This is a title and this is too

Stefano Coretta[1]

The University of Manchester[a)]

Put your abstract here.

[a)] stefano.coretta@manchester.ac.uk; other info

I. 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 (Belasco, 1953; Chen, 1970; Durvasula and Luo, 2012; Esposito, 2002; Farnetani and Kori, 1986; Fowler, 1992; Halle and Stevens, 1967; Heffner, 1937; House and Fairbanks, 1953; Hussein, 1994; Javkin, 1976; Klatt, 1973; Kluender *et al.*, 1988; Laeufer, 1992; Lampp and Reklis, 2004; Lisker, 1974; Maddieson and Gandour, 1976; Peterson and Lehiste, 1960; Raphael, 1975; Warren and Jacks, 2005). 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 and 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*.

Several proposal have been put forward as to where to search for 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.[1] A notable production account, which will be the focus of this study, is the compensatory temporal adjustment account (Lehiste, 1970a,b; Lindblom, 1967; Slis and Cohen, 1969a,b). According to this account, the voicing effect derives 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. It is well known that the closure of voiceless stops is longer than that of voiced stops (Davis and Van Summers, 1989; De Jong, 1991; Lisker, 1957;

Van Summers, 1987). 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 proposed two prosodic units as the scope of the temporal adjustment: the syllable (or, more neutrally, the VC sequence Lindblom 1967), and the word (Lehiste, 1970a,b; Slis and Cohen, 1969a,b). 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 findings in Chen (1970) and Jacewicz *et al.* (2009). Chen (1970) rejects a syllable-based compensatory account on 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 dependent on the voicing of the coda consonant. Second, although the results in Slis and Cohen (1969b) 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, so 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 and Gandour, 1976), by which vowels are longer when followed by aspirated stops than when followed by non-aspirated 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 and 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 consonant are different. Maddieson and 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 and 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 and 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* was calculated as the interval between the closure of /t^h/ and the release of /k/). This measure includes the burst and (eventual) aspiration of the consonant following the target vowel. Slis and Cohen (1969a), however, states that the inverse correlation between vowel duration and the following consonant applies to *closure* duration, and not the entire *consonant* duration. If a correlation exists between vowel and closure duration, the inclusion of burst and/or aspiration duration clearly alters this relationship.

Indeed, the study on Hindi voicing and aspiration effects conducted by Durvasula and Luo (2012) indicates that closure duration, properly measured, decreases according to the hierarchy voiceless unaspirated > voiced unaspirated > voiceless aspirated > voiced aspirated, which closely resembles the order of increasing vowel duration in Maddieson and Gandour (1976). Nonetheless, Durvasula and 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 and Gandour (1976), and in face of the results in Slis and Cohen (1969b) 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.

A. 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 and 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.[2] 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.[3] Moreover, the design of the study has been constrained by the use of such articulatory techniques (see Section II). 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 of the effect of stop voicing on vowel duration. Italian has been unanimously reported as a voicing-effect language (Caldognetto *et al.*, 1979; Esposito, 2002; Farnetani and Kori, 1986). The mean difference in vowel duration when followed by voiceless vs. voiced consonants ranges between 22 and 24 ms, with longer vowels followed by voiced consonants.[4] On the other hand, 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. Moreover, Malisz and 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 confirms the existence of a voicing effect in Italian and Polish, and suggests 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 and Pouplier, 2014) and I discuss the insights such account provides in relation to our understanding of the gestural organisation of speech.

II. METHOD

A. Participants

Seventeen subjects in total participated in this exploratory 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 Poznań and 4 speakers 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.

B. 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. Recordings were made at sample rate of 22050 MHz (16-bit), in a proprietary format (`.aa0`). A FocusRight Scarlett Solo pre-amplifier and a Movo LV4-O2 Lavalier microphone were used for audio recording. The microphone was placed around at the level of the participant's mouth on one side, at a distance of 10 cm.[5]

C. 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.).[6] 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 and Hewlett 2007).

The target words were embedded in a frame sentence, *Dico X lentamente* 'I say X slowly' in Italian (following Hajek and 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 stress and emphasis similar across languages.

D. 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 for 6 times (with the exceptions of IT02, who repeated the 12 sentences 5 times, and IT07, with whom words containing /u/ were not recorded due to technical difficulties relating to ultrasound data collection).[7] with a grand total of 1224 tokens (792 from Italian, 432 from Polish).

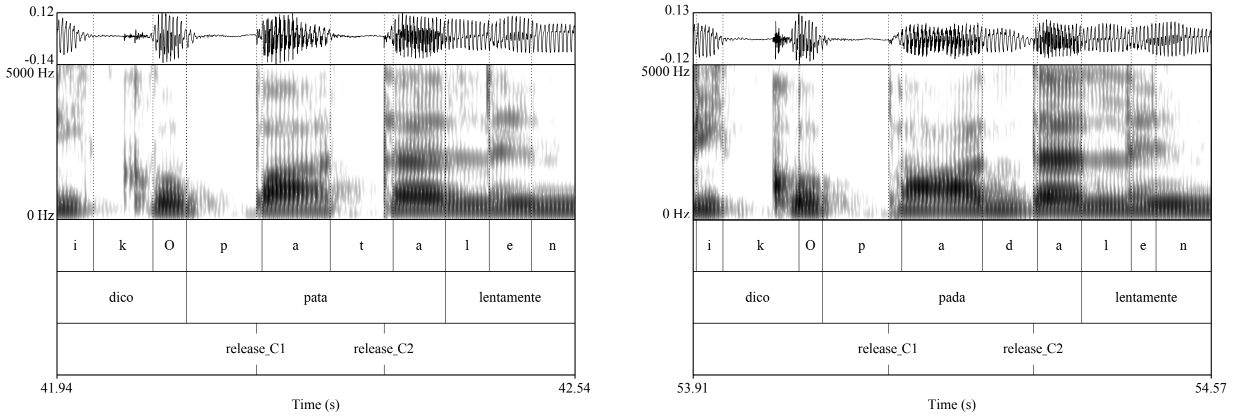
E. Data processing and measurements

The audio recordings were exported from AAA in `.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, according to the criteria in Table I. The releases of C1 and C2 were detected automatically by means of a Praat scripting implementation of the algorithm described

TABLE I. List of measurements as extracted from acoustics.

