

This is a title and this is too

Stefano Coretta¹

The University of Manchester^{a)}

¹ 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; Laeuffer, 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 (1969a), Slis and Cohen (1969b), and Lehiste (1970) (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) and the word (Lehiste, Slis) 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, [Chen \(1970\)](#) and [Maddieson and Gandour \(1976\)](#) criticise the compensatory temporal adjustment account on empirical grounds. [Chen \(1970\)](#) shows that the duration of the syllable is affected by consonant voicing (compatible with findings in [Jacewicz *et al.*, 2009](#)), contrary to Lindblom's expectations. [Maddieson and Gandour \(1976\)](#) reject any compensatory account based on data from a phenomenon parallel to the voicing effect, the aspiration effect, by which vowel tend to be longer when followed by aspirated stops than when followed by non-aspirated stops. They measured consonant duration and they found no compensatory pattern in relation to vowel duration: the consonant /t/, which has the shortest duration, is preceded by the shortest vowel, and the vowels before /d/ and /t/ have the same duration although the durations of the two consonant are different.

They show that word duration is not affected by voicing of C2 but they don't discuss the internal structure of the word. I will show that the Release to Release is invariant and that this is compatible with a gestural timing in which the C2 is right-edge aligned with C1/V. I will also offer an interpretation of [Maddieson and Gandour \(1976\)](#) that is compatible with a compensatory temporal adjustment account.

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 Hewlett-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 six 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 72 sentences, with a grand total of 576 tokens (288 per language). The reading task lasted between 15 and 20 minutes, with optional short breaks between one repetition and the other.

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)

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). 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).

F. Statistical analysis

The durational measurements were analysed with linear mixed-effects models using `lme4` v1.1-17 in R v3.5.0 ([Bates *et al.*, 2015](#); [Team, 2018](#)). All factors were coded as 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,](#)

110 1995; Wagenmakers, 2007; ?). The approximation is calculated according to the equation
 111 in 1 (Wagenmakers, 2007, p. 796).

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

112 where $\Delta BIC_{10} = BIC_1 - BIC_0$, BIC_1 is the BIC of the full model, and BIC_0 is the BIC
 113 of the null model. Values of $BF_{01} > 1$ indicate a preference of H_0 over H_1 . The interpretation
 114 of the Bayes factors follows the recommendations in Raftery (1995, p. 139).

115 III. RESULTS

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

117 A. Vowel duration

118 A linear mixed-effects model was fitted with the following terms: vowel duration as the
 119 outcome variable; fixed effects for C2 voicing, C2 place of articulation, vowel identity, lan-
 120 guage, and speech rate (as syllables per second); by-speaker and by-word random intercept
 121 with by-speaker random slopes for C2 voicing. All logical interactions were included. The
 122 following terms and interactions are significant: C2 voicing, vowel identity, language, speech
 123 rate, the interaction between C2 voicing and vowel, and C2 place and vowel. The vowels /a,
 124 o/ are 14 ms longer (se = 3 ms) when followed by a voiced stop. The effect is smaller for
 125 /u/ (around 6 ms; $\hat{\beta} = -8$ ms, se = 4 ms). Polish has on average shorter vowels than Italian
 126 ($\hat{\beta} = -25.5$ ms, se = 7.5 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. The following terms and interactions were significant: C2 voicing, C2 place of articulation, vowel identity, language, and interactions between language and C2 place, language and vowel identity, C2 voicing and place, C2 voicing and vowel, and a three-way interaction between C2 voicing, place and vowel identity. Stop closure is 15 ms shorter ($se =$) if the stop is voiced. Vowel identity has an effect on closure duration in voiced stops, but not in voiceless stops, and more so in voiced velar than in voiced coronal stops: closure after /a/ is the shortest, while after /u/ is the longest, with closure after /o/ in the middle.

C. Vowel and closure duration

The model with the following specification was fit to assess the correlation between vowel and closure durations: vowel duration as the outcome variable; fixed effects for closure duration, vowel identity, C2 place of articulation, and speech rate; all interactions between closure duration, vowel, and C2 place; by-speaker and by-word random intercepts, and by-speaker random slope for closure duration. The results of terms which were included in the analysis of vowel duration in XXX will not be discussed here (see Supplementary material).

Closure duration has a non-significant effect on vowel duration when the vowel is /a/ ($\hat{\beta} = -0.2$ ms, se = 0.06 ms) and /o/ ($\hat{\beta} = -0.25$ ms, se = 0.06 ms), but it has a significant effect on the duration of /u/ ($\hat{\beta} = -0.5$ ms, se = 0.06 ms). In general, closure duration is inversely correlated with vowel duration when 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 C2 voicing affects word duration. 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 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 duration

The duration of the interval between the release of C1 and the release of C2 is not affected by C2 voicing.

165 IV. DISCUSSION

166 A major drawback of the analysis in Maddieson and Gandour (1976) is that the conso-
 167 nant duration in fact was measured from the closure of the relevant consonant to the release
 168 of the following consonant, due to difficulties in detecting the release of the consonant of
 169 interest (e.g., in *ab s̄ath kaho*, the duration of /t / in *s̄ath* was calculated as the interval
 170 between the closure of /t / and the release of /k/). This measure includes the burst and
 171 (eventual) aspiration of the consonant. Slis and Cohen (1969a), however, states that the in-
 172 verse correlation between vowel duration and the following consonant raises when consonant
 173 *closure* duration is taken into account, and not entire *consonant* duration. If the correlation
 174 exists between vowel and closure duration, the inclusion of burst/aspiration duration clearly
 175 alters this relationship.

