An exploratory study of voicing-related differences in vowel duration as compensatory temporal adjustment in Italian and Polish

Stefano Coretta

1 Introduction

Almost a hundred years of research have consistently shown that consonantal voicing has an effect on preceding vowel duration: vowels followed by voiced obstruents are longer than when followed by voiceless ones (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). This so called 'voicing effect' has been found in a considerable variety of languages, including (but not limited to) English, German, Hindi, Russian, Italian, Arabic, and Korean (see Maddieson & Gandour 1976 for a more comprehensive, but still not exhaustive list). Despite of the plethora of evidence in support of the *existence* of the voicing effect, agreement hasn't been reached regarding its *source*.[1]

Several proposals have been put forward in relation to the possible source of the voicing effect (see Sóskuthy 2013 and Beguš 2017 for an overview). The majority of the proposed accounts place the source of the voicing effect in properties of speech production.[2] A notable production account, which will be the focus of this study, is the compensatory temporal adjustment account (Lindblom, 1967; Slis & Cohen, 1969b,a; Lehiste, 1970b,a). According to this account, the voicing effect follows from the reorganisation of gestures within a unit of speech that is not affected by stop voicing. The duration of such unit is held constant across voicing contexts, while the duration of voiceless and voiced obstruents differs. The closure of voiceless stops is longer than that of voiced stops (Lisker, 1957; Van Summers, 1987; Davis & Van Summers, 1989; de Jong, 1991). As a consequence, vowels followed by voiceless stops (which have a long closure) are shorter than vowels followed by voiced stops (which have a short closure). Advocates of the compensatory account propose two prosodic units as the scope of the temporal adjustment: the syllable (or, more neutrally, the VC sequence, Lindblom 1967), and the word (Slis & Cohen, 1969b,a; Lehiste, 1970b,a). However, the compensatory temporal adjustment account has been criticised in subsequent work.

Empirical evidence and logic challenge the proposal that the syllable or the word have a constant duration and hence drive compensation. First, Lindblom's 1967 argument that the duration of the syllable is constant is not supported by the findings in Chen (1970) and Jacewicz et al. (2009). Chen

[1] Here, I take 'source' as a general term without any particular theoretical commitment to whether it should be intended as a strictly synchronic or diachronic cause.

[2] Two accounts that posit a perceptual cause are Javkin (1976) and Kluender et al. (1988). To the best of my knowledge, Javkin's (1976) proposal remains to be empirically tested, while see Fowler (1992) for arguments against Kluender et al. (1988).

(1970) rejects a syllable-based compensatory account in the light of the fact that the duration of the syllable is affected by consonant voicing. Jacewicz et al. (2009) further show that the duration of monosyllabic words in American English changes depending on the voicing of the coda consonant. Second, although the results in Slis & Cohen (1969a) suggest that the duration of disyllabic words in Dutch is constant whether the second stop is voiceless or voiced, it does not follow from this fact that compensation should necessarily target the vowel preceding the stop. Indeed, it is logically possible that the following unstressed vowel could be the target of the compensation, therefore differences in preceding vowel duration still call for an explanation.

The compensatory temporal adjustment account has been further challenged on the basis of the so called 'aspiration effect' (Maddieson & Gandour, 1976), by which vowels are longer when followed by aspirated stops than when followed by unaspirated stops. In Hindi, vowels before voiceless unaspirated stops are the shortest, followed by vowels before voiced unaspirated and voiceless aspirated stops, which have similar duration, followed by vowels before voiced aspirated stops, which are the longest. Maddieson & Gandour (1976) find no compensatory pattern between vowel and consonant duration: the consonant /t/, which has the shortest duration, is preceded by the shortest vowel, and vowels before /d/ and /t^h/ have the same duration although the durations of the two consonants are different. Maddieson & Gandour (1976) argue that a compensatory explanation for differences in vowel duration cannot be maintained.

However, a re-evaluation of the way consonant duration is measured in Maddieson & Gandour (1976) might actually turn their findings in favour of a compensatory account. Due to difficulties in detecting the release of the consonant of interest, consonant duration in Maddieson & Gandour (1976) is measured from the closure of the relevant consonant to the release of the following, (e.g., in *ab sāth kaho*, the duration of /t^h/ in *sāth* is calculated as the interval between the closure of /t^h/ and the release of /k/). This measure includes the burst and aspiration (if present) of the consonant following the target vowel. Slis & Cohen (1969b), however, state that the inverse correlation between vowel duration and the following consonant applies to *closure* duration, and not to the entire *consonant* duration. If a correlation exists between vowel and closure duration, the inclusion of burst and/or aspiration clearly alters this relationship.

Indeed, the study on Hindi voicing and aspiration effects conducted by Durvasula & Luo (2012) indicates that closure duration, measured from closure onset to closure offset, decreases according to the hierarchy voiceless unaspirated > voiced unaspirated > voiceless aspirated > voiced aspirated, which closely resembles the order of increasing vowel duration in Maddieson & Gandour (1976). Nonetheless, Durvasula & Luo (2012) do not find a negative correlation between vowel duration and consonant closure duration, but rather a (small) *positive effect*. Vowel duration increases with closure duration when voicing and aspiration are taken into account. However, as noted in Beguš (2017), it is likely that this result is a consequence of not controlling for speech rate. A small negative effect of closure duration can turn positive if the effect of speech rate (which is positive) is greater, given the cumulative nature of these effects (Beguš, 2017, p. 2177).

Further evidence for a compensatory account comes from the effect of a third type of consonants, namely ejectives. Beguš (2017) finds that in Georgian (which contrasts aspirated, voiced, and ejective consonants) vowels are short when followed by voiceless aspirated stops, longer before ejective stops, and longest when followed by voiced stops. Crucially, stop closure duration follows the reversed pattern: closure duration is short in voiced stops, longer in ejectives, and longest in voiceless aspirated stops. Moreover, vowel duration is inversely correlated with closure duration across the three phonation types. Beguš (2017) argues that these findings support a temporal compensation account (although not univocally, see Beguš 2017, Section V).

To summarise, a compensatory temporal adjustment account has been proposed to explain the voicing effect. According to such account, the difference in vowel duration before consonants varying in voicing (and possibly other phonation types) is the outcome of a compensation between vowel and closure duration. After a careful review of the critiques advanced by Chen (1970) and Maddieson & Gandour (1976), and in face of the results in Slis & Cohen (1969a) and Beguš (2017), a compensatory account gains credibility. However, issues about the actual implementation of the compensation mechanism still remain. In conclusion, while the compensatory temporal adjustment account is plausible on the light of the reviewed literature, we are left with the necessity of finding a constant speech interval within which compensation is logically implemented.

1.1 The present study

This paper reports on selected results from a broader exploratory study that investigates the relationship between vowel duration and consonant voicing from both an acoustic and articulatory perspective. Synchronised recordings of audio, ultrasound tongue imaging, and electroglottography were carried out to enable a data-driven approach to the analysis of features related to the voicing effect in the context of disyllabic (CV́CV) words in Italian and Polish.[3] Given its exploratory nature, this study was not devised to test the compensatory account, but rather to collect articulatory and acoustic data on the voicing effect. Moreover, the design of the study has been constrained by the use of such articulatory techniques (see Section 2). Since the tongue imaging and electroglottographic data don't bear on the main argument put forward here, only the results from acoustics will be discussed.

Italian and Polish reportedly differ in the magnitude (or presence) of the effect of stop voicing on vowel duration, while they are both classified as voicing languages (languages in which the laryngeal opposition in consonants is between voiceless unaspirated and voiced consonants, Beckman et al. 2013).[4] For this reason, these two language offer the opportunity to investigate differences that could reveal mechanisms underlying the voicing effect. Moreover, given that Italian and Polish share—on a general level—some features of the segmental and prosodic make-up of their phonological systems, the design of the experimental material and comparison of the results is facilitated.

Italian has been unanimously reported as a voicing-effect language (Caldognetto et al., 1979; Farnetani & Kori, 1986; Esposito, 2002). The mean difference in vowel duration when followed by voiceless vs. voiced consonants ranges between 22 and 24 ms in these studies, with longer vowels followed by voiced consonants. The mean differences are based on 3 speakers in Farnetani & Kori 1986 and 7 speakers in Esposito 2002. Caldognetto et al. 1979 does not report estimates of vowel duration, just the direction of the effect, but the study is based on 10 speakers. On the other hand,

[3] As per Cysouw & Good (2013), the glossonyms *Italian* and *Polish* as used here to refer, respectively, to the languoids Italian [GLOTTOCODE: `ital1282`] and Polish [GLOTTOCODE: `poli1260`].

[4] The distinction between voicing and aspirating varieties of Polish discussed in @cyran2011 is based on phonological considerations rather than empirical data on VOT duration.

the results regarding the presence and magnitude of the effect in Polish are mixed. While Keating (1984) reports no effect of voicing on vowel duration in data from 24 speakers, Nowak (2006) finds that vowels followed by voiced stops are 4.5 ms longer in the 4 speakers recorded. Malisz & Klessa (2008) argue based on data from 40 speakers that the magnitude of the voicing effect in Polish is highly idiosyncratic, and claim their results are inconclusive on this matter.

