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 100 years of research have repeatedly 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). Evidence for such so called 'voicing effect' has been found in a 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, still after 100 years agreement hasn't been reached regarding the source of this effect.

Several proposal have been put forward as to where to look for the possible cause of the voicing effect (see Maddieson and Gandour, 1976, and Sóskuthy (2013) for an overview). Most of the proposed accounts place the source of the voicing effect in properties of speech production.[1] One of these production accounts, which will be the focus of this study, relates the voicing effect to some constant property of speech that is held constant across contexts while the local property of voiceless vs. voiced obstruents varies, thus creating a trade-off solution within the constant property. Lindblom (1967), Slis and Cohen (1969b), and Lehiste (1970b) (among others) argue that the relevant invariant property of speech is a constant durational interval within which segments of different duration results in different duration of other segments. Both the syllable/VC sequence (Lindblom, 1967) and the word (Lehiste, 1970a,b; Slis and Cohen, 1969a,b) has been proposed as the fixed interval. The closure of voiced stops is shorter than that of voiceless stops. It follows that vowels followed by shorter closures (like in the case of voiced stops) are longer than vowels followed by longer closures (like in the case of voiceless stops). However, the compensatory temporal adjustment account has been criticised in several occasions.

The proposal of the syllable or the word as the targets for compensation and encounter difficulties when confronted with empirical evidence and when scrutinised by logic. First, Lindblom's (1967) argument that the syllable is the interval within which compensation happens is not supported by the 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. More recently, Jacewicz *et al.* (2009) further shows that the duration of monosyllabic words in American English does not change dependent on the voicing of the coda consonant. Second, although Slis and Cohen (1969b) results confirm that the word does not change in duration whether the stop following the stressed vowel is voiceless or voiced, it does not follow from this that compensation should necessarily happen on the stressed vowel. Indeed, it could also happen in the following unstressed vowel.

Maddieson and Gandour (1976) reject any compensatory account for the voicing effect based on data from Hindi on the so called 'aspiration effect', by which vowels tend to be longer when followed by aspirated stops than when followed by non-aspirated stops. Vowels before voiceless unaspirated stops are the shortest, followed by vowels before voiced unaspirated and voiceless aspirated stops, which have similar duration between each other, 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 /tʰ/ have the same duration although the durations of the two consonant are different.

However, an reevaluation of the way consonant duration is measured in Maddieson and Gandour (1976) might actually turn the situation in favour of a compensatory account. Consonant duration is in fact measured from the closure of the relevant consonant to the release of the following consonant, due to difficulties in detecting the release of the consonant of interest (e.g., in *ab sāth kaho*, the duration of /tʰ/ in *sāth* was calculated as the interval between the closure of /tʰ/ and the release of /k/). This measure includes the burst and (eventual) aspiration of the consonant. Slis and Cohen (1969a), however, states that the inverse correlation between vowel duration and the following consonant raises when consonant *closure* duration is taken into account, and not the entire *consonant* duration. If the correlation exists between vowel and closure duration, the inclusion of burst/aspiration duration clearly alters this relationship. Indeed, the data in Durvasula and Luo (2012) show that closure duration, appropriately measured, decreases from voiceless unaspirated > voiced > voiceless aspirated > voiced aspirated, which closely resembles the order of increasing vowel duration in Maddieson and Gandour (1976).[2]

To summarise, a compensatory temporal adjustment account of the voicing effect is possible after a careful review of the critiques advanced by Chen (1970) and Maddieson and Gandour (1976), although issues about the actual implementation of the compensation still remain. In conclusion, for the compensatory account to gain plausibility, an invariant interval within which compensation is implemented needs to be better defined, on the light of empirical data.

A. The present study

This paper reports on results from a broader exploratory study that investigates the relationship between vowel duration and consonant voicing from an 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 (CVCV) words in Italian and Polish. The design of the study has been constrained by the use of these articulatory techniques (see Section II). Moreover, given the exploratory nature of the study, the experimental design was not implemented to directly test the compensatory account. Here, only the results from acoustic will be discussed.

Italian and Polish reportedly differ in the magnitude of the voicing effect. 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, Esposito, 2002; Farnetani and Kori, 1986).[3] On the other hand, the results regarding the presence and magnitude of the effect in Polished 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 to be inconclusive on this matter. The difference in presence or magnitude of the voicing effect in Italian vs. Polish should enable us to find an underlying property that differs in the two languages and that might indicate a possible source for the voicing effect.