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)

in Ananthapadmanabha *et al.* (2014). 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. Figure 1 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 and Smith, 2018), and was calculated as the number of syllables divided by the duration of the sentence in seconds (8 syllables in Italian, 6



(a) (b)

FIG. 1. Segmentation example.

in Polish). All further data processing and visualisation was done in R v3.5.0 (R Core Team, 2018; Wickham, 2017).

F. 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) (Gelman and Loken, 2013; Kerr, 1998; 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 ($\alpha = 0.05$).[8]

Bayes factors were used to specifically test the null hypotheses that word and RR 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 BIC approximation was then used to obtain Bayes factors (Jarosz and Wiley, 2014; Raftery, 1995, 1999; Wagenmakers, 2007). 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).

The extracted measurements were filtered before statistical analysis. Measures of vowel duration, closure duration, word duration, and RR duration that are 3 standard deviations lower or higher than the respective means were excluded from the final dataset. This operation (which generally corresponds to a loss of around 2.5% of the data) yields a total of 920 tokens of vowel and closure durations, 1176 tokens of word duration, and 848 tokens of RR duration.

III. 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 cluttering generated by model parameters and alleviate the burden of the reader, the full output of statistical models and respective *p*-values are included in Appendix A.

A. 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, but the effect is discernible in Polish /u/. In Italian, vowels have a mean duration of 106 ms (sd = 27) before voiceless stops, and a mean duration of 118 ms (sd = 33) before voiced stops. Polish vowels are on average 75 ms long (sd = 16) when followed by a voiceless stop, and 83 ms long (sd = 19) if a voiced stop follows. The difference in vowel duration based on the raw means is 12 ms in Italian and 8 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 intercept with by-speaker random slopes for C2 voicing. All possible interactions between C2 voicing, vowel, and language were included. The following terms are significant

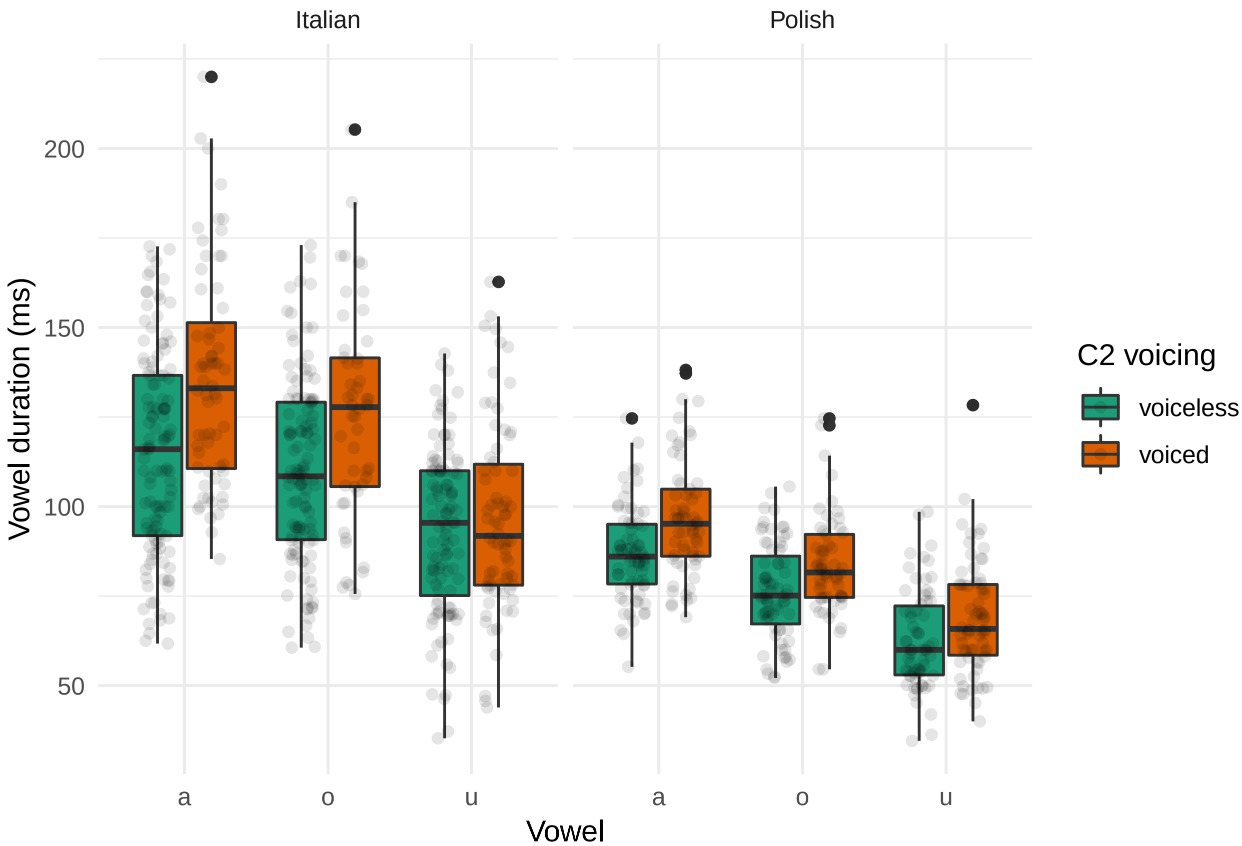


FIG. 2. Raw data and boxplots of the duration in milliseconds of vowels in Italian (left) and Polish (right), for the vowels /a, o, u/ when followed by a voiceless (green) or voiced (orange) stop.