The acoustic data from the exploratory study discussed here suggests the existence of a voicing effect both in Italian and Polish, and that the duration of the interval between two consecutive stop releases (the Release to Release interval) is not affected by the voicing of the second consonant in both languages. This finding is compatible with a compensatory temporal adjustment account by which the timing of the stop closure onset within said interval determines the respective durations of vowel and closure. I further propose that the constant duration of the Release to Release interval is congruent with current views on gestural timing (Goldstein & Pouplier, 2014) and I discuss the insights such an account provides in relation to our understanding of the gestural organisation of speech.

2 Method

2.1 Participants

For this exploratory study, we set a target of 10 speakers per language.[5] The stopping rule was to reach 10 speakers in both languages or to end data collection within 15 months from the start, due to resource and time limits. Participants were sought in Manchester, UK, and in Verbania, Italy. Seventeen subjects in total participated in this study. Eleven subjects are native speakers of Italian (5 female, 6 male), while six are native speakers of Polish (3 female, 3 male). The Italian speakers are from the North and Centre of Italy (8 speakers from Northern Italy, 3 from Central Italy). The Polish group has 2 speakers from Western Poland, 3 speakers from Central Poland, and 1 speaker from Eastern Poland. For more information on the sociolinguistic details of the speakers, see Appendix B. Ethical clearance for this study was obtained from the University of Manchester (REF 2016-0099-76). The participants signed a written consent and received a monetary compensation of £10.

2.2 Equipment

The acquisition of the audio signal was achieved with the software Articulate Assistant Advanced™ (AAA, v2.17.2, Articulate Instruments Ltd™ 2011) running on a Hawlett-Packard ProBook 6750b laptop with Microsoft Windows 7. Audio recordings were sampled at 22050 Hz (16-bit), in a proprietary format (`.aa0`). A FocusRight Scarlett Solo pre-amplifier and a Movo LV4-O2 Lavalier microphone

[5] A low target number of participants was chosen since processing of ultrasound data takes longer than standard acoustic analysis.

were used for audio recording. The microphone was placed at the level of the participant's mouth on one side, at a distance of about 10 cm.[6]

2.3 Materials

The target stimuli were disyllabic words with $C_1V_1C_2V_2$ structure, where C_1 = /p/, V_1 = /a, o, u/, C_2 = /t, d, k, g/, and $V_2 = V_1$ (e.g. /pata/, /pada/, /poto/, etc.).[7] Most are nonce words, although inevitably some combinations produce real words both in Italian (4 words) and Polish (2 words, see Appendix C). The lexical stress of the target words was placed by speakers of both Italian and Polish on V_1, as intended.

The make-up of the target words was constrained by the design of the experiment, which included ultrasound tongue imaging (UTI). Front vowels are difficult to image with UTI, since their articulation involves tongue positions which are particularly far from the ultrasonic probe, hence reducing the visibility of the tongue contour. For this reason, only central and back vowels were included. Since one of the variables of interest in the exploratory study was the closing gesture of C_2, only lingual consonants were used. A labial stop was chosen as the first consonant to reduce possible coarticulation with the following vowel (although see Vazquez-Alvarez & Hewlett 2007). The number of target words was kept low to reduce the time required for completing the task.[8]

The target words were embedded in a frame sentence. Controlling for meaning, segmental and prosodic make-up betweeen languages proved to be difficult. The frames are *Dico X lentamente* 'I say X slowly' in Italian (following Hajek & Stevens, 2008), and *Mówię X teraz* 'I say X now' in Polish, and presented according to the respective writing conventions. These sentences were chosen in order to keep the placement of focus similar across languages (on the target word). Possible influences of the semantic content of the frame sentence on duration estimates will be discussed in Section 4.

2.4 Procedure

The participant was asked to read the sentences with the target words which were sequentially presented on the computer screen. The order of the sentences was randomised for each participant. Participants read the list of randomised sentence stimuli 6 times. Due to software constraints, the order of the list was kept the same across the six repetitions within each participant. The reading task lasted between 15 and 20 minutes, with optional short breaks between one repetition and the other. The total experiment time lasted around 45 minutes. Each speaker read a total of 12 sentences

[6]The microphone was clipped onto a metal headset wore by the participant, which is part of the ultrasound equipment.

[7]Italian has both a mid-low [ɔ] and a mid-high [o] back vowel in its vowel inventory. These vowels are traditionally described as two distinct phonemes (Krämer, 2009), although both their phonemic status and their phonetic substance are subject to a high degree of geographical and idiosyncratic variability (Renwick & Ladd, 2016). As a rule of thumb, stressed open syllables in Italian (like the ones used in this study) have [ɔː] (vowels in penultimate stressed open syllables are long) rather than [oː] (Renwick & Ladd, 2016). On the other hand, Polish has only a mid-low back vowel phoneme /ɔ/ (Gussmann, 2007). For sake of typographical simplicity, the symbol /o/ will be used here for both languages.

[8]The ultrasonic equipment usually gets uncomfortable when wore for more than 20 minutes.

Table 1: Criteria for the identification of acoustics landmarks.

landmark		**criteria**
vowel onset	(V1 onset)	Appearance of higher formants in the spectrogram following the release of /p/ (C1)
vowel offset	(V1 offset)	Disappearance of the higher formants in the spectrogram preceding the target consonant (C2)
consonant onset	(C2 onset)	Corresponds to V1 offset
closure onset	(C2 closure onset)	Corresponds to V1 offset
consonant offset	(C2 offset)	Appearance of higher formants of the vowel following C2 (V2); corresponds to V2 onset
consonant release	(C1/C2 release)	Automatic detection + manual correction (Ananthapadmanabha et al., 2014)

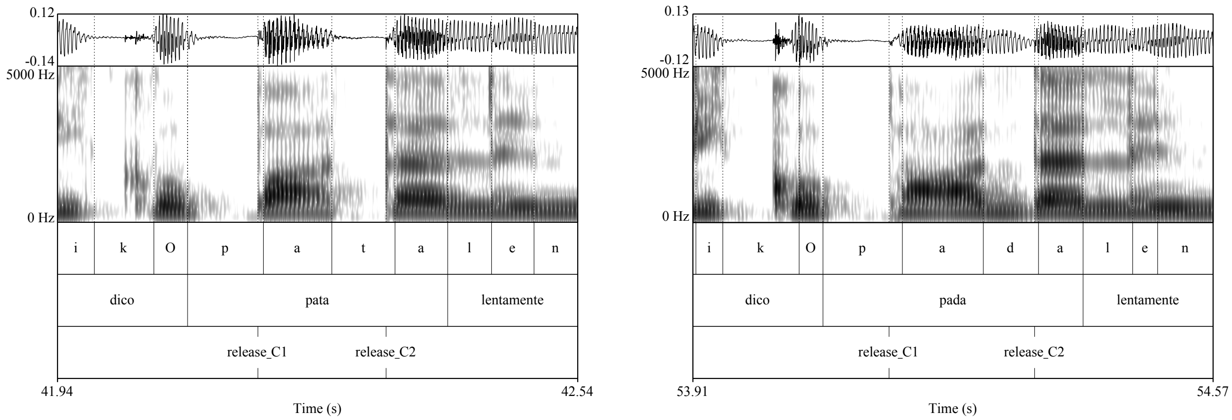
for 6 times (with the exceptions of IT02, who repeated the 12 sentences 5 times), which yields to a grand total of 1212 tokens (792 from Italian, 420 from Polish).[9]

The experiment was carried out in two locations: in the sound attenuated booth of the Phonetics Laboratory at the University of Manchester, and in a quiet room in a field location in Italy (Verbania, Northern Italy). In both locations the equipment and procedures were the same. Data collection started in December 2016 and ended in March 2018.

2.5 Data processing and measurements

The audio recordings were exported from AAA in the `.wav` format at the same sample and bit rate for further processing. A forced aligned transcription was accomplished through the SPeech Phonetisation Alignment and Syllabification software (SPPAS) (Bigi, 2015). The outcome of the automatic annotation was manually corrected for the relevant boundaries, according to the criteria in Table 1 based on Machač & Skarnitzl (2009). Segmentation boundaries not used in the analyses have not been checked to speed up processing. The releases of C1 and C2 were detected automatically by means of a Praat scripting implementation of the algorithm described in Ananthapadmanabha et al. (2014), and subsequently corrected if necessary. The identification of the stop release was not possible in 99 tokens (8%) of C1 and 265 tokens (22%) of C2 out of 1212. This was due either to the absence of a clear burst in the waveform and spectrogram, or the realisation of voiced stops as voiced fricatives Most of the fricativised tokens come from the three speakers of Central Italian, IT12, IT13, and IT14, a variety of Italian known to show this phenomenon Hualde & Nadeu 2011. Moreover IT12 and IT14 produced several tokens of voiceless stops with voicing during closure (in some cases the closure is completely voiced). These tokens have been kept in the analysis, because (1) the actual presence or absence of voicing during closure does not bear on the compensatory account discussed here (which concerns supraglottal gestures) and laryngeal gestures can be imple-

[9] IT01 and IT02 (the first two participants of this study) also read sentences with words starting with /b/, which were later excluded from the experimental design. The data from /b/-initial words are not included in the analysis reported in this paper.