The acoustic data from the exploratory study reported here reveal that the duration of the interval between the releases of the two consonants in CVCV words (the Release to Release interval) is not affected by the voicing of the second consonant. 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 the vowel and the stop closure. I further propose that the invariant duration of the Release to Release interval is congruent with current views on gestural timing (Goldstein and Pouplier, 2014) and I discuss the insights it provides in relation to our understanding of gestural organisation in speech.

II. METHOD

A. Participants

Seventeen subjects in total participated to this exploratory study. Eleven participants were native speakers of Italian (5 female, 6 male), while six were native speakers of Polish (3 female, 3 male). The Italian speakers were from the North and Centre of Italy (8 speakers from Northern Italy, 3 from Central Italy). The Polish group had 2 speakers from Poznań and 4 speakers from Eastern Poland. For more information on the speakers, see Appendix A. Ethical clearance was obtained for this study from the University of Manchester (REF 2016-0099-76). The participants signed a written consent and received a monetary compensation.

B. Equipment

The acquisition of the audio signal was achieved with the software Articulate Assistant Advanced™ (AAA, v2.17.2) running on a Hawlett-Packard ProBook 6750b laptop with Microsoft Windows 7, with a sample rate of 22050 MHz (16-bit) in a proprietary format. A FocusRight Scarlett Solo pre-amplifier and a Movo LV4-O2 Lavalier microphone were used for audio recording.

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.). 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. These sentences were chosen in order to keep the placement of stress and emphasis similar across languages, so to ensure comparability of results.

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 sentence stimuli was randomised for each participant. Each participant 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. 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 the ultrasound data collection).[4] with a grand total of 1224 tokens (792 from Italian, 432 from Polish). The reading task lasted between 15 and 20 minutes, with optional short breaks between one repetition and the other.

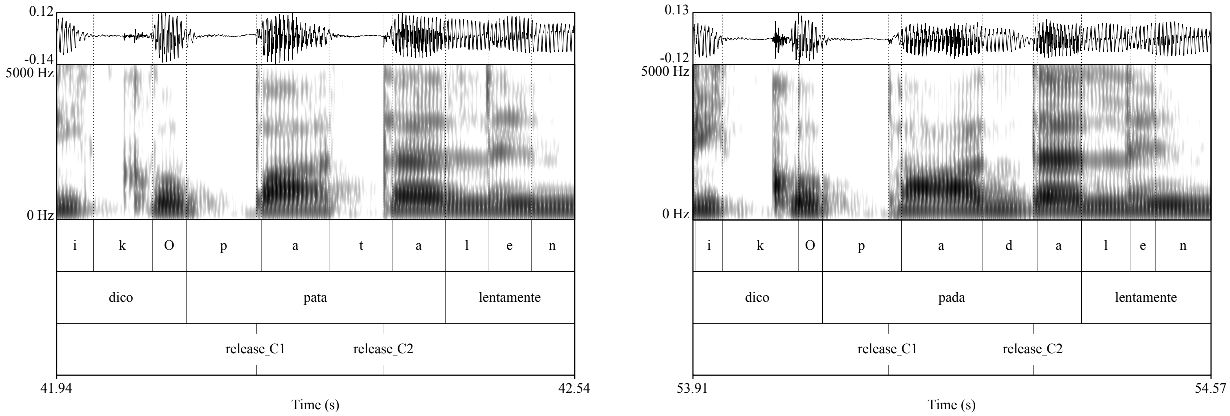
E. Data processing and measurements

The audio recordings were exported from AAA in `.wav` format 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 when necessary, 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 in Ananthapadmanabha *et al.* (2014). The durations in milliseconds of the following intervals were extracted from the annotated acoustic landmarks with Praat scripting: sentence duration, word duration, vowel duration (V1 onset to V1 offset), consonant closure duration (V1 offset to C2 burst), and Release-to-Release duration (RR duration, C1 release to C2 release). Figure 1 shows an exam-

TABLE I. List of measurements as extracted from acoustics.

landmark		criteria
vowel onset	(V1 onset)	appearance of higher formants in the spectrogram following the burst 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)

ple 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) for duration normalisation, and was calculated as the number of syllables divided by the duration of the sentence (8 syllables in Italian, 6 in Polish).



(a) (b)

FIG. 1. Segmentation example.

F. Statistical analysis

Given the exploratory nature of the study, all statistical analyses reported here are to be considered data-driven or hypothesis-generating rather than hypothesis-driven (Gelman and Loken, 2013; Kerr, 1998). The durational measurements were analysed with linear mixed-effects models using `lme4` v1.1-17 in R v3.5.0 (Bates *et al.*, 2015; R Core Team, 2018). All factors were coded with treatment contrasts. *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; **?**). *P*-values below the alpha level 0.05 were considered significant. The estimates of the relevant effects are then calculated by refitting the models including only the significant terms (step-down approach, Diggle *et al.*, 2002; Zuur *et al.*, 2009, pp. 121–122). 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 Maximum Likelihood estimation (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 yields a total of 920 tokens of vowel and closure durations, 1176 tokens of word duration, and 848 tokens of RR duration.