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 19 ms longer (se = 4.4) when followed by a voiced stop (C2 voicing). The effect of C2 voicing is smaller with /u/ (around 5 ms, $\hat{\beta}$ = -14.4 ms, se = 6). Polish has on average shorter vowels than Italian ($\hat{\beta}$ = -28 ms, se = 8), and the effect of voicing is estimated to be about 11 ms (although note that the interaction between language and C2 voicing is deemed as not significant). Speech rate has a negative effect on vowel duration, such that faster rates correlate with shorter vowel durations ($\hat{\beta}$ = -15 ms, se = 1).

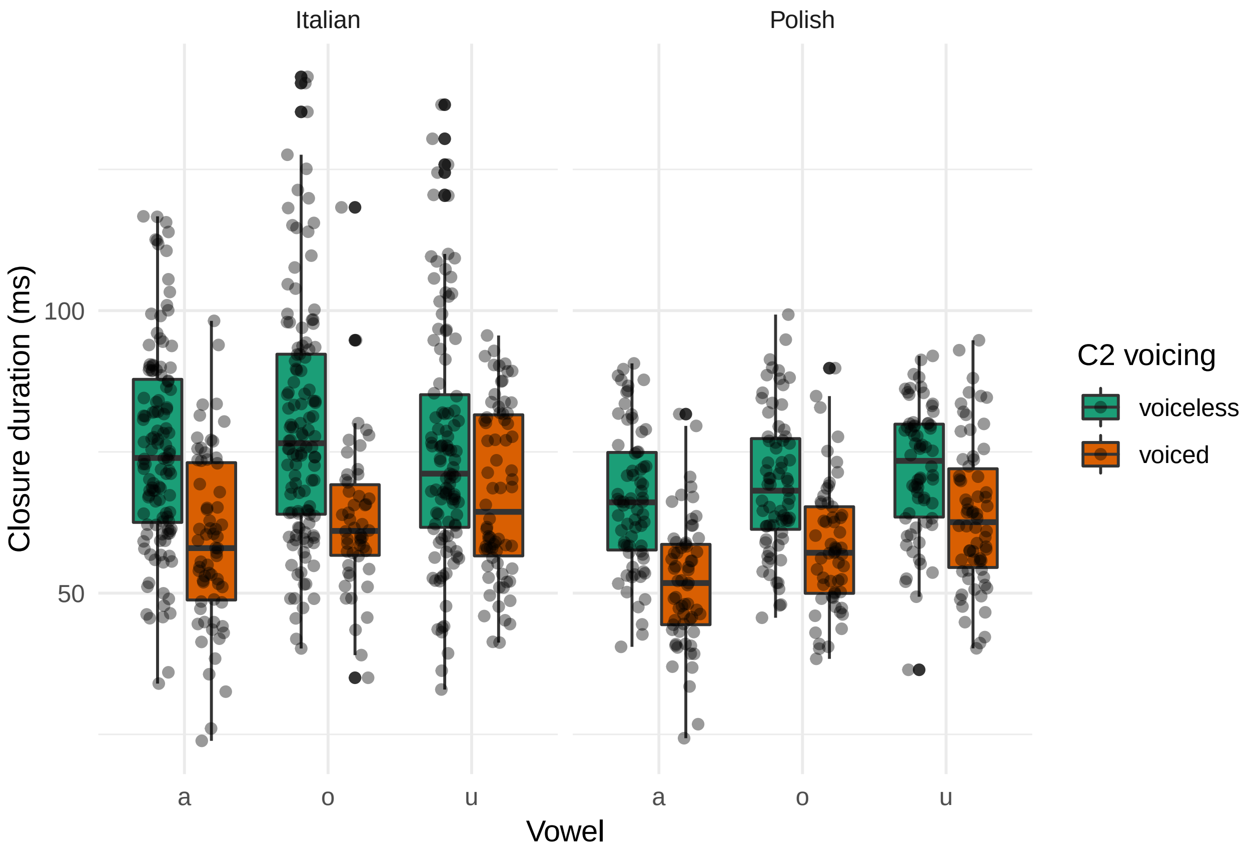


FIG. 3. Raw data and boxplots of closure duration in milliseconds of voiceless (green) and voiced (orange) stops in Italian (left) and Polish (right) when preceded by the vowels /a, o, u/.