(a) /pata/ (b) /pada/

Figure 1: Segmentation example of the words *pata* and *pada* uttered by the Italian speaker IT09 (the times on the *x*-axis refer to the times in the concatenated audio file).

mented almost entirely independently from oral gestures, and (2) the voicing effect has been shown to exist even in whispered speech, where vocal fold vibration is entirely absent (Sharf, 1964).

The durations in milliseconds of the following intervals were extracted with Praat scripting from the annotated acoustic landmarks: word duration, vowel duration (V1 onset to V1 offset), consonant closure duration (V1 offset to C2 release), and Release to Release duration (C1 release to C2 release). Sentence duration was measured in seconds. **??** shows an example of the segmentation of /pata/ (a) and /pada/ (b) from an Italian speaker. Syllable rate (syllables per second) was used as a proxy to speech rate (Plug & Smith, 2018), and was calculated as the number of syllables divided by the duration of the sentence in seconds (8 syllables in Italian, 6 in Polish). All further data processing and visualisation was done in R v3.5.0 (R Core Team, 2018; Wickham, 2017).

2.6 Statistical analysis

Given the data-driven nature of the study, all statistical analyses reported here are to be considered exploratory (hypothesis-generating) rather than confirmatory (hypothesis-driven) (Kerr, 1998; Gelman & Loken, 2013; Roettger, 2018). The durational measurements were analysed with linear mixed-effects models using `lme4` v1.1-17 in R (Bates et al., 2015), and model estimates were extracted with the `effects` package v4.0-2 (Fox, 2003). All factors were coded with treatment contrasts and the following reference levels: voiceless (vs. voiced), /a/ (vs. /o/, /u/), coronal (vs. velar), Italian (vs. Polish). The models were fitted by Restricted Maximum Likelihood estimation (REML). The estimates in the results section refer to these reference levels unless interactions are discussed. *P*-values for the individual terms were obtained with `lmerTest` v3.0-1, which uses the Satterthwaite's approximation to degrees of freedom (Kuznetsova et al., 2017; Luke, 2017). A result is considered significant if the *p*-value is below the alpha level ($\alpha = 0.05$).[10]

Bayes factors were used to specifically test the null hypotheses that word and Release to Release duration are not affected by C2 voicing (i.e., the effect of C2 voicing on duration is 0). For each set of null/alternative hypotheses, a full model (with the predictor of interest) and a null model (excluding it) were fitted separately using the Maximum Likelihood estimation (ML) (Bates et al., 2015, p. 34). The Bayes Information Criterion (BIC) approximation was then used to obtain Bayes factors (Raftery, 1995, 1999; Wagenmakers, 2007; Jarosz & Wiley, 2014). The approximation is calculated according to the equation in 1 (Wagenmakers, 2007, p. 796).

$$BF_{01} \approx exp(\Delta BIC_{10}/2) \tag{1}$$

where $\Delta BIC_{10} = BIC_1 - BIC_0$, BIC_1 is the BIC of the full model, and BIC_0 is the BIC of the null model. Values of $BF_{01} > 1$ indicate a preference of H_0 over H_1. The interpretation of the Bayes factors follows the recommendations in Raftery (1995, p. 139): 1–3 = weak evidence, 3–20 = positive evidence, 20–150 = strong evidence, $>$ 150 = very strong evidence.

The extracted measurements were filtered before statistical analysis. Measures of vowel duration, closure duration, word duration, and Release to Release duration that are 3 standard deviations lower or higher than the respective means were excluded from the final dataset (this procedure generally corresponds to a loss of around 2.5% of the data). One sentence (sentence 48 of IT07, *Dico pada lentamente*) included a speech error and has been excluded. Further excluding missing measurements, these operations yields a total of 920 tokens of vowel and closure durations, 1176 tokens of word duration, and 848 tokens of Release to Release duration.

While the study has been devised to compare Italian and Polish, the low number of Polish speaker (6, especially compared with the number of Italian speakers 11), makes statistical comparison difficult (see Kirby & Sonderegger 2018 and references therein for a discussion on statistical power). The raw mean differences, presented in conjunction with the estimates from statistical modelling, can still inform us on the cross-linguistic differences and thus they will contribute to the discussion of the results.

2.7 Open Science statement

Following recommendations for Open Science in Crüwell et al. (2018) and Berez-Kroeker et al. (2018) the data and code used to produce the analyses discussed in this paper are available on the Open Science Framework at https://osf.io/bfyhr/?view_only=391ef2dcc2834039a90f739ddb6f137a (Coretta, 2018).

[10]Luke (2017) argues that the common approach of using likelihood ratio tests for statistical inference with mixed models leads to inflated Type I error rates. Luke (2017, 1501) also warns that 'results should be interpreted with caution, regardless of the method adopted for obtaining *p*-values'.

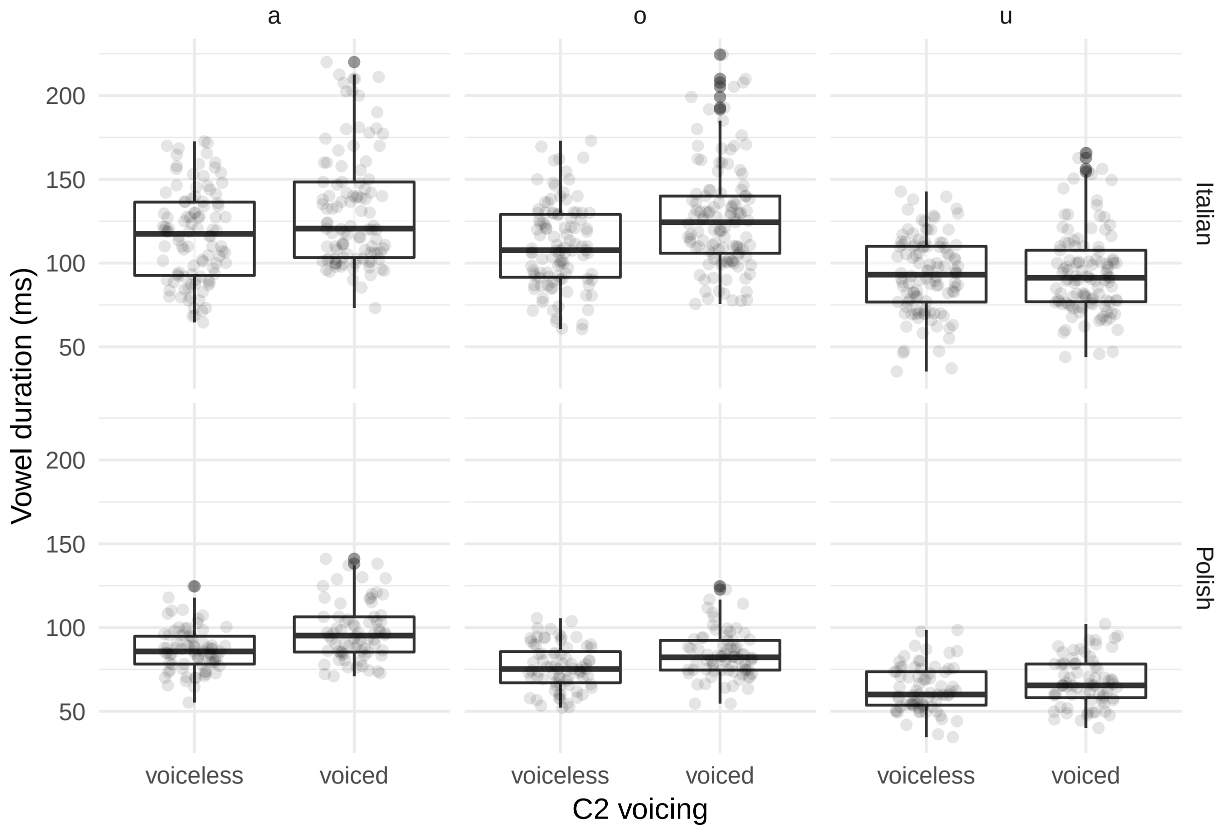


Figure 2: Raw data and boxplots of the duration in milliseconds of vowels in Italian (top row) and Polish (bottom row), for the vowels /a, o, u/ when followed by a voiceless or voiced stop.

