Pre-registration for Compensatory aspects of the effect of voicing on vowel duration in English

Stefano Coretta

16/01/2019

1 Study information

1.1 Title

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

1.2 Authorship

Stefano Coretta (The University of Manchester).

1.3 Research questions

Vowels in English (and other languages) are longer when followed by voiced stops than when followed by voiceless stops (Heffner, 1937). The origin of this 'voicing effect' is still object of debate. According to the compensatory temporal adjustment account, the difference in vowel duration is the consequence of the difference in stop closure durations (Lindblom, 1967; Slis & Cohen, 1969b,a; Lehiste, 1970b,a). Voiceless stops have longer closures than voiced stops (Lisker, 1957; Van Summers, 1987; Davis & Van Summers, 1989; de Jong, 1991), so that vowels are longer when followed by the shorter closure of voiced stops.

An exploratory study discussed in Coretta (2018) indicates that the duration of the interval between two consecutive stops releases is not affected by consonant voicing in CV́CV words in Italian and Polish. Coretta argues that the temporal stability of the release to release interval plus the difference in consonant closure durations drives the differences in vowel durations by a mechanism of compensation between the duration of the vowel and that of the closure.

This study sets out to test whether English exhibits the same temporal pattern found in Italian and Polish. The aim of the study is to answer the following questions:

- Q1: 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?
- Q2: 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?

A third question pertains to differences between mono- and disyllabic words:

- Q3: What is the estimated difference in the effect of voicing on vowel and stop closure duration between monosyllabic and disyllabic words?

1.4 Hypotheses

In relation to Q1, the following hypotheses will be tested:

- 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.

H1b is proposed based on Jacewicz et al. (2009), who report that monosyllabic words are longer when C2 is voiced in American English.

The exploratory study of Italian and Polish indicates that the intrinsic duration of the vowel and the consonant contribute to the duration of the release to release interval. More specifically, the release to release with a high vowel is shorter than that with a low vowel. It is well known that cross-linguistically high vowels tend to be shorter than low vowels (Hertrich & Ackermann, 1997; Esposito, 2002; Mortensen & Tøndering, 2013; Toivonen et al., 2015; Kawahara et al., 2017). As for consonant place of articulation, if the consonant is velar the interval is shorter. The closure of velar stops is shorter than that of labial stops (see for example Sharf 1962). It is possible that some of the durational difference of the release to release interval seen in the exploratory study depend on intrinsic vowel and consonant duration.

Considering the relation between intrinsic vowel and consonant duration and release to release, we can formulate the following hypotheses in relation to Q2 above:

- 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 /ɑː/, /əː/, /iː/.
- H2c: The release to release interval is shorter when C2 is velar.

As for Q3:

- H3: The effect of voicing on vowel duration is greater in monosyllabic than in disyllabic words.

There is no specific hypothesis concerning differences in voiceless vs. voiced stop closure duration between mono- and disyllabic words.

2 Sampling plan

2.1 Existing data

Registration prior to creation of data: As of the date of submission of this research plan for preregistration, the data have not yet been collected, created, or realised.

2.2 Explanation of existing data

NA.

2.3 Data collection procedures

2.3.1 Participants

Inclusion rule: Native speakers of English from the Manchester area, 18+ yo, with no reported hearing or speaking disorders, and with normal or corrected to normal vision. The participants will sign a written consent and will receive a monetary compensation of £5.

2.3.2 Procedure

The participants will be recorded while reading sentences with CVC and CV́CVC target words presented on a computer screen with PsychoPy (Peirce, 2009), in a sound attenuated room in the Phonetics Laboratory at the University of Manchester. The test words are $C_1V_1C_2$(VC) words, where C_1 = /t/, V_1 = /iː, əː, ɑː/, C_2 = /p, b, k, g/, and (VC) = /əs/. This structure leads to 24 possible combinations.

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

Each target word is combined with 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 will be read once. Acoustic recording will be obtained 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 recordings will be subject to force alignment with SPPAS (Bigi, 2015) for analysis in Praat (Boersma & Weenink, 2018).

2.4 Sample size

24 words, 5 repetitions (= 120 tokens per participant). We set a minimum of 20 participants with additional participants if needed (see following sections).