B. 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 77 ms long (sd = 20), while the voiced stops have a mean closure duration of 63 ms (sd = 15). In Polish, the closure duration is 69 ms (sd = 12) in voiceless stops and 58 ms (sd = 13) in voiced stops. The difference in closure duration based on the raw means is 14 ms in Italian and 11 ms in Polish. The same model specification as with vowel duration has been fitted with consonant closure durations as the outcome variable. C2 voicing, C2

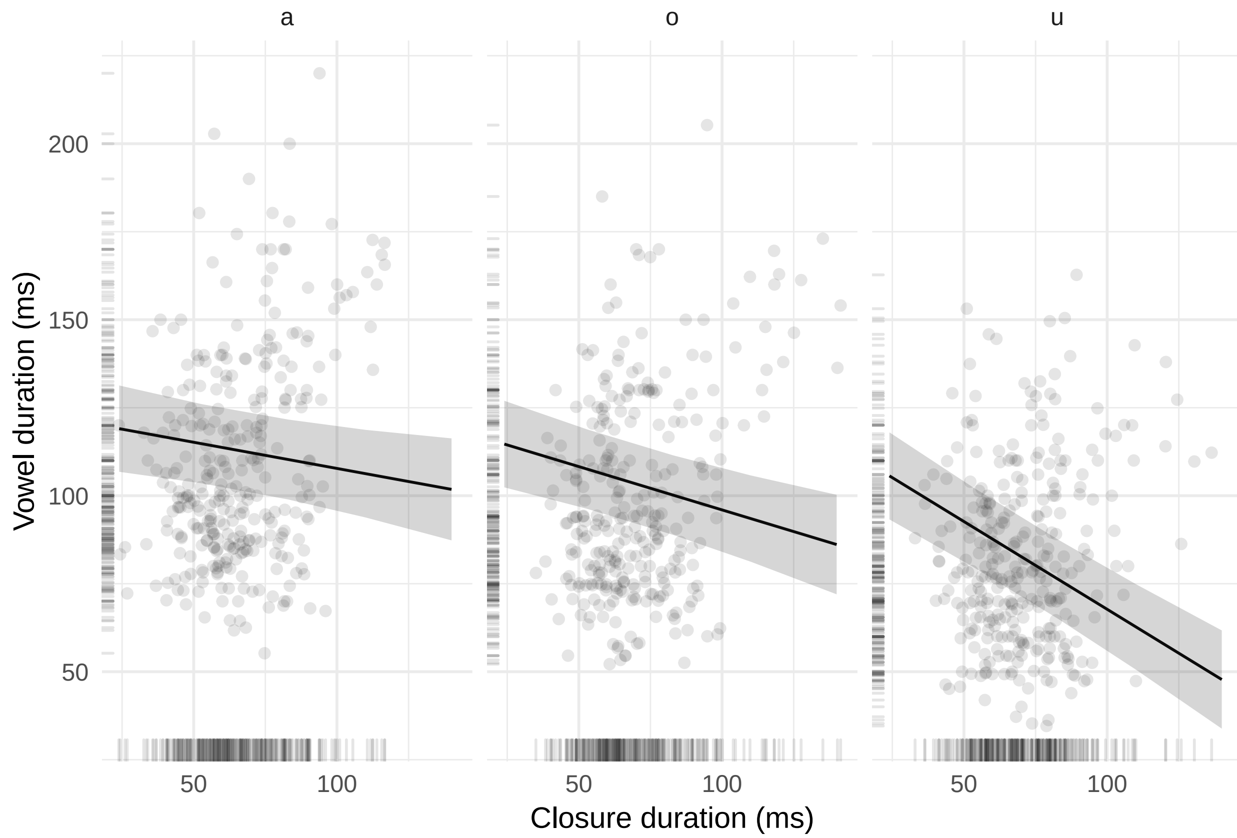


FIG. 4. Raw data and estimated regression lines of the effect of closure duration on vowel duration for the vowels /a, o, u/ (from a mixed-effects model fitted to data pooled from Italian and Polish).

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).

C. 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.

D. 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 24. Thus, the null model (in which there is no effect of C2 voicing, β = 0) is 24 times more likely under the observed data than the full model. This indicates that there is strong evidence for a null effect of C2 voicing on word duration.