3 Results

The following sections report the results of the study in relation to the durations of vowels, consonant closure, word, and the Release to Release interval. When discussing the output of statistical modelling, only the relevant predictors and interactions will be presented. To avoid the visual cluttering of parameters tables and alleviate the burden of the reader, the full output of statistical models (including confidence intervals and *p*-values) are given in Appendix A.

3.1 Vowel duration

Figure 2 shows boxplots and raw data of vowel duration for the three vowels /a, o, u/ when followed by voiceless or voiced stops in Italian and Polish. Vowel tend to be longer when followed by a voiced stop in both languages. The effect appears to be greater in Italian than in Polish, especially for the vowels /a/ and /o/. There is no evident effect of C2 voicing in /u/ in Italian (as in Ferrero et al. 1978), but the effect is discernible in Polish /u/. In Italian, vowels have a mean duration of 106.16 ms (SD = 27.08) before voiceless stops, and a mean duration of 117.66 ms (SD = 34.63) before voiced stops. Polish vowels are on average 75.57 ms long (SD = 16.16) when followed by a voiceless stop, and 83.11 ms long (SD = 19.37) if a voiced stop follows. The difference in vowel duration based on the raw means is 11.5 ms in Italian and 7.54 ms in Polish.

A linear mixed-effects model with vowel duration as the outcome variable was fitted with the following predictors: fixed effects for C2 voicing (voiceless, voiced), C2 place of articulation (coronal, velar), vowel (a, o, u), language (Italian, Polish), and speech rate (as syllables per second); by-speaker and by-word random intercepts with by-speaker random slopes for C2 voicing. All possible interactions between C2 voicing, vowel, and language were included. The following terms are significant according to *t*-tests with Satterthwaite's approximation to degrees of freedom: C2 voicing, vowel, language, and speech rate. Only the interaction between C2 voicing and vowel is significant. Vowels are 16.28 ms longer (SE = 4.42) when followed by a voiced stop (C2 voicing). The effect of C2 voicing is smaller with /u/ (around 3 ms, $\hat{\beta}$ = -13.1 ms, SE = 5.56). Polish has on average shorter vowels than Italian ($\hat{\beta}$ = -24.05 ms, SE = 7.83), and the effect of voicing is estimated to be about 10.55 ms (although note that the interaction between language and C2 voicing is not significant). Speech rate has a negative effect on vowel duration, such that faster rates correlate with shorter vowel durations ($\hat{\beta}$ = -16.23 ms, SE = 1.26).

3.2 Consonant closure duration

Figure 3 illustrates stop closure durations with boxplots and individual raw data points. A pattern opposite to that with vowel duration can be noticed: closure duration is shorter for voiced than for voiceless stops. The closure of voiceless stops in Italian is 106.16 ms long (SD = 27.08), while the voiced stops have a mean closure duration of 117.66 ms (SD = 34.63). In Polish, the closure duration is 75.57 ms (SD = 16.16) in voiceless stops and 83.11 ms (SD = 19.37) in voiced stops. The difference in closure duration based on the raw means is 13.33 ms in Italian and 10.87 ms in Polish. The same model specification as with vowel duration has been fitted with consonant closure duration as the outcome variable. C2 voicing, C2 place, and speech rate are significant. Stop closure is 16.5 ms shorter (SE = 3) if the stop is voiced and 3.5 ms longer (SE = 1.5) if velar. Finally, faster speech rates correlate with shorter closure durations ($\hat{\beta}$ = -8.5 ms, SE = 1 ms).

3.3 Vowel and closure duration

A model addressing the relationship between vowel and stop closure duration was fitted with the following terms and interactions: vowel duration as the outcome variable; as fixed effects, closure duration, vowel, speech rate; an interaction between closure duration and vowel; by-speaker and by-word random intercepts, and by-speaker random slopes for C2 voicing. Closure duration has a significant effect on vowel duration ($\hat{\beta}$ = -0.15 ms, SE = 0.06 ms). The effect with /u/ is greater than with /a/ and /o/ ($\hat{\beta}$ = -0.35 ms, SE = 0.06 ms). In general, closure duration is inversely proportional to vowel duration. However, such correlation is quite weak, as shown by the small estimates. A 1 ms increase in closure duration corresponds to a 0.2–0.5 ms decrease in vowel duration. Figure 4 shows for each vowel /a, o, u/ the individual data points and the regression lines with confidence intervals extracted from the mixed-effects model.

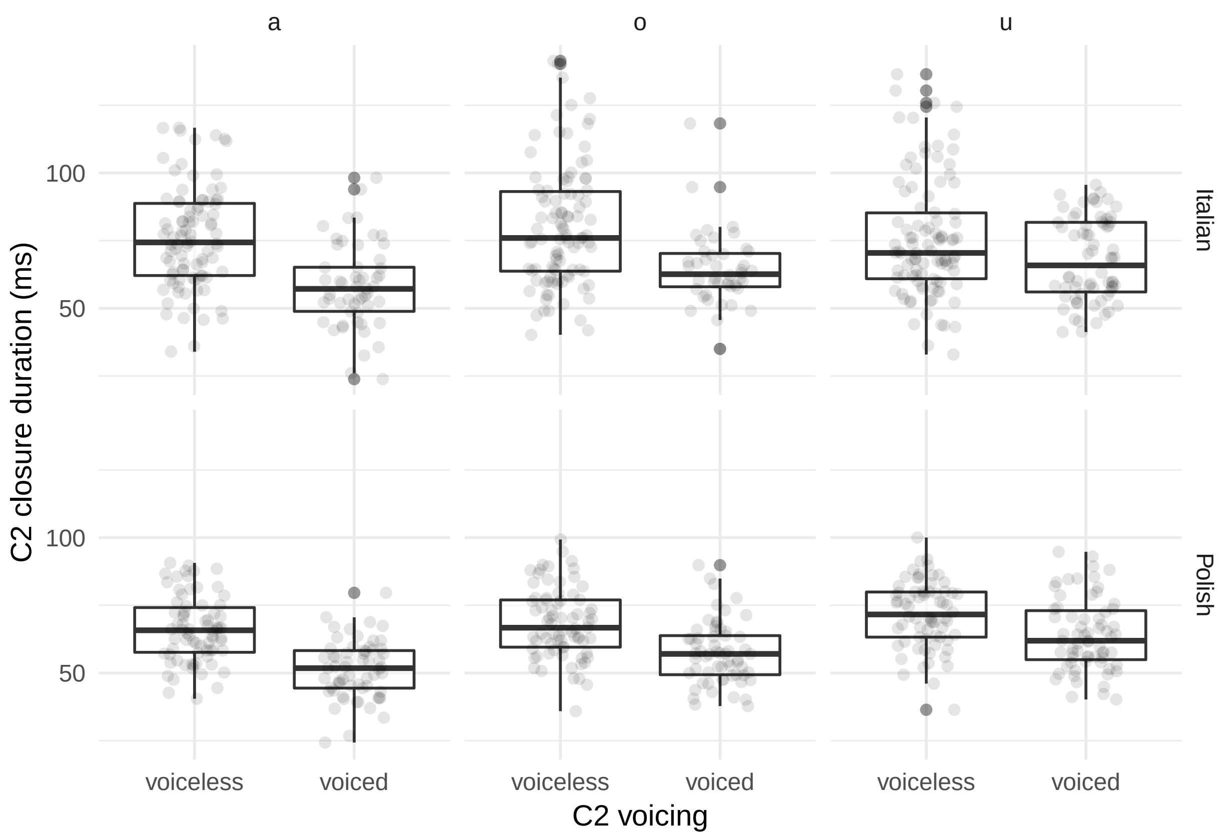


Figure 3: Raw data and boxplots of closure duration in milliseconds of voiceless and voiced stops in Italian (top row) and Polish (bottom row) when preceded by the vowels /a, o, u/.

