This is a title and this is too

Stefano Coretta[1]

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; 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 (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 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 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

89 of the algorithm described in Ananthapadmanabha *et al.* (2014). The durations in millisec-

90 onds of the following intervals were extracted from the annotated acoustic landmarks with

91 Praat scripting: sentence duration, word duration, vowel duration (V1 onset to V1 offset),

92 consonant closure duration (V1 offset to C2 burst), and Release-to-Release duration (RR

93 duration, C1 release to C2 release). Syllable rate (syllables per second) was used as a proxy

94 to speech rate (Plug and Smith, 2018) for duration normalisation, and was calculated as

95 the number of syllables divided by the duration of the sentence (8 syllables in Italian, 6 in

96 Polish).

97 **F. Statistical analysis**

98 The durational measurements were analysed with linear mixed-effects models using `lme4`

99 v1.1-17 in R v3.5.0 (Bates *et al.*, 2015; Team, 2018). All factors were coded as treatment

100 contrasts. *P*-values for the individual terms were obtained with `lmerTest` v3.0-1, which uses

101 the Satterthwaite's approximation to degrees of freedom (Kuznetsova *et al.*, 2017). *P*-values

102 below the alpha level 0.05 were considered significant. The estimates of the relevant effects

103 are then calculated by refitting the models including only the significant terms (step-down

104 approach, Diggle *et al.*, 2002; Zuur *et al.*, 2009, pp. 121–122). Bayes factors were used

105 to specifically test the null hypotheses that word and RR duration are not affected by C2

106 voicing (i.e., the effect of C2 voicing on duration is 0). For each set of null/alternative

107 hypotheses, a full model (with the predictor of interest) and a null model (excluding it)

108 were fitted separately using Maximum Likelihood estimation (Bates *et al.*, 2015, p. 34).

109 The BIC approximation was then used to obtain Bayes factors (Wagenmakers, 2007). The

110 approximation is calculated according to the equation in 1 (Wagenmakers, 2007, p. 796).

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

111 where $\Delta BIC_{10} = BIC_1 - BIC_0$, BIC_1 is the BIC of the full model, and BIC_0 is the BIC

112 of the null model. Values of $BF_{01} > 1$ indicate a preference of H_0 over H_1. The interpretation

113 of the Bayes factors follows the recommendations in **?**, p. 139.

114 **III. RESULTS**

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

116 **A. Vowel duration**

117 A linear mixed-effects model was fitted with the following terms: vowel duration as the

118 outcome variable; fixed effects for C2 voicing, C2 place of articulation, vowel identity, lan-

119 guage, and speech rate (as syllables per second); by-speaker and by-word random intercept

120 with by-speaker random slopes for C2 voicing. All logical interactions were included. The

121 following terms and interactions are significant: C2 voicing, vowel identity, language, speech

122 rate, the interaction between C2 voicing and vowel, and C2 place and vowel. The vowels /a,

123 o/ are 14 ms longer (se = 3 ms) when followed by a voiced stop. The effect is smaller for

124 /u/ (around 6 ms; $\hat{\beta}$ = -8 ms, se = 4 ms). Polish has on average shorter vowels than Italian

125 ($\hat{\beta}$ = -25.5 ms, se = 7.5 ms), although the effect of voicing is estimated to be the same in

126 both languages (the interaction of language and C2 voicing is not significant). Speech rate

127 has a negative effect on vowel duration, such that faster rates correlate with shorter vowel

128 durations ($\hat{\beta}$ = -15 ms, se = 1 ms).

B. Consonant closure duration

130 The same maximally specified model as with vowel duration has been fitted to conso-

131 nant closure durations as the outcome variable. The following terms and interactions were

132 significant: C2 voicing, C2 place of articulation, vowel identity, language, and interactions

133 between language and C2 place, language and vowel identity, C2 voicing and place, C2 voic-

134 ing and vowel, and a three-way interaction between C2 voicing, place and vowel identity.

135 Stop closure is 15 ms shorter (se =) if the stop is voiced. Vowel identity has an effect on

136 closure duration in voiced stops, but not in voiceless stops, and more so in voiced velar than

137 in voiced coronal stops: closure after /a/ is the shortest, while after /u/ is the longest, with

138 closure after /o/ in the middle.

C. Vowel and closure duration

140 The model with the following specification was fit to assess the correlation between vowel

141 and closure durations: vowel duration as the outcome variable; fixed effects for closure

142 duration, vowel identity, C2 place of articulation, and speech rate; all interactions between

143 closure duration, vowel, and C2 place; by-speaker and by-word random intercepts, and by-

144 speaker random slope for closure duration. The results of terms which were included in the

145 analysis of vowel duration in XXX will not be discussed here (see Supplementary material).