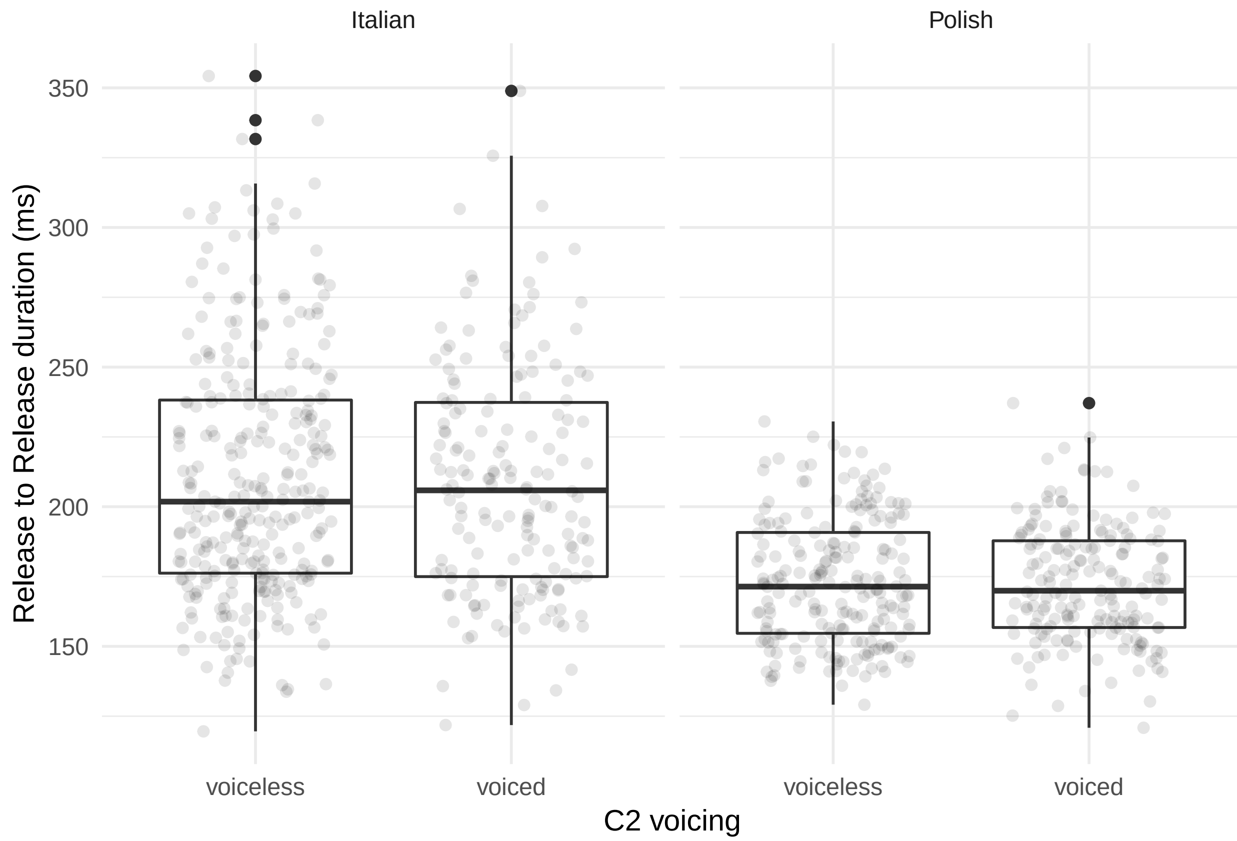


FIG. 5. Raw data and boxplots of the duration in milliseconds of the Release to Release interval in Italian (left) and Polish (right) when C2 is voiceless or voiced.

E. 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 23, which means that the null model (without C2 voicing) is 23 times more likely than the model with C2 voicing as a predictor. The data suggests there is positive evidence that duration of the RR interval is not affected by C2 voicing.

F. Summary

IV. DISCUSSION

The data and statistical analyses of this exploratory study suggest that the duration of 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 (Lehiste, 1970b; Slis and Cohen, 1969b), 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.

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 such account, as proposed by Slis and Cohen (1969b) and Lehiste (1970b), is the lack of a precise identification of the word-internal mechanics of compensation. As already discussed in Section I, it is not clear, for example, why the adjustment should target the preceding stressed vowel, rather then the

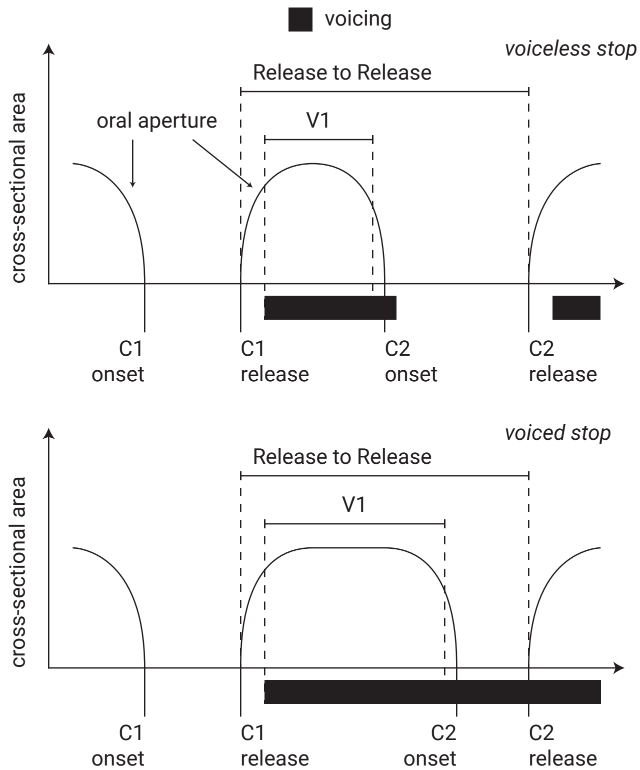


FIG. 6. A schematic representation of the voicing effect as a compensatory temporal adjustment phenomenon. The schematic show the gestural unfolding of a CV́C sequence when C2 is voiceless (top panel), or voiced (bottom panel). Oral cavity aperture (on the *y*-axis, as the inverse of oral constriction) through time (on the *x*-axis) is represented with a changing black line that represents the movement trajectory of an articulator. 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.

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.

