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

Stefano Coretta

1 Introduction

Consonants and vowels are known to excise 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 plethora of languages across different linguistic families. English is possibly by far the most investigated language in relation to the voicing effect. Several hypotheses have been proposed as to the origin of this phenomenon, however no one particular hypothesis has gained unequivocal consensus.

Coretta (2018) proposes to seek the source of the voicing effect in a compensatory mechanism between vowel and consonant closure duration. In an exploratory study of acoustic durations in 17 Italian and Polish speakers, 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, Coretta investigated disyllabic words of the form CV́CV (with lexical stress on the first syllable), where the second consonant was either voiceless or voiced. The results of that study suggest that, for example, the duration of the release-to-release interval in a word like /pata/ is non-significantly different from that in /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, which states that vowels are shorter when followed by voiceless stops because the latter have long closure durations, and vice versa with voiced stops, which have shorter closure durations (Lindblom 1967; Slis & Cohen 1969b,a; Lehiste 1970b,a). Coretta also reviews some of the shortcomings of the compensatory account and concludes that the release-to-release invariance offers a resolution to those. In particular, previous versions of the account are not clear on within which speech interval the compensation is implemented. Both the syllable (Lindblom 1967; Farnetani & Kori 1986) and the word (Slis & Cohen 1969b,a; Lehiste 1970b,a) have been proposed but 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 the effect of voicing on closure duration. Other properties of speech production and perception can of course further contribute to the emergence and enhancement of the voicing effect (see Beguš 2017 for an overview).

English is generally the language given as an example in which the voicing effect has a big magnitude. This fact is normally attributed to phonologisation of the voicing effect in this language Sharf (1964); de Jong (2004). Indeed, previous studies on English report a difference in vowel duration before voiceless vs. voiced stops which ranges between 20 and 150 ms. A Bayesian meta-analysis of the English voicing effect (See Supplemental materials) indicates that 95% credible interval for the effect of voicing in monosyllabic words is between 55 and 95 ms. This means that, based on the data, the true effect lies within that interval at a probability of 95%, on in other words we can be 95% sure that the effect is between 55-95 ms.

1.1 Research hypotheses

The principal aim 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 be observed in English. Jacewicz et al. (2009) report that monosyllabic words in American English 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 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 to be close to the difference in vowel duration.

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 that a well known cross-linguistic tendency is that high vowels are 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 interval is shorter in Italian and Polish if the second consonant is velar compared to when it is coronal. The closure of velar stops is shorter than that of other stops (see for example Sharf 1962). It can thus be expected that intervals with a velar stop in English will be shorter than intervals with consonants of other places of articulation.

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

- 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?
  - 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.
- 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?
  - 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.
- Q3: 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 research design and data analyses of this study has been pre-registered at the Open Science Framework.

2.1 Participants

The participants of this study were 15 native speakers of British English, who were born and brought up in the Greater Manchester area. All the speakers were 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.

Sample size and stopping rule were determined prior to data collection by the Region Of Practical Equivalence (ROPE) method (Kruschke 2015; Vasishth et al. 2018). A ‘no-effect’ region of values around 0 is first identified. The null region (the ROPE) can be thought of as the Bayesian 95% credible interval from a distribution, the values within which can be interpreted as a 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).

Once a ROPE width is set, the goal is to collect data 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). 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 April 5th 2019, independent of the the ROPE condition.

2.2 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 cloths, about 20 cm under their mouth, displaced a few centimetres on one side.

2.3 Materials

The test words were $C_1V_1C_2$(VC) words, where C_1 = /t/, V_1 = /iː, əː, ɑː/, C_2 = /p, b, k, g/, and (VC) = /əs/. This structure specification generates 24 test words, shown in **??**. 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 would read 120 sentence stimuli (24 words × 5 frames).

2.4 Procedure

The participants were first debriefed on the experimental procedure. Prior to recording, the participants familiarised themselves with the materials by reading them aloud and were instructed not to insert pauses anywhere within the sentence stimuli and to try and keep a similar intonation contour for the total duration of the experiment. They were also given the change to take any number of breaks at any point during recording. Misreadings or speech errors were corrected by letting the participant repeat the stimulus. The reading task took around 6 to 10 minutes.

2.5 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 number of syllables per second (number of syllables in the sentence divided by the sentence duration in seconds, Plug & Smith 2018).

2.6 Statistical analysis

All statistical analyses were performed in R v3.5.2 (R Core Team 2018). Bayesian regression models were fit with brms (Bürkner 2017, 2018). All factors were coded using treatment contrasts with the first level as the reference level: number of syllables (disyllabic, monosyllabic), vowel (/ɑː/, /ɜː/, /iː/), C2 voicing (voiceless, voiced), C2 place of articulation (velar, labial). Speech rate has been centred when included in the models so that the intercept can be interpreted as the mean intercept at mean speech rate.

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.

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

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.

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.

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

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.

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.

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.

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.

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.

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.

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.

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.

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.

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, Daniela Mertzen, Lena A. Jäger & Andrew Gelman. 2018. The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language* 103. 151–175.