III. RESULTS

Only the most relevant terms will be presented. For the others see tables and appendixes.

A. Vowel duration

A linear mixed-effects model was fitted with the following terms: vowel duration as the outcome variable; fixed effects for C2 voicing (voiceless, voiced), 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.[5] All possible interactions between C2 voicing, vowel, and language

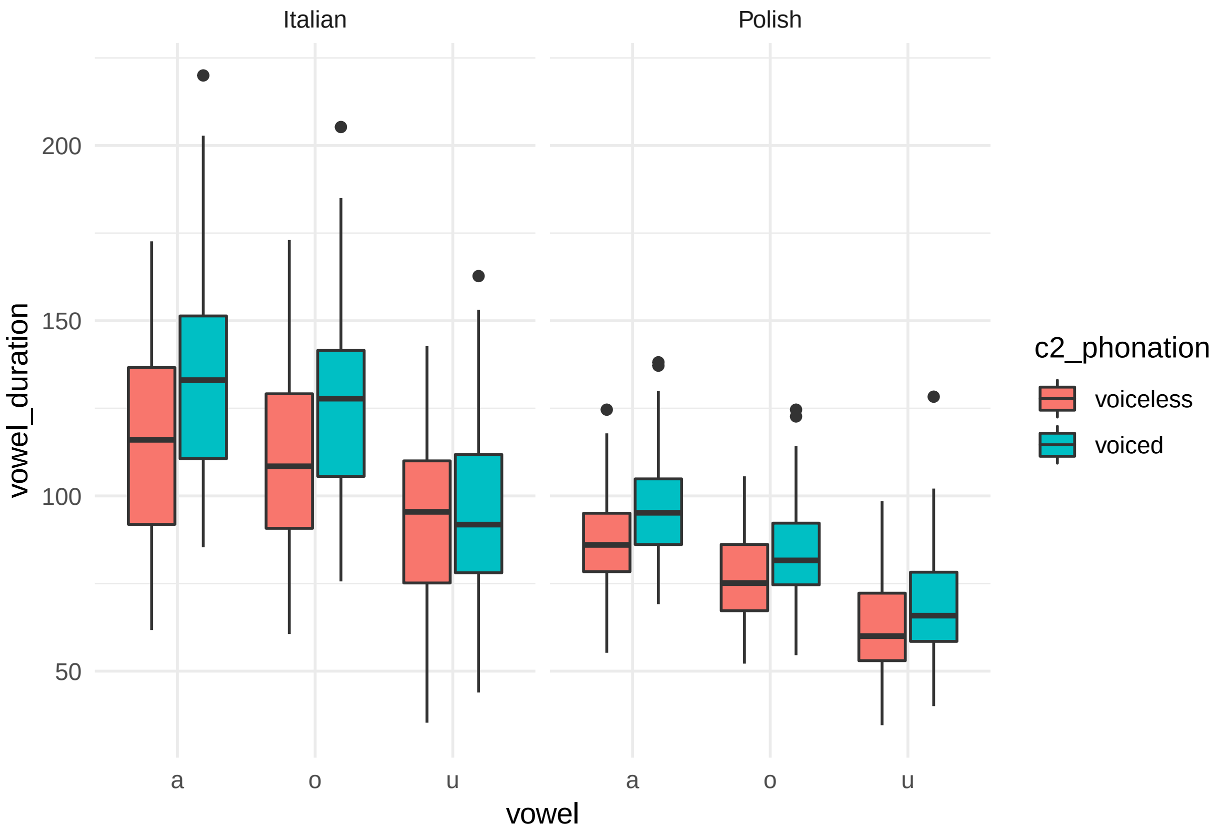


FIG. 2. Vowel duration in Italian and Polish.

were included. The following terms are significant: C2 voicing, vowel, language, and speech rate. No interaction was significant. The vowel /a/ (when followed by voiceless stops) has a duration of 202.5 ms (se = 8.5 ms). The vowel /o/ is 9.5 ms shorter (se = 3 ms), while the vowel /u/ is 30.5 ms shorter (se = 3). Vowels are 11 ms longer (se = 3) when followed by a voiced stop. Polish has on average shorter vowels than Italian ($\hat{\beta}$ = -28 ms, se = 8 ms), although the effect of voicing is estimated to be the same in both languages (the interaction of 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 ($\hat{\beta}$ = -15 ms, se = 1 ms).