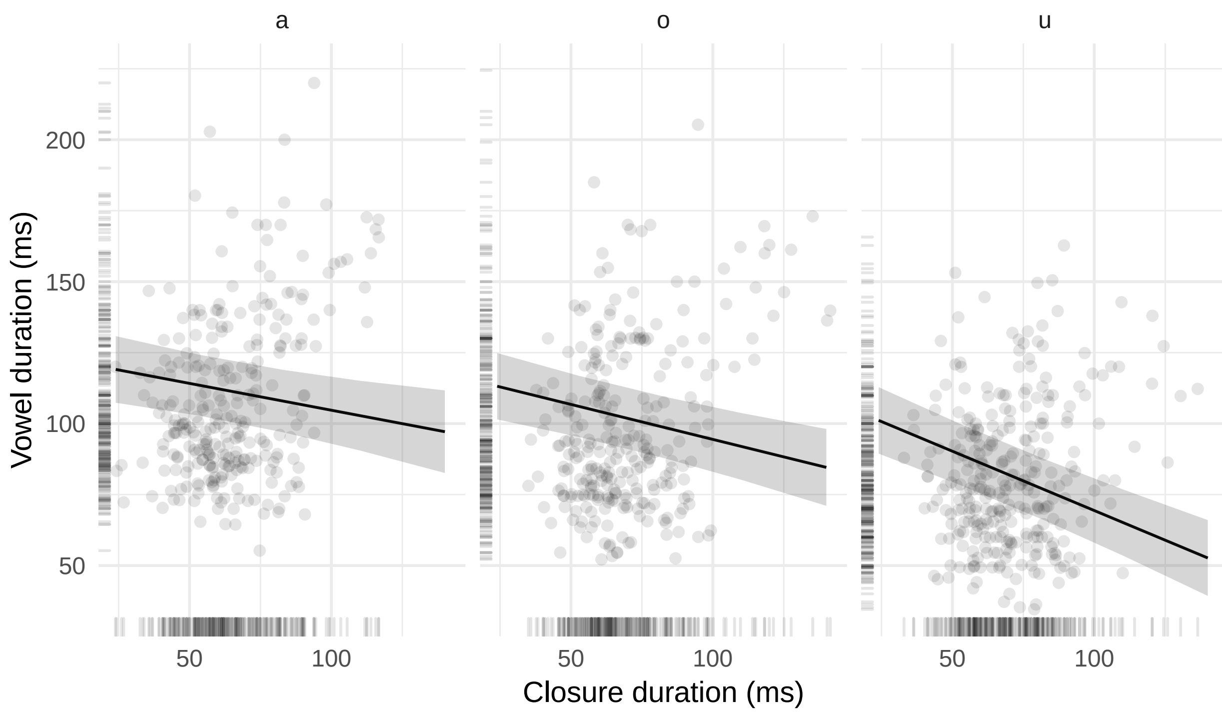


Figure 4: Raw data, estimated regression lines, and 95 per cent confidence intervals of the effect of closure duration on vowel duration for the vowels /a, o, u/ (from a mixed-effects model fitted to data pooled from Italian and Polish, see text for details).

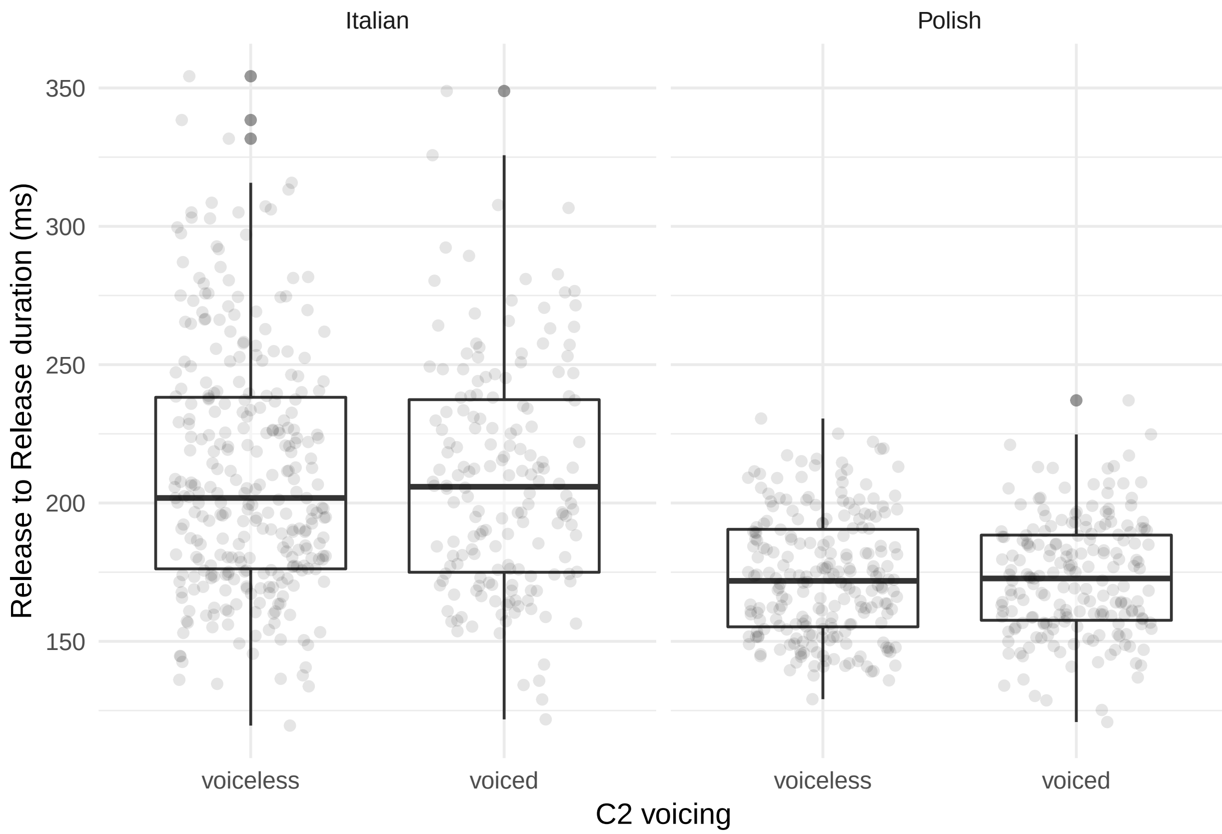


Figure 5: Raw data and boxplots of the duration in milliseconds of the Release to Release interval in Italian (left) and Polish (right) when C2 is voiceless or voiced.

3.4 Word duration

Words with a voiceless C2 are on average 397 ms long (SD = 81) in Italian and 356 ms long (SD = 39) in Polish. Words with a voiced stop have a mean duration of 396 ms (SD = 72) in Italian and 362 ms (SD = 39) in Polish. The following full and null models were fitted to test the effect of C2 voicing on word duration. The full model is made up of the following fixed effects: C2 voicing, C2 place, vowel, language, and speech rate. The model also includes by-speaker and by-word random intercepts, and a by-speaker random slope for C2 voicing. The null model is the same as the full model with the exclusion of the fixed effect of C2 voicing. The Bayes factor of the null against the full model is 19. Thus, the null model (in which there is no effect of C2 voicing, $\beta = 0$) is 19 times more likely under the observed data than the full model. This indicates that there is positive evidence for a null effect of C2 voicing on word duration.

3.5 Release to Release interval duration

In Figure 5, boxplots and raw data points show the duration of the Release to Release interval in words with a voiceless vs. a voiced C2 stop, in Italian and Polish. It can be seen that the distributions, medians, and quartiles of the durations in the voiceless and voiced condition do not differ much in either language. In Italian, the mean duration of the Release to Release interval is 210 ms (SD = 44) if C2 is voiceless, and 209 ms (SD = 41) if voiced. In Polish, the mean durations are respectively 173 (SD = 22) and 172 (SD = 21) ms. The specifications of the null and full models for the Release to Release duration are the same as for word duration. The Bayes factor of the null model against the full model is 21, which means that the null model (without C2 voicing) is 21 times more likely than the model with C2 voicing as a predictor. The Bayes factor suggests there is strong evidence that duration of the Release to Release interval is not affected by C2 voicing.

3.6 Summary

Seventeen participants were recorded while reading CV́CV words embedded in a frame sentence. The stressed vowel was either /a, o, u/, and C2 was one of /t, d, k, g/. Of the seventeen participants, 11 are native speakers of Italian and 6 of Polish. The durations of the vowel, stop closure, word, and Release to Release interval were measured from the acoustic signal. The analyses of the durational data suggest that:

(a) Stressed vowels in $C_1\acute{V}_1C_2V$ words in Italian and Polish are 19 ms longer (SE = 4.4) when C2 is voiced.

(b) C2 closure is 16.5 ms shorter (SE = 3) if the stop is voiced.

(c) V1 duration negatively correlates with closure duration, such that shorter closures correspond to longer vowels.

(d) Both word duration and Release to Release duration are not affected by the underlying voicing specification of C2.

4 Discussion

4.1 Voicing effect in Italian and Polish

4.2 Compensatory temporal adjustment

The data and statistical analyses of this exploratory study suggest that the duration of the interval between the releases of two consecutive consonants in CV́CV words (the Release to Release interval) is insensitive to the phonological voicing of the second consonant (C2) in Italian and Polish. In accordance with a compensatory temporal adjustment account (Slis & Cohen, 1969a; Lehiste, 1970b), the difference in vowel duration before voiceless vs. voiced stops can be seen as the outcome of differences in stop closure duration. More specifically, the timing of the closure onset of C2 within the invariant Release to Release interval determines the duration of the preceding vowel. An earlier closure onset relative to the onset of the preceding vowel (like in the case of voiceless stops) causes the vowel to be shorter. On the other hand, a later closure onset (like with voiced stops) produces a longer vowel. Figure 6 illustrates this mechanism.

