the Acoustical Society of America* 78(6). 1949–1957. doi:10.1121/1.392651.

Machač, Pavel & Radek Skarnitzl. 2009. *Principles of phonetic segmentation*. Epocha.

Maddieson, Ian & Jack Gandour. 1976. Vowel length before aspirated consonants. In *UCLA Working papers in Phonetics*, vol. 31, 46–52.

Marin, Stefania & Marianne Pouplier. 2010. Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control* 14(3). 380–407. doi:10.1123/mcj.14.3.380.

Marin, Stefania & Marianne Pouplier. 2014. Articulatory synergies in the temporal organization of liquid clusters in Romanian. *Journal of Phonetics* 42. 24–36. doi:10.1016/j.wocn.2013.11.001.

McElreath, Richard. 2015. *Statistical rethinking: A bayesian course with examples in R and Stan*. CRC Press.

Meyer, Ernst Alfred. 1904. Zur vokaldauer im deutschen. In *Nordiska studier tillegnade A. Noreen*, 347–356. K.W. Appelbergs Boktryckeri: Uppsala.

Mortensen, Johannes & John Tøndering. 2013. The effect of vowel height on Voice Onset Time in stop consonants in CV sequences in spontaneous Danish. In *Proceedings of Fonetik 2013*, Linköping, Sweden: Linköping University.

Munafò, Marcus R., Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie Du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware & John P. A. Ioannidis. 2017. A manifesto for reproducible science. *Nature Human Behaviour* 1(1). 0021. doi:10.1038/s41562-016-0021.

Nicenboim, Bruno, Timo B. Roettger & Shravan Vasishth. 2018. Using meta-analysis for evidence synthesis: The case of incomplete neutralization in german. *Journal of Phonetics* 70. 39–55. doi:10.1016/j.wocn.2018.06.001.

Nooteboom, Sieb G. & Gert J. N. Doodeman. 1980. Production and perception of vowel length in spoken sentences. *The Journal of the Acoustical Society of America* 67(1). 276–287. doi:10.1121/1.383737.

O’Dell, Michael L. & Tommi Nieminen. 2008. Coupled oscillator model for speech timing: Overview and examples. In *Nordic prosody: Proceedings of the Xth conference*, 179–190.

Ohala, John J, Catherine P Browman & Louis M Goldstein. 1986. Towards an articulatory phonology. *Phonology* 3. 219–252.

Öhman, Sven E. G. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America* 39(1). 151–168. doi:10.1121/1.1909864.

Öhman, Sven E. G. 1967. Numerical model of coarticulation. *The Journal of the Acoustical Society of America* 41(2). 310–320. doi:10.1121/1.1910340.

Pape, Daniel & Luis MT Jesus. 2014. Production and perception of velar stop (de)voicing in European Portuguese and Italian. *EURASIP Journal on Audio, Speech, and Music Processing* 2014(1). 6.

Peterson, Gordon E. & Ilse Lehiste. 1960. Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America* 32(6). 693–703. doi:10.1121/1.1908183.

Plug, Leendert & Rachel Smith. 2018. Segments, syllables and speech tempo perception. In *Proceedings of the 9th international conference on speech prosody 2018*, 279–283. doi:10.21437/SpeechProsody.2018-57.

Port, Robert F & Jonathan Dalby. 1982. Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics* 32(2). 141–152.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Raphael, Lawrence J. 1972. Preceding vowel duration as a cue to the perception of the voicing characteristic of word final consonants in American English. *The Journal of the Acoustical Society of America* 51(4B). 1296–1303. doi:10.1121/1.1912974.

Ratnikova, E. I. 2017. Towards a log-normal model of phonation units lengths distribution in the oral utterances. *International Research Journal* 3(57). 46–49. doi:10.23670/IRJ.2017.57.103.

Roettger, Timo B. 2019. Researcher degrees of freedom in phonetic sciences. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 10(1). 1–27. doi:10.5334/labphon.147.

Rosen, Kristin M. 2005. Analysis of speech segment duration with the lognormal distribution: A basis for unification and comparison. *Journal of Phonetics* 33(4). 411–426. doi:10.1016/j.wocn.2005.02.001.

Saltzman, Elliot, Hosung Nam, Jelena Krivokapic & Louis Goldstein. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In *Proceedings of the 4th international conference on speech prosody (speech prosody 2008), campinas, brazil*, 175–184.

Sharf, Donald J. 1962. Duration of post-stress intervocalic stops and preceding vowels. *Language and speech* 5(1). 26–30.

Sharf, Donald J. 1964. Vowel duration in whispered and in normal speech. *Language and speech* 7(2). 89–97.

Slis, Iman H. & Antonie Cohen. 1969a. On the complex regulating the voiced-voiceless distinction II. *Language and speech* 12(3). 137–155. doi:10.1177/002383096901200301.

Slis, Iman Hans & Antonie Cohen. 1969b. On the complex regulating the voiced-voiceless distinction I. *Language and speech* 12(2). 80–102. doi:10.1177/002383096901200202.

Stevens, Kenneth N. & Samuel Jay Keyser. 1989. Primary features and their enhancement in consonants. *Language* 81–106.

Tilsen, Sam. 2013. A dynamical model of hierarchical selection and coordination in speech planning. *PLoS ONE* 8(4). e62800. doi:10.1371/journal.pone.0062800.

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

Toivonen, Ida, Lev Blumenfeld, Andrea Gormley, Leah Hoiting, John Logan, Nalini Ramlakhan & Adam Stone. 2015. Vowel height and duration. In Ulrike Steindl, Thomas Borer, Huilin Fang, Alfredo García Pardo, Peter Guekguezian, Brian Hsu, Charlie O'Hara & Iris Chuoying Ouyang (eds.), *Proceedings of the 32nd west coast conference on formal linguistics*, vol. 32, 64–71. Somerville, MA: Cascadilla Proceedings Project.

Van Summers, W. 1987. Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *The Journal of the Acoustical Society of America* 82(3). 847–863. doi:10.1121/1.395284.

Vasishth, Shravan, M. Beckman, B. Nicenboim, Fangfang Li & Eun Jong Kong. 2018a. Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics* 71. 147–161. doi:10.1016/j.wocn.2018.07.008.

Vasishth, Shravan, Daniela Mertzen, Lena A. Jäger & Andrew Gelman. 2018b. The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language* 103. 151–175. doi:10.1016/j.jml.2018.07.004.

Warren, Willis & Adam Jacks. 2005. Lip and jaw closing gesture durations in syllable final voiced and voiceless stops. *The Journal of the Acoustical Society of America* 117(4). 2618–2618. doi:10.1121/1.4778168.