2.5 Sample size rationale

The approach of the region of practical equivalence (ROPE) will be used for determining the participant sample size (Kruschke, 2015; Vasishth et al., 2018b). This approach is based on establishing a region around 0 which could theoretically be interpreted as a 'no effect' region. Based on estimates from the literature on the just noticeable difference (Huggins, 1972; Nooteboom & Doodeman, 1980), the chosen ROPE width is 20 milliseconds (from -10 to +10 milliseconds).

2.6 Stopping rule

Data collection will stop either:

(a) When the width of the 95% CI of the effect of C2 voicing will be less than 20 milliseconds, or

(b) If the ROPE target hasn't been reached by April 5th 2019, due to time constraints imposed on the project.

3 Variables

3.1 Manipulated variables

- **Vowel**: /iː/, /əː/, /ɑː/.
- **Voicing of C2**: voiceless (/p, k/), voiced (/b, g/).
- **Place of articulation of C2**: labial (/p, b/), velar (/k, g/).
- **Number of syllables**: monosyllabic, disyllabic.
- **Frame sentence**: 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.

See Section 2.3.2 for a list of word stimuli.

3.2 Measured variables

- From the acoustic signal:
  - **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 number of syllables per second (number of syllables in the sentence divided by the sentence duration in seconds, Plug & Smith 2018).

C1/C2 release detection will be performed with an automatic procedure in Praat based on Ananthapadmanabha et al. (2014). The output of the automatic force-alignment (see Section 2.3.2) and the release detection algorithm will be checked and corrected manually if needed.

3.3 Indices

NA.

4 Design plan

4.1 Study type

Experiment—A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomised controlled trials.

4.2 Blinding

No blinding is involved in this study.

4.3 Study design

Repeated measures, mixed design.

4.4 Randomisation

The sentence stimuli will be randomised within participant by means of the built-in randomisation procedure in PsychoPy (Peirce, 2009).

5 Analysis plan

5.1 Statistical models

Bayesian linear mixed models (Vasishth et al., 2018a; McElreath, 2015; Kruschke, 2015) will be fitted with brms (Bürkner, 2017b,a) in R (R Core Team, 2018).

5.1.1 Release to release

The following Bayesian regression model will be used to model the duration of the release to release interval. As fixed effects: C2 voicing (factor, levels = 'voiceless', 'voiced'), number of syllables (factor, levels = 'disyllabic', 'monosyllabic'), centred speech rate, interaction between C2 voicing and number of syllables. Factors are coded with treatment contrasts. A by-speaker and by-word random intercept, and a by-speaker random coefficient for C2 voicing.

```
rr_1 <- brm(
  rr ~
    c2_voice +
    n_syl +
    c2_voice:n_syl +
    speech_rate_c +
    (1 + c2_voice | speaker) +
    (1 | word),
  family = gaussian()
)
```

The following priors will be used. For the intercept of the release to release interval duration, a normal distribution with mean 200 ms and SD = 50, based on the data from the Italian/Polish exploratory study. For the effect of C2 voicing, a weakly informative prior as a normal distribution with mean 0 ms and SD = 25, based on results from the exploratory study. For the effect of number of syllables, a weakly informative prior as a normal distribution with mean 50 ms and SD = 25, based on differences in vowel duration between mono- and disyllabic words, which range between 30 and 100 ms (Sharf, 1962; Klatt, 1973). The same prior is chosen for the interaction of C2 voicing and number of syllables, based on a similar range of reported differences in vowel duration in monosyllabic words (30-100 ms). For the effect of speech rate (centred), a normal distribution with mean -25 and SD = 10, based on results from the exploratory study. For the random effects, a half Cauchy distribution (location = 0, scale = 25) for the standard deviation and a LKJ(2) distribution for the correlation. For the residual standard deviation, a half Cauchy distribution with location 0 ms and scale 25.

```
c(
  set_prior("normal(200, 50)", class = "Intercept"),
  set_prior("normal(0, 25)", class = "b", coef = "c2_voicevoiced"),
  set_prior("normal(50, 25)", class = "b", coef = "n_sylmono"),
  set_prior("normal(50, 25)", class = "b", coef = "c2_voicevoiced:n_sylmono"),
  set_prior("normal(-25, 10)", class = "b", coef = "speech_rate_c"),
  set_prior("cauchy(0, 25)", class = "sd"),
  set_prior("lkj(2)", class = "cor"),
  set_prior("cauchy(0, 25)", class = "sigma")
)
```