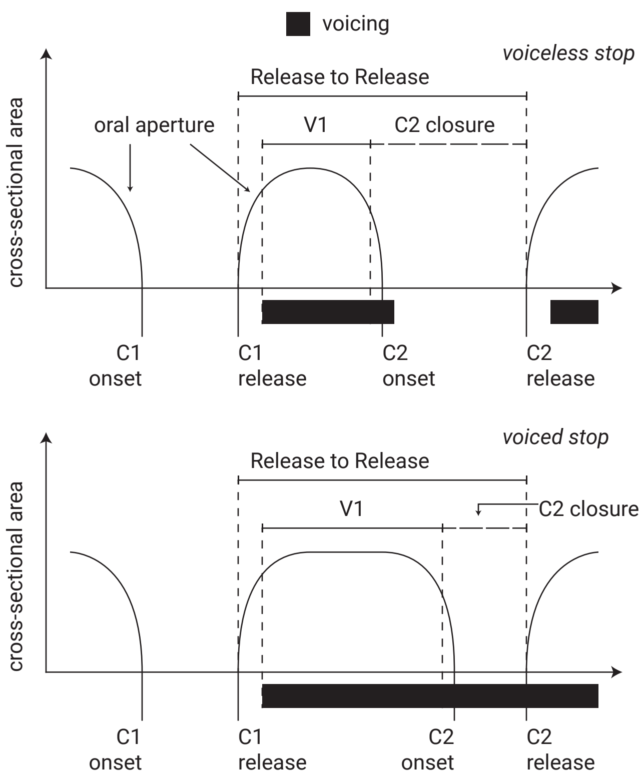


Figure 6: A schematic representation of the voicing effect as a compensatory temporal adjustment phenomenon. The schematic shows the gestural unfolding of a CV́C sequence when C2 is voiceless (top panel), or voiced (bottom panel). Oral cavity aperture (on the *y*-axis, as the inverse of oral constriction) through time (on the *x*-axis) is represented by the black line. Lower values represent a more constricted oral tract (a contoid configuration), while higher values indicate a more open oral tract (a vocoid configuration). The black bars below the time axis represent voicing (vocal fold vibration). Various landmarks and intervals are indicated in the schematic. Design based on Esposito (2002).

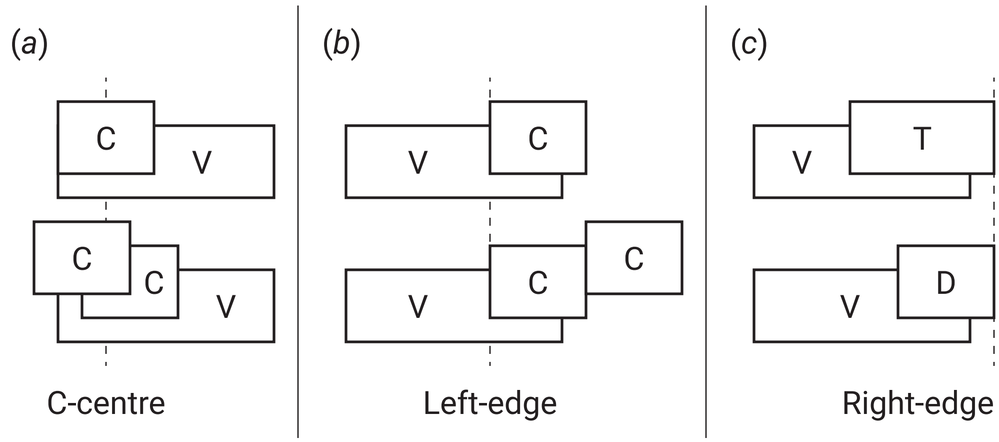


Figure 7: Gestural organisation patterns for onsets (a), codas (b), heterosyllabic onsets (c). C = consonant, V = vowel, T = voiceless stop, D = voiced stop. See Section 4.3 for details. Based on Marin & Pouplier (2010).

The invariance of the Release to Release interval allows us to refine the logistics of the compensatory account by narrowing the scope of the temporal adjustment action. A limitation of this account, as proposed by Slis & Cohen (1969a) and Lehiste (1970b), is the lack of a precise identification of the word-internal mechanics of compensation. As already discussed in Section 1, it is not clear, for example, why the adjustment should target the preceding stressed vowel, rather then the following unstressed vowel or any other segment in the word. Since the Release to Release interval includes just the vowel (broadly defined as a vocoid gesture) and the consonant closure, it follows that differences in closure duration must be reflected in differences in the duration of the preceding vowel. It is worth noting, though, that other accounts—which could be compatible with other aspects of production and perception—aren't ultimately ruled out (for example, perceptual factor might play a role in the enhancement of the effect, see Kingston & Diehl 1994, Port & Dalby 1982, and Luce & Charles-Luce 1985).

On the one hand, the voicing effect can be re-interpreted as a by-product of gestural timing, rather then a consequence of intrinsic features of voicing *per SE*, with a constant Release to Release interval as the explanans.

On the other hand, the Release to Release invariance is in turn an explanandum. In the following section, I offer a gestural organisation account that allows the invariance of such interval to follow from the relative timing of the articulatory gestures in a CV́C sequence.

4.3 Gestural alignment

According to the coupled oscillator model of syllabic structure (Browman & Goldstein, 1988, 2000; Goldstein et al., 2006; Goldstein & Pouplier, 2014), articulatory gestures can be timed according to two coupling modes: in-phase (synchronous) mode, by which two gestures start in synchrony, or anti-phase (sequential) mode, in which a gesture starts when the preceding one has reached its target. Marin & Pouplier (2010) showed that onset consonants in American English are in-phase with respect to the vowel nucleus and anti-phase with each other. Such phasing pattern establishes a stable relationship between the centre of the consonant (or consonants in a cluster) and the following vowel. Independent of the number of onset consonants, the temporal midpoint of the onset (the so-called ‘C-centre’) is maintained at a fixed distance from the vowel, such that an increasing number of consonants in the onset does not change the distance between the vowel and the onset C-centre (Figure 7a). On the other hand, coda consonants are timed anti-phase with the preceding vowel and between themselves. Temporal stability in codas is found in the lag between the vowel and the left-most edge of the coda, which is not affected by the number of coda consonants (Figure 7b). Other studies found further evidence for the synchronous and sequential coupling modes (see extensive review in Marin & Pouplier 2010 and Marin & Pouplier 2014), although the use of one mode over the other depends on the language and the consonants under study.

Consonants can thus be said to follow either a C-centre or a left-edge organisation pattern. In both cases, of course, the pattern is relative to the tautosyllabic vowel (the following vowel for onsets, the preceding vowel for codas). To the best of my knowledge, no study has reported the timing of onset consonants relative to the *preceding* (heterosyllabic) vowel. The results from this acoustic study on Italian and Polish are compatible with a right-edge organisation pattern for onset consonants relative to the preceding stressed vowel (Figure 7c). In CV́CV words, the timing of the C2 release (the acoustic parallel of the articulatory right edge of C2) is fixed relative to V1.

A consequence of a right-edge organisation pattern of C2 relative to V1 in CV́CV words is that differences in C2 closure duration do not affect the lag between V1 and the release of C2, as shown by the results of this study. The invariance of the lag between the release of C1 and that of C2 then can be seen to follow from the invariance in timing between, on the one hand, C1 (which is always /p/ in this study) and V1, and, on the other, between V1 and the right edge of C2.

A right-edge organisation account is compatible with findings from a variety of sources. Celata et al. (2018) show with ultrasound tongue imaging data that, in Italian, vowels followed by single consonants are longer than when followed by geminates (for example, /ba.ta/ vs. /bat.ta/). However, vowels followed by a tautosyllabic cluster have the same duration as vowels followed by a heterosyllabic cluster (/pa.tron/ vs. /bat.man/). Celata et al. (2018) argue that these results corroborate a rhythmic account in which the relevant unit is not the traditional syllable, but rather the VC(C) sequence. The importance of the VC(C) sequence as relevant speech unit has been previously recognised in the work of Farnetani & Kori (1986) (who called it the ‘rhythmic syllable’) and Steriade (2012) (who uses the term ‘vowel to vowel interval’, see also Hirsch 2014 and Lunden 2017). The duration of the rhythmic syllable or interval is constant across the phonological contexts, while the duration of the traditional syllable is not. This reflects a gestural organisation in which the timing of the right edge of the consonant is fixed relative to the vowel.

de Jong (1991) reports that the closing gesture of voiceless stops (following stressed vowels) is faster than that of voiced stops, and that also it is timed earlier with respect to the opening gesture of the stressed vowel. According to de Jong (1991), the differences in vowel duration are driven by the timing of the consonantal closing gesture relative to the vocalic opening gesture (also see Hertrich & Ackermann 1997). Moreover, the data in de Jong (1991) show that the final portion of the opening gesture is prolonged before voiced stops.

This pattern fits the one reported in an electromyographic study by Raphael (1975). The electromyographic signal corresponding to the vocalic gesture reaches its plateau at the same time relative to the preceding consonant in the voiceless and voiced context, but the plateau is held for longer in the case of vowels followed by voiced stops. This indicates that muscular activation in the vocalic gesture before voiced stops is held for longer. Raphael (1975) further notes that the durational difference in muscular activation corresponds to the difference in the acoustic duration of vowels before voiceless vs. voiced stops (see also Warren & Jacks 2005).