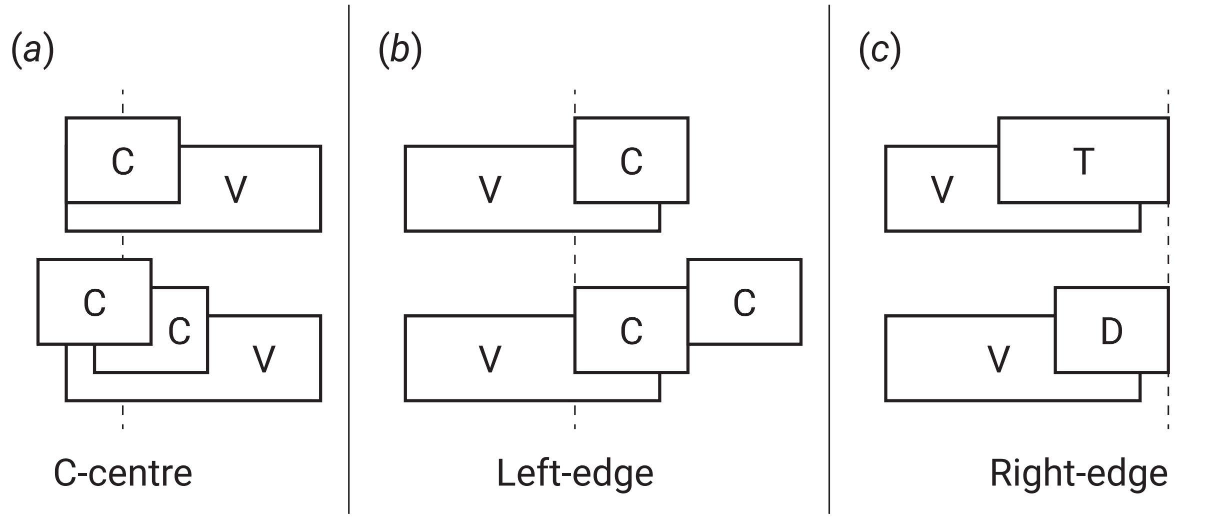


FIG. 7. Gestural organisation patterns for onsets (a), codas (b), heterosyllabic onsets (c). See Section IV A for details. Based on Marin and Pouplier (2010).

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.

A. Gestural alignment

According to the coupled oscillator model of syllabic structure (Browman and Goldstein, 1988, 2000; Goldstein *et al.*, 2006; Goldstein and 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 one gesture starts when the preceding one has reached its target. Marin and 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 consonant cluster and the following vowel. Independent of the number of onset consonants, the 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 C-centre to vowel distance (Figure 7a). On the other hand, coda consonants are timed anti-phase with the preceding vowel and between themselves. Stability in codas is seen 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 and Pouplier 2010 and Marin and 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 organisation pattern 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 and preceding stressed vowels (Figure 7c). The release of C2 (which is the onset of the second syllable in CV́CV words)—which can be thought as the acoustic parallel of the articulatory right edge of C2—is invariantly timed relative to V1 (which is the nucleus of the first syllable).

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 electromyographic, x-ray microbeam, and ultrasonic data by, respectively, Raphael (1975), De Jong (1991), and Celata *et al.* (2018). Celata *et al.* (2018) show that vowels before tautosyllabic clusters have the same duration as before heterosyllabic clusters. However, vowels followed by geminates are shorter than when followed by singletons, although from a syllabic structure point of view geminates correspond to heterosyllabic clusters and singletons to tautosyllabic clusters (i.e., V-final syllables followed by singletons and tautosyllabic clusters are open, while those followed by geminates and heterosyllabic clusters are closed). Celata *et al.* (2018) argue that these results corroborate a rhythmic account in which the relevant unit is the rhythmic syllable, i.e. the VC(C) sequence (independent of the traditional syllabic structure), which is kept constant. Such view reflects a gestural timing view in which the timing of the right edge of the consonant is held constant 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 and Ackermann 1997). Moreover, the data in De Jong (1991) show that the final portion of the vocalic opening gesture is prolonged before voiced stops. This finding corresponds to what Raphael (1975) reported based on electromyographic data. The electromyographic signal corresponding to the vocalic gesture reaches its plateaux at the same time in the voiceless and voiced context, but the plateaux is held for longer in the case of vowels followed by voiced stops, indicating that muscular activation is kept for longer.

These studies taken together, plus the results from this study, bring evidence to the view that two factors contribute to the difference in vowel duration observed before consonants varying in their voicing specification. These two factors are: (1) the right-edge alignment of coda consonants following stressed vowels relative to the latter, and (2) the differential timing of the closing gesture onset for voiceless vs. voiced stops. These two factors together can be synthesised into a compensatory temporal adjustment account, in which the fixed interval is generated by factor (1) and the temporal adjustment is brought about by factor (2).

B. Limitations and future work

The generalisations reported 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 the vowels preceding the pre-stress stops have different durations (Davis and Van Summers, 1989). According to the gestural interpretation given here, the absence in differences of closure duration should correspond to no difference in vowel duration. 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 II C. Given these constraints, temporal information from other vowels (like front vowels) and places of articulation is a desideratum. Section IV A 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 RR 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.

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 and 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.

V. CONCLUSION

ACKNOWLEDGMENTS

Thanks to...

APPENDIX A: OUTPUT OF STATISTICAL MODELS

1. Vowel duration

term	estimate	std.error	df	statistic	p.value	conf.low	conf.high
(Intercept)	202.5289	8.6169	134.7948	23.5036	0.0000	185.6400	219.4178
c2_phonationvoiced	18.9669	4.3898	12.7785	4.3207	0.0009	10.3631	27.5707
vowelo	-6.1457	3.9512	8.6900	-1.5554	0.1555	-13.8899	1.5985
vowelu	-26.3039	3.9772	8.9199	-6.6136	0.0001	-34.0991	-18.5087
languagePolish	-24.2194	8.1708	21.7230	-2.9642	0.0072	-40.2338	-8.2050
c2_placevelar	-8.1827	1.6984	10.5938	-4.8178	0.0006	-11.5116	-4.8539
syl_rate	-15.2920	1.2679	775.7483	-12.0608	0.0000	-17.7771	-12.8070
c2_phonationvoiced:vowelo	-2.0453	5.8662	10.5314	-0.3487	0.7342	-13.5428	9.4522
c2_phonationvoiced:vowelu	-14.4536	5.8040	10.0977	-2.4903	0.0318	-25.8292	-3.0780
c2_phonationvoiced:languagePolish	-7.9928	6.4252	14.2528	-1.2440	0.2336	-20.5860	4.6005
vowelo:languagePolish	-3.6121	5.7389	9.6704	-0.6294	0.5437	-14.8601	7.6360
vowelu:languagePolish	1.6149	5.7695	9.8777	0.2799	0.7853	-9.6931	12.9230
c2_phonationvoiced:vowelo:languagePolish	-2.9987	8.3627	10.8862	-0.3586	0.7268	-19.3894	13.3920
c2_phonationvoiced:vowelu:languagePolish	7.9601	8.3077	10.6040	0.9582	0.3593	-8.3227	24.2428