176 ACKNOWLEDGMENTS

177 Thanks to...

178 APPENDIX A: SOCIO-LINGUISTIC INFORMATION OF PARTICIPANTS

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

A subtitle goes on another line

Participant ID	Age	Sex	Native language	Other languages	City of birth	Spoken language
it01	29	Male	Italian	English, Spanish	Verbania	V
it02	26	Male	Italian	Friulian, English, Ladin-Venetan	Udine	T
it03	28	Female	Italian	English, German	Verbania	V
it04	54	Female	Italian	Calabrese	Verbania	V
it05	28	Female	Italian	English	Verbania	V
it07	29	Male	Italian	English	Tradate	C
it09	35	Female	Italian	English	Vignola (MO), Italy	V
it11	24	Male	Italian	english	Monza	M
it13	20	Female	Italian	English, French, Arabic, Farsi	Ancona	C
it14	32	Male	Italian	English, Spanish	Frosinone	F
pl02	32	Female	Polish	English, Norwegian, French, German, Dutch	Koło	P
pl03	26	Male	Polish	Russian, English, French, German	Nowa Sol	P
pl04	34	Female	Polish	Spanish, English, French	Warsaw	W
pl05	42	Male	Polish	English, French	Przasnysz	W
pl06	33	Male	Polish	English	Zgierz	Z
pl07	32	Female	Polish	English, Russian	Bielsk Podlaski	B

- 183 Ananthapadmanabha, T. V., Prathosh, A. P., and Ramakrishnan, A. G. (2014). “Detection
184 of the closure-burst transitions of stops and affricates in continuous speech using the plosion
185 index,” The Journal of the Acoustical Society of America **135**(1), 460–471.
- 186 Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). “Fitting linear mixed-effects
187 models using lme4,” Journal of Statistical Software **67**(1), 1–48.
- 188 Belasco, S. (1953). “The influence of force of articulation of consonants on vowel duration,”
189 The Journal of the Acoustical Society of America **25**(5), 1015–1016.
- 190 Bigi, B. (2015). “SPPAS - Multi-lingual approaches to the automatic annotation of speech,”
191 The Phonetician **111–112**, 54–69.
- 192 Chen, M. (1970). “Vowel length variation as a function of the voicing of the consonant
193 environment,” *Phonetica* **22**(3), 129–159.
- 194 Diggle, P., , Heagerty, P., Liang, K.-Y., and Zeger, S. (2002). *Analysis of longitudinal data*
195 (Oxford University Press).
- 196 Durvasula, K., and Luo, Q. (2012). “Voicing, aspiration, and vowel duration in Hindi,”
197 Proceedings of Meetings on Acoustics **18**, 1–10.
- 198 Esposito, A. (2002). “On vowel height and consonantal voicing effects: Data from Italian,”
199 *Phonetica* **59**(4), 197–231.
- 200 Farnetani, E., and Kori, S. (1986). “Effects of syllable and word structure on segmental
201 durations in spoken Italian,” *Speech communication* **5**(1), 17–34.
- 202 Fowler, C. A. (1992). “Vowel duration and closure duration in voiced and unvoiced stops:
203 There are no contrast effects here,” *Journal of Phonetics* **20**(1), 143–165.

- 204 Hajek, J., and Stevens, M. (2008). “Vowel duration, compression and lengthening in stressed
205 syllables in central and southern varieties of standard italian,” ISCA.
- 206 Halle, M., and Stevens, K. (1967). “Mechanism of glottal vibration for vowels and conso-
207 nants,” The Journal of the Acoustical Society of America **41**(6), 1613–1613.
- 208 Heffner, R.-M. (1937). “Notes on the length of vowels,” American Speech **12**, 128–134.
- 209 House, A. S., and Fairbanks, G. (1953). “The influence of consonant environment upon the
210 secondary acoustical characteristics of vowels,” The Journal of the Acoustical Society of
211 America **25**(1), 105–113.
- 212 Hussein, L. (1994). “Voicing-dependent vowel duration in Standard Arabic and its acquisi-
213 tion by adult american students,” Ph.D. thesis, The Ohio State University.
- 214 Jacewicz, E., Fox, R. A., and Lyle, S. (2009). “Variation in stop consonant voicing in two
215 regional varieties of American English,” Journal of the International Phonetic Association
216 **39**(3), 313–334, doi: [10.1017/S0025100309990156](https://doi.org/10.1017/S0025100309990156).
- 217 Jarosz, A. F., and Wiley, J. (2014). “What are the odds? a practical guide to computing
218 and reporting Bayes factors,” The Journal of Problem Solving **7**(1), 2–9, doi: [10.7771/
219 1932-6246.1167](https://doi.org/10.7771/1932-6246.1167).
- 220 Javkin, H. R. (1976). “The perceptual basis of vowel duration differences associated with
221 the voiced/voiceless distinction,” Report of the Phonology Laboratory, UC Berkeley **1**,
222 78–92.
- 223 Klatt, D. H. (1973). “Interaction between two factors that influence vowel duration,” The
224 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](https://doi.org/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). “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.
- 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).

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

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. (1969a). “On the complex regulating the voiced-voiceless distinction I,” *Language and speech* **12**(2), 80–102.

Slis, I. H., and Cohen, A. (1969b). “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.

Team, R. C. (2018). “R: A language and environment for statistical computing” R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org>.

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).