9

146 Closure duration has a non-significant effect on vowel duration when the vowel is /a/ ($\hat{\beta} =$

147 -0.2 ms, se = 0.06 ms) and /o/ ($\hat{\beta}$ = -0.25 ms, se = 0.06 ms), but it has a significant effect

148 on the duration of /u/ ($\hat{\beta}$ = -0.5 ms, se = 0.06 ms). In general, closure duration is inversely

149 correlated with vowel duration when controlling for speech rate. However the correlation

150 is quite weak: a 1 ms increase in closure duration corresponds to a 0.2–0.5 ms decrease in

151 vowel duration.

D. Word duration

153   The following full and null models were fitted to test whether C2 voicing affects word

154 duration. In the full model, I entered as fixed effects: C2 voicing, C2 place, vowel, speech

155 rate, and language. The model also included by-speaker and by-word random intercepts,

156 plus a by-speaker random slope for C2 voicing. The null model was the same as the full

157 model, with the exclusion of the fixed effect of C2 voicing. The Bayes factor of the null

158 model against the full model is 21.5. Thus, the null model (in which the effect of C2 voicing

159 is 0) is 21.5 times more likely than the full model given the data. This indicates that there

160 is strong evidence for word duration not being affected by C2 voicing.

E. Release to Release interval duration

162   The duration of the interval between the release of C1 and the release of C2 is not affected

163 by C2 voicing.

IV. DISCUSSION

A major drawback of the analysis in Maddieson and Gandour (1976) is that the conso-
nant duration in fact was 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 /t / in *sāth* was calculated as the interval
between the closure of /t / and the release of /k/). This measure includes the burst and
(eventual) aspiration of the consonant. Slis and Cohen (1969a), however, states that the in-
verse correlation between vowel duration and the following consonant raises when consonant
closure duration is taken into account, and not entire *consonant* duration. If the correlation
exists between vowel and closure duration, the inclusion of burst/aspiration duration clearly
alters this relationship.

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).

11

Participant ID	Age	Sex	Native language	Other languages	City of birth	S
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	Engligh	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

182 Ananthapadmanabha, T. V., Prathosh, A. P., and Ramakrishnan, A. G. (**2014**). "Detection

183 of the closure-burst transitions of stops and affricates in continuous speech using the plosion

184 index," The Journal of the Acoustical Society of America **135**(1), 460–471.

185 Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects

186 models using lme4," Journal of Statistical Software **67**(1), 1–48.

187 Belasco, S. (**1953**). "The influence of force of articulation of consonants on vowel duration,"

188 The Journal of the Acoustical Society of America **25**(5), 1015–1016.

189 Bigi, B. (**2015**). "SPPAS - Multi-lingual approaches to the automatic annotation of speech,"

190 The Phonetician **111–112**, 54–69.

191 Chen, M. (**1970**). "Vowel length variation as a function of the voicing of the consonant

192 environment," Phonetica **22**(3), 129–159.

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

194 (Oxford University Press).

195 Durvasula, K., and Luo, Q. (**2012**). "Voicing, aspiration, and vowel duration in Hindi,"

196 Proceedings of Meetings on Acoustics **18**, 1–10.

197 Esposito, A. (**2002**). "On vowel height and consonantal voicing effects: Data from Italian,"

198 Phonetica **59**(4), 197–231.

199 Farnetani, E., and Kori, S. (**1986**). "Effects of syllable and word structure on segmental

200 durations in spoken Italian," Speech communication **5**(1), 17–34.

201 Fowler, C. A. (**1992**). "Vowel duration and closure duration in voiced and unvoiced stops:

202 There are no contrast effects here," Journal of Phonetics **20**(1), 143–165.

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.

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.

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.

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**). "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).

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.

246 Slis, I. H., and Cohen, A. (**1969**a). "On the complex regulating the voiced-voiceless distinc-

247   tion I," Language and speech **12**(2), 80–102.

248 Slis, I. H., and Cohen, A. (**1969**b). "On the complex regulating the voiced-voiceless distinc-

249   tion II," Language and speech **12**(3), 137–155.

250 Sóskuthy, M. (**2013**). "Phonetic biases and systemic effects in the actuation of sound

251   change," Ph.D. thesis, University of Edinburgh.

252 Team, R. C. (**2018**). "R: A language and environment for statistical computing" R Foun-

253   dation for Statistical Computing, Vienna, Austria, `https://www.R-project.org`.

254 Vazquez-Alvarez, Y., and Hewlett, N. (**2007**). "The 'trough effect': an ultrasound study,"

255   Phonetica **64**(2-3), 105–121.

256 Wagenmakers, E.-J. (**2007**). "A practical solution to the pervasive problems of p values,"

257   Psychonomic bulletin & review **14**(5), 779–804.

258 Warren, W., and Jacks, A. (**2005**). "Lip and jaw closing gesture durations in syllable final

259   voiced and voiceless stops," The Journal of the Acoustical Society of America **117**(4),

260   2618–2618.

261 Zuur, A. F., Saveliev, A. A., Ieno, E. N., Smith, G. M., and Walker, N. (**2009**). *Mixed*

262   *Effects Models and Extensions in Ecology with R* (Springer).