2. Closure duration

term	estimate	std.error	df	statistic	p.value	conf.low	conf.high
(Intercept)	119.7338	7.2100	128.2742	16.6065	0.0000	105.6023	133.8652
c2_phonationvoiced	-16.5825	4.3129	17.8144	-3.8449	0.0012	-25.0356	-8.1294
vowelo	3.6830	3.4951	9.0918	1.0538	0.3192	-3.1672	10.5333
vowelu	-1.9898	3.5174	9.3243	-0.5657	0.5849	-8.8837	4.9041
languagePolish	-6.9400	6.8688	22.0443	-1.0104	0.3233	-20.4027	6.5226
c2_placevelar	3.4024	1.4976	10.9532	2.2719	0.0443	0.4672	6.3376
syl_rate	-8.4278	1.0550	557.6472	-7.9887	0.0000	-10.4954	-6.3601
c2_phonationvoiced:vowelo	1.1040	5.1738	10.8916	0.2134	0.8350	-9.0364	11.2445
c2_phonationvoiced:vowelu	9.9882	5.1257	10.4981	1.9486	0.0786	-0.0581	20.0344
c2_phonationvoiced:languagePolish	1.6759	6.5019	20.0145	0.2578	0.7992	-11.0675	14.4194
vowelo:languagePolish	-0.2681	5.0672	10.0440	-0.0529	0.9588	-10.1997	9.6635
vowelu:languagePolish	7.1432	5.0932	10.2505	1.4025	0.1903	-2.8393	17.1256
c2_phonationvoiced:vowelo:languagePolish	1.5022	7.3707	11.2269	0.2038	0.8422	-12.9441	15.9485
c2_phonationvoiced:vowelu:languagePolish	-3.2088	7.3279	10.9696	-0.4379	0.6700	-17.5711	11.1536

3. Vowel and closure duration

term	estimate	std.error	df	statistic	p.value	conf.low	conf.high
(Intercept)	219.3142	10.4477	123.5512	20.9917	0.0000	198.8371	239.7913
closure_duration	-0.1487	0.0632	50.3807	-2.3532	0.0226	-0.2726	-0.0249
vowelo	-2.0462	5.4702	81.5530	-0.3741	0.7093	-12.7675	8.6751
vowelu	-5.0236	5.5582	86.7938	-0.9038	0.3686	-15.9176	5.8703
syl_rate	-17.5364	1.2855	896.1529	-13.6415	0.0000	-20.0559	-15.0168
closure_duration:vowelo	-0.0973	0.0615	876.5971	-1.5835	0.1137	-0.2178	0.0231
closure_duration:vowelu	-0.3500	0.0619	895.3921	-5.6582	0.0000	-0.4712	-0.2288

TABLE II. Participants' sociolinguistic information.

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
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 III. Target 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

APPENDIX B: SOCIO-LINGUISTIC INFORMATION OF PARTICIPANTS

APPENDIX C: TARGET WORDS

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

[2]As per **?**, the glossonyms *Italian* and *Polish* as used here refer, respectively, to the languoids Italian [Glottocode: `ital1282`] and Polish [Glottocode: `poli1260`].

[3]To the best of my knowledge, this is the first attempt to gather synchronised acoustic, tongue imaging and electroglottographic data in relation to the voicing effect.

[4]The mean differences are based on 3 speakers in Farnetani and Kori 1986 and 7 speakers in Esposito 2002. Caldognetto *et al.* 1979 does not report estimates of vowel duration, but the study is based on 10 speakers.

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

[6]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 (**?**), although both their phonemic status and their phonetic substance are subject to a high degree of geographical and idiosyncratic variability **?**. 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] **?**. On the

other hand, Polish has only a mid-low back vowel phoneme /ɔ/ (**?**). For sake of typographical simplicity, the symbol /o/ will be used here for both languages.

[7] 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.

[8] 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'.

Ananthapadmanabha, T. V., Prathosh, A. P., and Ramakrishnan, A. G. (**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.

Articulate Instruments Ltd™ (**2011**). "Articulate Assistant Advanced user guide. Version 2.16" .

Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," Journal of Statistical Software **67**(1), 1–48.

Beguš, G. (**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, S. (**1953**). "The influence of force of articulation of consonants on vowel duration," The Journal of the Acoustical Society of America **25**(5), 1015–1016.

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

Browman, C. P., and Goldstein, L. (**1988**). “Some notes on syllable structure in articulatory phonology,” Phonetica **45**(2-4), 140–155.

Browman, C. P., and Goldstein, L. (**2000**). “Competing constraints on intergestural coordination and self-organization of phonological structures,” Bulletin de la communication parlée (5), 25–34.

Caldognetto, E. M., Ferrero, F., Vagges, K., and Bagno, M. (**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, C., Meluzzi, C., and Bertini, C. (**2018**). “Stressed vowel durational variations and articulatory cohesiveness: Italian data” Poster presented at LabPhon 16, Lisbon.

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

Davis, S., and Van Summers, W. (**1989**). “Vowel length and closure duration in word-medial VC sequences,” The Journal of the Acoustical Society of America **17**, 339–353.

De Jong, K. (**1991**). “An articulatory study of consonant-induced vowel duration changes in english,” Phonetica **48**(1), 1–17.