B. Consonant closure duration

The same maximally specified model as with vowel duration has been fitted to consonant closure durations as the outcome variable. Only C2 voicing and speech rate were significant. Stop closure is 13 ms shorter (se = 3 ms) if the stop is voiced. Finally, faster speech rates correlate with shorter closure durations ($\hat{\beta}$ = -8 ms, se = 1 ms).

C. Vowel and closure duration

The full model was specified 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 when the vowel is /a/ ($\hat{\beta}$ = -0.2 ms, se = 0.06 ms). The effect with the vowel /o/ does not significantly differ from the one with /a/, while with /u/ the effect is -0.5 ms ($\hat{\beta}$ = -0.3 ms, se = 0.06 ms). In general, then, closure duration is inversely correlated with vowel duration after controlling for speech rate. However the correlation is quite weak. A 1 ms increase in closure duration corresponds to a 0.2–0.5 ms decrease in vowel duration.

D. Word duration

The following full and null models were fitted to test whether word duration is affected by C2 voicing. In the full model, I entered as fixed effects: C2 voicing, C2 place, vowel, speech rate, and language. The model also included by-speaker and by-word random intercepts, plus a by-speaker

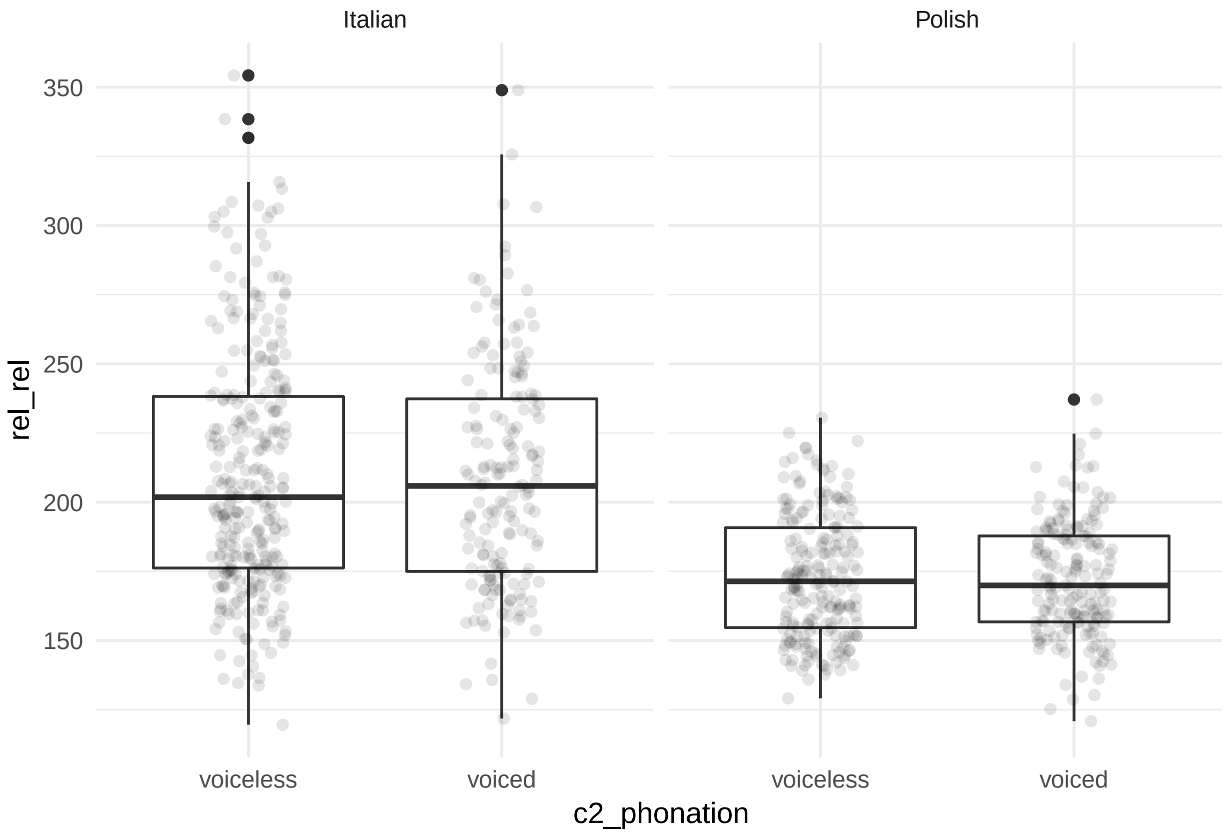


FIG. 3. Release to Release interval duration.

random slope for C2 voicing. The null model was the same as the full model, with the exclusion of the fixed effect of C2 voicing. The Bayes factor of the null model against the full model is 21.5. Thus, the null model (in which the effect of C2 voicing is 0) is 21.5 times more likely under the observed data than the full model. This indicates that there is strong evidence for word duration not being affected by C2 voicing.