A separate model will test the effect of vowel and C2 place on the duration of the release to release duration. As fixed effects: vowel (factor, levels = ‘ee’, ‘er’, ‘ar’), C2 place (factor, levels = ‘labial’, ‘velar’), centred speech rate, interaction between vowel and C2 place. Factors are coded with treatment contrasts. By-speaker and by-word random intercepts.

```
rr_2 <- brm(
  rr ~
    vowel +
    c2_place +
    speech_rate_c +
    (1 | speaker) +
    (1 | word),
  family = gaussian()
)
```

The following priors will be used. For the intercept of the release to release interval duration, a normal distribution with mean 200 ms and SD = 50, as above. For the effect of vowel, a weakly informative prior as a normal distribution with mean 0 ms and SD = 30, based on differences in vowel duration reported in Heffner (1937), House & Fairbanks (1953), and Hertrich & Ackermann (1997). For the effect of C2 place of articulation, a weakly informative prior as a normal distribution with mean 0 ms and SD = 30, based on differences in closure durations between labial and velar stops, which range between 10 and 30 ms (Sharf, 1962). The remaining priors are the same as the ones of the previous model.

```
c(
  set_prior("normal(200, 50)", class = "Intercept"),
  set_prior("normal(0, 30)", class = "b", coef = "voweler"),
  set_prior("normal(0, 30)", class = "b", coef = "vowelar"),
  set_prior("normal(0, 30)", class = "b", coef = "c2_placevelar"),
  set_prior("normal(-25, 10)", class = "b", coef = "speech_rate_c"),
  set_prior("cauchy(0, 25)", class = "sd"),
  set_prior("cauchy(0, 25)", class = "sigma")
)
```

5.1.2 Vowel duration

The following Bayesian regression model will be used to model the duration of V1. As fixed effects: C2 voicing (factor, levels = ‘voiceless’, ‘voiced’), vowel (factor, levels = ‘ee’, ‘er’, ‘ar’), number of syllables (factor, levels = ‘disyllabic’, ‘monosyllabic’), centred speech rate, all logical interactions between C2 voicing, vowel, and number of syllables. Factors are coded with treatment contrasts. A by-speaker and by-word random intercept, and a by-speaker random coefficient for C2 voicing.

```
vow_1 <- brm(
  v1_duration ~
    c2_voice +
    vowel +
    n_syl +
    c2_voice:vowel +
    c2_voice:n_syl +
    vowel:n_syl +
    c2_voice:vowel:n_syl +
    speech_rate_c +
    (1 + c2_voice | speaker) +
    (1 | word),
  family = gaussian()
)
```

The following priors will be used.

```
c(
  set_prior("normal(145, 30)", class = "Intercept"),
  set_prior("normal(50, 20)", class = "b", coef = "c2_voicevoiced"),
  set_prior("normal(0, 30)", class = "b", coef = "voweler"),
  set_prior("normal(0, 30)", class = "b", coef = "vowelar"),
  set_prior("normal(50, 25)", class = "b", coef = "n_sylmono"),
  set_prior("normal(0, 20)", class = "b", coef = "c2_voicevoiced:voweler"),
  set_prior("normal(0, 20)", class = "b", coef = "c2_voicevoiced:vowelar"),
  set_prior("normal(50, 25)", class = "b", coef = "c2_voicevoiced:n_sylmono"),
  set_prior("normal(0, 30)", class = "b", coef = "voweler:n_sylmono"),
  set_prior("normal(0, 30)", class = "b", coef = "vowelar:n_sylmono"),
  set_prior("normal(0, 30)", class = "b", coef = "c2_voicevoiced:voweler:n_sylmono"),
  set_prior("normal(0, 30)", class = "b", coef = "c2_voicevoiced:vowelar:n_sylmono"),
  set_prior("normal(-25, 10)", class = "b", coef = "speech_rate_c"),
  set_prior("cauchy(0, 25)", class = "sd"),
  set_prior("lkj(2)", class = "cor"),
  set_prior("cauchy(0, 25)", class = "sigma")
)
```