The results of the studies just discussed, together with the results from this study, bring support to a view in which two aspects of gestural organisation contribute to the difference in vowel duration observed before consonants varying in their voicing specification. These aspects are: (1) the *right-edge alignment* of coda consonants following a stressed vowel relative to the latter, and (2) the *differential timing of the closing gesture onset* for voiceless vs. voiced stops. The interplay of these two aspects can be synthesised into a compensatory temporal adjustment account, which requires a temporally constant interval, produced by (1), and a temporal reorganisation, brought about by (2).

4.4 Limitations and future work

The generalisations put forward in this paper strictly apply to disyllabic words with a stressed vowel in the first syllable. It is possible that the organisation pattern found in this context does not occur in sequences including an unstressed vowel. For example, it is known that the difference in closure duration between voiceless and voiced stops is not stable when the stops precede a stressed vowel, although vowels preceding pre-stress stops have slightly different durations (Davis & Van Summers, 1989). According to the gestural interpretation given here, the absence of differences in closure duration should correspond to no difference in vowel duration. Moreover, the magnitude of the effect is modulated by prosodic characteristics, like the number of syllables in the word, presence/absence of focus, and position within the sentence (Sharf, 1962; Klatt, 1973; Laeufer, 1992; de Jong, 2004). Data from different contexts and different languages is thus needed to assess the generality of the claims put forward in this paper.

The constraints on experimental material enforced by the use of ultrasound tongue imaging have been previously mentioned in Section 2.3. Given these constraints, temporal information from other vowels (like front vowels) and places of articulation is a desideratum. Section 4.3 discusses the interpretation of the Release to Release invariance in CV́CV words as a consequence of the timing of C2 rather than of a holistic CV́C motor plan in which the Release to Release interval is held constant. Although beyond the scope of this paper, disambiguating between these two interpretations on articulatory grounds is fundamental for a general understanding of a theory of gestural organisation (a promising venue of research might be the activation-spin model by Tilsen 2013).

The compensatory temporal adjustment account presented here extends to other durational effects discussed in the literature. In particular, the account bears predictions on the direction of the durational difference led by phonation types different from voicing, like aspiration and ejection. For example, the mix of results with regard to the effect of aspiration (Durvasula & Luo, 2012) suggests that the conditions for a temporal adjustment might differ across the contexts and languages studied. In light of the results in Beguš (2017), future studies will have to investigate the durational invariance of speech intervals in relation to a variety of phonation contrasts.

5 Conclusions

The results of an exploratory study on the effect of voicing on vowel duration are congruent with a compensatory temporal adjustment account of such effect. Acoustic data from seventeen speakers of Italian and Polish show that the temporal distance between two consecutive stop releases is not affected by the voicing of the second stop in CV́CV words. The temporal invariance of the Release to Release interval, together with a difference in stop closure duration of voiceless and voiced stops, causes vowels to be shorter when followed by voiceless stops (which have a long closure) and longer when followed by voiced stops (the closure of which is short). I proposed that the Release to Release invariance is a consequence of the gestural organisation of the CV́C sequence, in which the lag between the right-edge of the second consonant and the preceding stressed vowel is fixed.

6 Acknowledgements

I am grateful to Dr. Ricardo Bermúdez-Otero and Dr. Patrycja Strycharczuk for their immense support and patience in providing feedback on this project. I also want to thank the audience at the 16th Laboratory Phonology conference (LabPhon16) for their invaluable input. Thanks also go to my colleagues at the Phonetics Laboratory of the University of Manchester, who provided help in different ways. Any remaining errors are my own. This project has been funded by the School of Arts, Languages, and Cultures Graduate School at the University of Manchester.

A Output of statistical models

See Table 2, Table 3, and Table 4.

B Socio-linguistic information of participants

See Table 5.

C Target words

See Table 6.

Table 2: Summary of the linear mixed-effects model fitted to vowel duration (see Section 3.1).

Predictor	Estimate	SE	CI low	CI up	df	t-value	p-value	$< \alpha$
Intercept	118.0609	4.9378	108.3830	127.7388	23.8904	23.9096	0.0000	*
C2 voi: voiced	16.2843	4.4222	7.6169	24.9517	15.3849	3.6824	0.0021	*
Vow: /o/	-7.5001	3.9324	-15.2075	0.2072	10.3076	-1.9073	0.0847	
Vow: /u/	-25.7081	3.9437	-33.4376	-17.9786	10.4274	-6.5188	0.0001	*
Lang: Polish	-24.0486	7.8339	-39.4027	-8.6945	22.3757	-3.0698	0.0055	*
C2 place: velar	-7.9513	1.6330	-11.1519	-4.7507	10.9945	-4.8691	0.0005	*
Speech rate	-16.2317	1.2568	-18.6950	-13.7684	854.6318	-12.9150	0.0000	*
Voiced, /o/	2.0925	5.5437	-8.7730	12.9580	10.1785	0.3775	0.7136	
Voiced, /u/	-13.0949	5.5599	-23.9921	-2.1978	10.2979	-2.3553	0.0396	*
Voiced, Polish	-5.7345	6.6125	-18.6947	7.2257	17.9976	-0.8672	0.3972	
/o/, Polish	-2.4959	5.6636	-13.5963	8.6046	11.0882	-0.4407	0.6679	
/u/, Polish	1.1219	5.6812	-10.0129	12.2568	11.2268	0.1975	0.8470	
Voiced, /o/, Polish	-6.1626	8.0043	-21.8506	9.5255	11.0591	-0.7699	0.4575	
Voiced, /u/, Polish	6.3984	8.0285	-9.3372	22.1340	11.1938	0.7970	0.4420	

Table 3: Summary of the linear mixed-effects model fitted to closure duration (see Section 3.2).

Predictor	Estimate	SE	CI low	CI up	df	t-value	p-value	$< \alpha$
Intercept	73.2456	4.2802	64.8566	81.6346	22.3799	17.1127	0.0000	*
C2 voi: voiced	-17.7025	4.0614	-25.6626	-9.7423	18.6311	-4.3587	0.0004	*
Vow: /o/	3.7499	3.2579	-2.6354	10.1352	9.4323	1.1510	0.2781	
Vow: /u/	-1.9112	3.2686	-8.3175	4.4952	9.5605	-0.5847	0.5723	
Lang: Polish	-7.0293	6.8198	-20.3959	6.3373	20.8234	-1.0307	0.3145	
C2 place: velar	3.7966	1.3835	1.0849	6.5082	10.9434	2.7442	0.0192	*
Speech rate	-7.8611	1.1332	-10.0820	-5.6402	488.5467	-6.9374	0.0000	*
Voiced, /o/	1.9116	4.8762	-7.6456	11.4689	11.7988	0.3920	0.7020	
Voiced, /u/	10.8846	4.7870	1.5023	20.2669	10.9693	2.2738	0.0441	*
Voiced, Polish	2.3026	6.0659	-9.5863	14.1915	19.8271	0.3796	0.7083	
/o/, Polish	-1.0442	4.6670	-10.1914	8.1030	9.9357	-0.2237	0.8275	
/u/, Polish	6.9419	4.6845	-2.2396	16.1233	10.0874	1.4819	0.1689	
Voiced, /o/, Polish	1.3626	6.8398	-12.0432	14.7683	11.4396	0.1992	0.8456	
Voiced, /u/, Polish	-3.0766	6.7738	-16.3530	10.1998	11.0085	-0.4542	0.6585	

Table 4: Summary of the linear mixed-effects model fitted to vowel duration with closure duration as predictor (see Section 3.3).

Predictor	Estimate	SE	CI low	CI up	df	t-value	p-value	$< \alpha$
Intercept	123.6154	6.7649	110.3564	136.8745	56.2418	18.2730	0.0000	*
C2 closure	-0.1892	0.0647	-0.3160	-0.0625	816.5293	-2.9259	0.0035	*
Vow: /o/	-4.5422	6.3055	-16.9006	7.8163	127.4624	-0.7204	0.4726	
Vow: /u/	-12.4652	6.3960	-25.0012	0.0707	134.6411	-1.9489	0.0534	
Speech rate	-5.1595	4.2790	-13.5461	3.2271	827.0350	-1.2058	0.2282	
C2 closure, /o/	-0.0573	0.0805	-0.2150	0.1004	829.3804	-0.7125	0.4764	
C2 closure, /u/	-0.2287	0.0812	-0.3878	-0.0695	831.4914	-2.8160	0.0050	*
C2 closure, sp. rate	-0.1977	0.0621	-0.3194	-0.0760	826.9698	-3.1831	0.0015	*
/o/, sp. rate	-3.7489	5.1889	-13.9190	6.4212	819.7897	-0.7225	0.4702	
/u/, sp. rate	-10.1333	5.4982	-20.9097	0.6430	822.5482	-1.8430	0.0657	
C2 clos, /o/, sp. rate	0.0861	0.0736	-0.0582	0.2304	820.7357	1.1698	0.2424	
C2 clos, /u/, sp. rate	0.1694	0.0791	0.0144	0.3243	823.8831	2.1421	0.0325	*

Table 5: Participants' sociolinguistic information. The column 'Spent most time in' gives the city in which the participant spent most of their life. The last column ('> 6 mo') indicates whether the participant has spent more than 6 months abroad.