Durvasula, K., and Luo, Q. (**2012**). “Voicing, aspiration, and vowel duration in Hindi,” Proceedings of Meetings on Acoustics **18**, 1–10.

Esposito, A. (**2002**). “On vowel height and consonantal voicing effects: Data from Italian,” Phonetica **59**(4), 197–231.

Farnetani, E., and Kori, S. (**1986**). “Effects of syllable and word structure on segmental durations in spoken Italian,” Speech communication **5**(1), 17–34.

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

Gelman, A., and Loken, E. (**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 .

Goldstein, L., Byrd, D., and Saltzman, E. (**2006**). “The role of vocal tract gestural action units in understanding the evolution of phonology,” in *Action to Language via the Mirror Neuron System*, edited by M. A. Arbib (Cambridge: Cambridge University Press), pp. 215–249.

Goldstein, L., and Pouplier, M. (**2014**). “The temporal organization of speech,” in *The Oxford handbook of language production*, edited by V. Ferreira, M. Goldrick, and M. Miozzo (Oxford: Oxford University Press).

Hajek, J., and Stevens, M. (**2008**). “Vowel duration, compression and lengthening in stressed syllables in central and southern varieties of standard italian,” ISCA.

Halle, M., and Stevens, K. (**1967**). “Mechanism of glottal vibration for vowels and consonants,” The Journal of the Acoustical Society of America **41**(6), 1613–1613.

Heffner, R.-M. (**1937**). “Notes on the length of vowels,” American Speech **12**, 128–134.

Hertrich, I., and Ackermann, H. (**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.

House, A. S., and Fairbanks, G. (**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.

Hussein, L. (**1994**). "Voicing-dependent vowel duration in Standard Arabic and its acquisition by adult american students," Ph.D. thesis, The Ohio State University.

Jacewicz, E., Fox, R. A., and Lyle, S. (**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, A. F., and Wiley, J. (**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, H. 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.

Keating, P. A. (**1984**). "Universal phonetics and the organization of grammars," UCLA Working Papers in Phonetics **59**.

Kerr, N. L. (**1998**). "HARKing: Hypothesizing after the results are known," Personality and Social Psychology Review **2**(3), 196–217.

Klatt, D. H. (**1973**). "Interaction between two factors that influence vowel duration," The Journal of the Acoustical Society of America **54**(4), 1102–1104.

Kluender, K. R., Diehl, R. L., and Wright, B. A. (**1988**). "Vowel-length differences before voiced and voiceless consonants: An auditory explanation.," Journal of Phonetics **16**, 153–169.

Kuznetsova, A., Bruun Brockhoff, P., and Haubo Bojesen Christensen, R. (**2017**). "`lmerTest` package: Tests in linear mixed effects models," Journal of Statistical Software **82**(13), doi:

10.18637/jss.v082.i13.

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

Lampp, C., and Reklis, H. (**2004**). “Effects of coda voicing and aspiration on Hindi vowels,” The Journal of the Acoustical Society of America **115**(5), 2540–2540.

Lehiste, I. (**1970**a). “Temporal organization of higher-level linguistic units,” The Journal of the Acoustical Society of America **48**(1A), 111–111.

Lehiste, I. (**1970**b). “Temporal organization of spoken language,” in *Working Papers in Linguistics*, Vol. 4, pp. 96–114.

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

Lisker, L. (**1957**). “Closure duration and the intervocalic voiced-voiceless distinction in English,” Language **33**(1), 42–49.

Lisker, L. (**1974**). “On “explaining” vowel duration variation,” in *Proceedings of the Linguistic Society of America*, pp. 225–232.

Luke, S. 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.

Maddieson, I., and Gandour, J. (**1976**). “Vowel length before aspirated consonants,” in *UCLA Working papers in Phonetics*, Vol. 31, pp. 46–52.

Malisz, Z., and Klessa, K. (**2008**). “A preliminary study of temporal adaptation in Polish vc groups,” in *Proceedings of Speech Prosody*, pp. 383–386.

Marin, S., and Pouplier, M. (**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, S., and Pouplier, M. (**2014**). "Articulatory synergies in the temporal organization of liquid clusters in Romanian," Journal of Phonetics **42**, 24–36.

Nowak, P. (**2006**). "Vowel reduction in Polish," Ph.D. thesis, University of California, Berkeley.

Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in english," The Journal of the Acoustical Society of America **32**(6), 693–703.

Plug, L., and Smith, R. (**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.

Raftery, A. E. (**1995**). "Bayesian model selection in social research," Sociological methodology 111–163.

Raftery, A. 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.

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

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

Slis, I. H., and Cohen, A. (**1969**a). "On the complex regulating the voiced-voiceless distinction I," Language and speech **12**(2), 80–102.

Slis, I. H., and Cohen, A. (**1969**b). “On the complex regulating the voiced-voiceless distinction II,” Language and speech **12**(3), 137–155.

Sóskuthy, M. (**2013**). “Phonetic biases and systemic effects in the actuation of sound change,” Ph.D. thesis, University of Edinburgh.

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, Y., and Hewlett, N. (**2007**). “The ‘trough effect’: an ultrasound study,” Phonetica **64**(2-3), 105–121.

Wagenmakers, E.-J. (**2007**). “A practical solution to the pervasive problems of p values,” Psychonomic bulletin & review **14**(5), 779–804.

Warren, W., and Jacks, A. (**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.

Wickham, H. (**2017**). “tidyverse: Easily install and load the ’tidyverse”’ R package version 1.2.1., https://CRAN.R-project.org/package=tidyverse.