5.1.3 Closure duration

The following Bayesian regression model will be used to model the duration of the closure of C2. As fixed effects: C2 voicing (factor, levels = ‘voiceless’, ‘voiced’), vowel (factor, levels = ‘ee’, ‘er’, ‘ar’), number of syllables (factor, levels = ‘disyllabic’, ‘monosyllabic’), centred speech rate, all logical interactions between C2 voicing, vowel, and number of syllables. Factors are coded with treatment contrasts. A by-speaker and by-word random intercept, and a by-speaker random coefficient for C2 voicing.

```
c2_1 <- brm(
  c2_duration ~
    c2_voice +
    c2_place +
    n_syl +
    c2_voice:c2_place +
    c2_voice:n_syl +
    c2_place:n_syl +
    c2_voice:c2_place:n_syl +
    speech_rate_c +
    (1 + c2_voice | speaker) +
    (1 | word),
  family = gaussian()
)
```

The following priors will be used. The means reported in Sharf (1962) and Luce & Charles-Luce (1985) indicate that stop closures in monosyllabic words are 10-30 ms shorter when the stop is voiced. A normal distribution with mean -20 ms and SD = 10 seems informative enough as a prior to the effect of C2 voicing on closure duration. The same studies indicate velar stops have a closure which is 10-20 ms shorter in monosyllabic words, hence the prior `normal(mean = -15, SD = 10)` for the effect of C2 place. Note that the estimates for C2 voicing and place refer to monosyllabic words, while the specified priors refer to disyllabic words (the reference level of `n_syl`).

```
c(
  set_prior("normal(90, 20)", class = "Intercept"),
  set_prior("normal(-20, 10)", class = "b", coef = "c2_voicevoiced"),
  set_prior("normal(-15, 10)", class = "b", coef = "c2_placevelar"),
  set_prior("normal(0, 25)", class = "b", coef = "n_sylmono"),
  set_prior("normal(0, 30)", class = "b", coef = "c2_voicevoiced:c2_placevelar"),
  set_prior("normal(0, 30)", class = "b", coef = "c2_voicevoiced:n_sylmono"),
  set_prior("normal(0, 30)", class = "b", coef = "c2_placevelar:n_sylmono"),
  set_prior("normal(0, 30)", class = "b", coef = "c2_voicevoiced:c2_placevelar:n_sylmono"),
  set_prior("normal(-25, 10)", class = "b", coef = "speech_rate_c"),
  set_prior("cauchy(0, 25)", class = "sd"),
  set_prior("lkj(2)", class = "cor"),
  set_prior("cauchy(0, 25)", class = "sigma")
)
```

5.2 Transformations

Speech rate will be centred to make the estimates of the intercepts interpretable. Centring will be obtained with the standard formula (speech rate - mean speech rate).

5.3 Follow-up analyses

NA.

5.4 Inference criteria

Bayesian posterior distributions, rather than point estimates, will be used for inference.

5.5 Data exclusion

Residual data with mistakes and speech errors will be discarded.

5.6 Missing data

Individual missing observations (due to excluded data or annotation difficulties) will be dropped.

5.7 Exploratory analysis

An exploratory analysis will look into differences in closure durations of voiceless and voiced stops in mono- and disyllabic words (see Section 5.1.3).

6 Script (Optional)

6.1 Analysis scripts (Optional)

See Section 5.1.

7 Other

NA.

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.

Bigi, Brigitte. 2015. SPPAS - Multi-lingual approaches to the automatic annotation of speech. *The Phonetician* 111–112. 54–69.

Boersma, Paul & David Weenink. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.40. http://www.praat.org/.

Bürkner, Paul-Christian. 2017a. Advanced Bayesian multilevel modeling with the r package brms. *The R Journal* 10(1). 395–411. doi:10.32614/RJ-2018-017.

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

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.

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

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

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.

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.

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.

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.

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.

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

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.

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.

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

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

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.

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.

Peirce, Jonathan W. 2009. Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics* 2(10).

Plug, Leendert & Rachel Smith. 2018. Segments, syllables and speech tempo perception. Talk presented at the 2018 Colloquium of the British Association of Academic Phoneticians (BAAP 2018).

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

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

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.

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.