E. Release to Release interval (RR) duration

The models specifications for the RR duration were the same as for word duration. The Bayes factor of the null model against the full model for RR duration is 19, which means that the null model

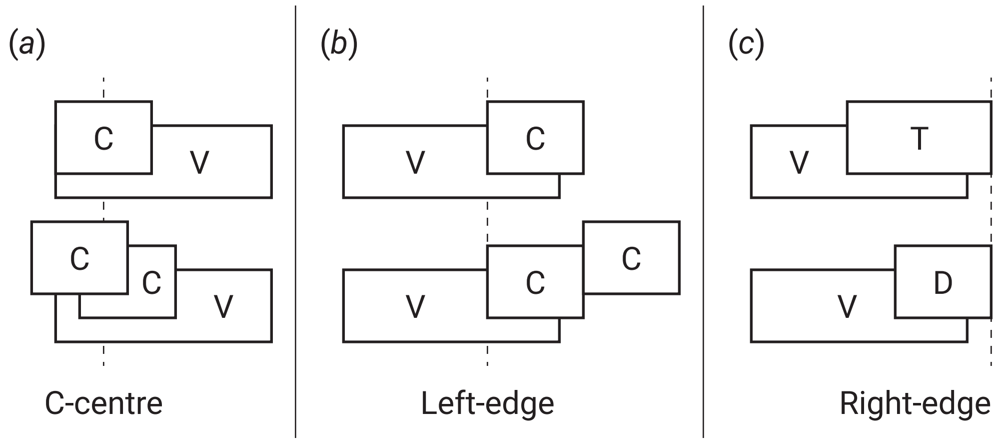


FIG. 4. Gestural organisation patterns for onsets (a), codas (b), heterosyllabic onsets (c). See Section IV A for details. Based on Marin and Pouplier (2010).

(without C2 voicing) is 19 times more likely than the full model. The data suggests there is positive evidence that duration of the RR interval is not affected by C2 voicing.

IV. DISCUSSION

In this exploratory study of Italian and Polish acoustic data, I found that the duration of interval between the releases of two consecutive consonants is insensitive to the phonological voicing of the second consonant. The difference in vowel duration before voiceless vs. voiced stops derives from differences in placement of the closure onset within the fixed interval between the releases of C1 and C2.

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 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 increasing number of consonants in the onset does not change the C-centre/vowel distance (Figure 4(a)). 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 4(b)). 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 4(c). 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 stress on 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, no differences in closure durations should correspond to no difference in vowel durations. The constraints on experimental material brought about by the use of ultrasound tongue imaging have been already mentioned in Section II C. In the previous section I mention that the invariance of the RR duration could be a consequence of the timing of C2 rather than of a holistic CVC motor plan in which the RR interval is held constant. Disambiguating between these two interpretations is not possible based on the data from this study.

V. CONCLUSION

ACKNOWLEDGMENTS

Thanks to...

APPENDIX A: SOCIO-LINGUISTIC INFORMATION OF PARTICIPANTS

[1] Two accounts that point to perceptual features are Javkin (1976) and Kluender *et al.* (1988). To the best of my knowledge, Javkin (1976)'s proposal remains empirically untested, while see Fowler (1992) for arguments against Kluender *et al.* (1988).

[2] Durvasula and Luo (2012) does 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 including an voicing and aspi-

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

ration as fixed effects. However, it is likely that this result is a consequence of not controlling for speech rate, so it will not be discussed here.

[3]These estimates should be taken as a gross approximation. There are several issues: number of speakers, different contexts, statistical modelling.

[4]IT01 and IT02 (the first two participants of this study) read also 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.

[5]The place of articulation of C2 was not included in any of the models discussed. This decision is based on the visual inspection of the data, which suggests C2 place does not have major effects on the outcome variables. Preliminary models including the effect of C2 place and interactions with other terms showed the model fit indeed does not improve.

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.

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.

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.

Diggle, P., , Heagerty, P., Liang, K.-Y., and Zeger, S. (**2002**). *Analysis of longitudinal data* (Oxford University Press).

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.

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

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.

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.

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.

Zuur, A. F., Saveliev, A. A., Ieno, E. N., Smith, G. M., and Walker, N. (**2009**). *Mixed Effects Models and Extensions in Ecology with R* (Springer).