ID	Age	Sex	Native L	Other Ls	City of birth	Spent most time in	> 6 mo
IT01	29	Male	Italian	English, Spanish	Verbania	Verbania	Yes
IT02	26	Male	Italian	Friulian, English, Ladin-Venetan	Udine	Tricesimo	Yes
IT03	28	Female	Italian	English, German	Verbania	Verbania	No
IT04	54	Female	Italian	Calabrese	Verbania	Verbania	No
IT05	28	Female	Italian	English	Verbania	Verbania	No
IT09	35	Female	Italian	English	Vignola	Vignola	Yes
IT11	24	Male	Italian	English	Monza	Monza	Yes
IT12	26	Male	Italian	English	Rome	Rome	Yes
IT13	20	Female	Italian	English, French, Arabic, Farsi	Ancona	Chiaravalle	Yes
IT14	32	Male	Italian	English, Spanish	Frosinone	Frosinone	Yes
PL02	32	Female	Polish	English, Norwegian, French, German, Dutch	Koło	Poznań	Yes
PL03	26	Male	Polish	Russian, English, French, German	Nowa Sol	Poznań	Yes
PL04	34	Female	Polish	Spanish, English, French	Warsaw	Warsaw	No
PL05	42	Male	Polish	English, French	Przasnysz	Warsaw	No
PL06	33	Male	Polish	English	Zgierz	Zgierz	Yes
PL07	32	Female	Polish	English, Russian	Bielsk Podlaski	Bielsk Podlaski	Yes

Table 6: Target words. Asterisks indicate real words.

Italian			Polish		
pata	poto*	putu	pata	poto	putu
pada	podo	pudu	pada*	podo	pudu
paca*	poco*	pucu	paka*	poko	puku
paga*	pogo	pugu	paga	pogo	pugu

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.

Articulate Instruments Ltd™. 2011. Articulate Assistant Advanced user guide. Version 2.16.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. doi:10.18637/jss.v067.i01.

Beckman, Jill, Michael Jessen & Catherine Ringen. 2013. Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics* 49(02). 259–284.

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.

Berez-Kroeker, Andrea L, Lauren Gawne, Susan Smythe Kung, Barbara F Kelly, Tyler Heston, Gary Holton, Peter Pulsifer, David I Beaver, Shobhana Chelliah & Stanley Dubinsky. 2018. Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics* 56(1). 1–18. doi:10.1515/ling-2017-0032.

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

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. 2000. Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la communication parlée* (5). 25–34.

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, Chiara Meluzzi & Chiara Bertini. 2018. Stressed vowel durational variations and articulatory cohesiveness: Italian data. Poster presented at LabPhon 16, Lisbon.

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 [Research compendium]. Open Science Framework, https://osf.io/bfyhr/?view_only=391ef2dcc2834039a90f739ddb6f137a.

Crüwell, Sophia, Johnny van Doorn, Alexander Etz, Matthew Makel, Hannah Moshontz, Jesse Niebaum, Amy Orben, Sam Parsons & Michael Schulte-Mecklenbeck. 2018. 8 easy steps to open science: An annotated reading list. PsyArXiv. doi:10.31234/osf.io/cfzyx.

Cysouw, Michael & Jeff Good. 2013. Languoid, doculect, and glossonym: Formalizing the notion 'language'. *Language Documentation & Conservation* 7. 331–359. doi:10125/4606.

Davis, Stuart & W Van Summers. 1989. Vowel length and closure duration in word-medial VC sequences. *The Journal of the Acoustical Society of America* 17. 339–353. doi:10.1121/1.2026892.

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.

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.

Ferrero, Franco E, Emanuela Magno-Caldognetto, Kiryaki Vagges & C Lavagnoli. 1978. Some acoustic characteristics of Italian vowels. *Journal of Italian Linguistics Amsterdam* 3(1). 87–94.

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.

Fox, John. 2003. Effect displays in R for generalised linear models. *ournal of Statistical Software* 8(15). 1–27. doi:10.18637/jss.v008.i15.

Gelman, Andrew & Eric Loken. 2013. The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking"and the research hypothesis was posited ahead of time. Department of Statistics, Columbia University, http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf.

Goldstein, Louis, Dani Byrd & Elliot Saltzman. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In Michael A. Arbib (ed.), *Action to language via the mirror neuron system*, 215–249. Cambridge: Cambridge University Press.

Goldstein, Louis & Marianne Pouplier. 2014. The temporal organization of speech. In V. Ferreira, M. Goldrick & M. Miozzo (eds.), *The oxford handbook of language production*, Oxford: Oxford University Press.

Gussmann, Edmund. 2007. *The phonology of Polish*. Oxford University Press.

Hajek, John & Mary Stevens. 2008. Vowel duration, compression and lengthening in stressed syllables in central and southern varieties of standard italian, ISCA.

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.

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.

Hirsch, Aron. 2014. What is the domain for weight computation: the syllable or the interval? In *Proceedings of the Annual Meetings on Phonology*, 1–12.

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.

Hualde, José Ignacio & Marianna Nadeu. 2011. Lenition and phonemic overlap in Rome Italian. *Phonetica* 68(4). 215–242.

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.

Jarosz, Andrew F & Jennifer Wiley. 2014. What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving* 7(1). 2–9. doi:10.7771/1932-6246.1167.

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.

Keating, Patricia A. 1984. Universal phonetics and the organization of grammars. *UCLA Working Papers in Phonetics* 59.

Kerr, Norbert L. 1998. HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review* 2(3). 196–217. doi:10.1207/s15327957pspr0203_4.

Kingston, John & Randy L. Diehl. 1994. Phonetic knowledge. *Language* 419–454.

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.

Krämer, Martin. 2009. *The phonology of Italian*. Oxford: Oxford University Press.

Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen. 2017. `lmerTest` package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). doi:10.18637/jss.v082.i13.

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.

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.

Luke, Steven G. 2017. Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods* 49(4). 1494–1502. doi:10.3758/s13428-016-0809-y.

Lunden, Anya. 2017. Syllable weight and duration: a rhyme/intervals comparison. *Proceedings of the Linguistic Society of America* 2. 33–1. doi:10.3765/plsa.v2i0.4084.

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.

Malisz, Zofia & Katarzyna Klessa. 2008. A preliminary study of temporal adaptation in Polish VC groups. In *Proceedings of speech prosody*, 383–386.

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.

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.

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

Nowak, Pawel. 2006. *Vowel reduction in Polish*: University of California, Berkeley dissertation.

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. Talk presented at the 2018 Colloquium of the British Association of Academic Phoneticians (BAAP 2018).

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.

Raftery, Adrian E. 1995. Bayesian model selection in social research. *Sociological methodology* 111–163. doi:10.2307/271063.

Raftery, Adrian E. 1999. Bayes factors and BIC: Comment on "A critique of the Bayesian information criterion for model selection". *Sociological Methods & Research* 27(3). 411–427. doi:10.1177/0049124199027003005.

Raphael, Lawrence J. 1975. The physiological control of durational differences between vowels preceding voiced and voiceless consonants in English. *Journal of Phonetics* 3(1). 25–33.

Renwick, Margaret & Robert D. Ladd. 2016. Phonetic distinctiveness vs. lexical contrastiveness in non-robust phonemic contrasts. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1). 1–29. doi:10.5334/labphon.17.

Roettger, Timo B. 2018. Researcher degrees of freedom in phonetic sciences. Pre-print available at PsyArXiv. doi:10.31234/osf.io/fp4jr.

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.

Sóskuthy, Márton. 2013. *Phonetic biases and systemic effects in the actuation of sound change*: University of Edinburgh dissertation.

Steriade, Donca. 2012. Intervals vs. syllables as units of linguistic rhythm. Talk presented at EALING, Paris: http://ealing.cognition.ens.fr/ealing2012/handouts/Steriade/.

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.

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.

Vazquez-Alvarez, Yolanda & Nigel Hewlett. 2007. The 'trough effect': an ultrasound study. *Phonetica* 64(2-3). 105–121. doi:10.1159/000107912.

Wagenmakers, Eric-Jan. 2007. A practical solution to the pervasive problems of *p* values. *Psychonomic bulletin & review* 14(5). 779–804. doi:10.3758/BF03194105.

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.

Wickham, Hadley. 2017. `tidyverse`: Easily install and load the 'Tidyverse'. R package version 1.2.1. https://CRAN.R-project.org/package=tidyverse.