The link between tongue root advancement and the voicing effect: an ultrasound study of Italian and Polish

Stefano Coretta[a,*]

[a] *University of Manchester, Linguistics and English Language*

1. Introduction

This paper reports a previously undocumented link between tongue root advancement in the production of voiced stops and longer duration of the preceding vowels. Using ultrasonic data from speakers of Italian and Polish, I demonstrate that the presence of tongue root advancement correlates with that of the so-called voicing effect, whereby vowels tend to be longer before voiced than before voiceless stops. I further suggest that such correlation points to a new articulatory account of the differential duration of vowels as a covariate of consonantal voicing.

A considerable number of studies show that vowels tend to be longer when followed by voiced obstruents than when they are followed by voiceless obstruents (House and Fairbanks 1953, Chen 1970, Klatt 1973, Lisker 1974, Farnetani and Kori 1986, Fowler 1992, Hussein 1994, Esposito 2002, Lampp and Reklis 2004, Durvasula and Luo 2012). This voicing effect has been reported 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 list). A common stance in the literature is that the magnitude of the voicing effect differs depending on the language (although see Laeufer 1992), and that this phenomenon is not a universal tendency, since the duration of vowels is not affected by the voicing of the following obstruents in some languages, like Polish and Czech (Keating 1984). Although several attempts have been made to explain the voicing effect, an account that survives all empirical data is still lacking (Durvasula and Luo 2012).

To provide a rationale for the language-specificity of the voicing effect, Kluender et al. (1988) propose that the source of the effect rests at the level of the perceptual system. They argue that speakers actively manipulate vowel durations to enhance the closure-duration difference which can cue voicing distinction in obstruents. However, Fowler (1992) shows by means of perceptual experiments that speakers judge vowels to be longer if the consonant closure duration is increased, and that, vice versa, consonant closure is perceived to be longer when vowel duration is increased. Fowler thus refutes on empirical grounds that speakers exploit durational contrasts to enhance voicing distinctions. Crucially, Davis and Van Summers (1989) demonstrate that closure duration is not employed as a cue to voicing, fact that further confutes any account based on closure-duration enhancement.

A desirable account of the voicing effect should then satisfy both of the following requirements: (1) it should allow for a language-specific implementation of the voicing effect, and (2) it should place the source of the effect in articulatory properties of voiced or voiceless obstruents that could favour longer or shorter vowel durations. One of the first articulatory accounts of the voicing effect to be proposed points to differences in adjustments of the glottis before voiced and voiceless stops (Halle and Stevens 1967, reiterated in Chomsky and Halle 1968). Halle and Stevens (1967) state that the glottal configuration during voicing in vowels is difference from the configuration needed during the closure of voiced consonants. The argument follows that vowels are longer before voiced consonants so that more time is available to the glottis to achieve a position suitable for main-

*Corresponding author
Email address: `stefano.coretta@manchester.ac.uk` (Stefano Coretta)

taining within-closure voicing, such that said configuration is secured before the onset of the consonant closure. Later studies, however, established that the glottal configuration of voicing in obstruents does not differ from the one in vowels (Lisker 1974, Kagaya and Hirose 1975). Hence, the claim that a glottal adjustment is required before voiced consonants is not supported empirically and must be abandoned.

A known articulatory difference between voiced and voiceless consonants regards a supra-glottal ... rather than laryngeal ..., namely the position of the tongue root along the front-back plane of the oral tract. It has been observed that the tongue root is in a more advanced position in voiced than in voiceless stops (Kent and Moll 1969, Perkell 1969, Westbury 1983). This gesture has been interpreted as a mechanism to ensure voicing can be maintained during closure. The realisation of vocal fold vibration (voicing) requires air pressure of the supra-glottal cavity to be lower than a certain threshold in comparison with the air pressure of the sub-glottal cavity. During the production of voiced obstruents, the supra-glottal pressure quickly increases. Such pressure increase can hinder the ability to maintain voicing during closure, to the point that voicing can cease if the pressure cut-off threshold is reached. An articulatory solution to counterbalance the increased pressure is to expand the supra-glottal cavity by advancing the root of the tongue.

Drawing on ultrasound tongue imaging data, Ahn and Davidson (2016) demonstrate that the root of the tongue is advanced during the articulation of voiced consonants in American English, even when voicing is not implemented during the closure in underlyingly voiced stops. Based on data from Korean, they further show that tongue root advancement not only favours voicing in stops, but it also facilitates short lag VOT in tense stops in Korean. The established link between voicing and tongue root advancement on one hand, and between voicing and vowel duration on the other, raise the question of whether a correlation exists between tongue root advancement and vowel duration. If tongue root advancement plays a role in the voicing effect, then it is expected that voiced consonants should be articulated with tongue root advancement in voicing-effect languages, but not in languages without the voicing effect.

For this study ultrasonic tongue data were recorded from speakers of Italian and Polish, two languages that have been reported to have and lack the voicing effect respectively. In a study investigating the general properties of segmental durations of Italian, Farnetani and Kori (1986) show that the first vowel in /lada/ is on average 35 milliseconds longer than the vowel in /lata/ ($\bar{x}$ = 223 ms, SD = 18 in /lata/; $\bar{x}$ = 258 ms, SD = 13 in /lada/, Farnetani and Kori 1986:26). Esposito (2002) extends Farnetanis's research to all the vowels and stops of Italian and confirms that vowels are longer when followed by a voiced stop, with an estimated mean duration similar to that reported in Farnetani and Kori (1986). As for Polish, although a difference of 2 milliseconds in vowel duration (167.5 ms in /rata/, 169.5 ms in /rada/) can be observed, such difference is deemed as non-significant according to statistical testing (Keating 1984). Vowels in Polish are thus not affected by the voicing of the following consonant.

Based on the hypothesis that tongue root advancement correlates with the voicing effect (and hence with vowel duration), the expectations with regard to Italian and Polish are the following: (1) since vowels are longer before voiced stops in Italian, voiced stops in this language should be articulated with tongue root advancement, and (2) given that vowels before voiceless and voiced stops in Polish are about the same duration, the position of the tongue root in voiced stops should not differ from that of voiceless stops.

2. Methodology

2.1. Participants

Eight native speakers of Italian (2 females, 2 males) and Polish (2 females, 2 males) were recorded in Manchester and in Italy (Table 1). The Italian speakers were from Northern Italy (three from the Northwest and one from Northeast). The Polish group was more heterogeneous, with two speakers from the

Table 1: Sociolinguistic information on participants. The right-most column indicates whether the participant spent more than 6 consecutive months abroad.

id	**sex**	**age**	**city**	> 6 mo
IT01	m	28	Verbania	yes
IT02	m	26	Udine	yes
IT03	f	27	Verbania	no
IT04	f	54	Verbania	no
PL02	f	32	Poznań	yes
PL03	m	26	Poznań	yes
PL04	f	34	Warsaw	no
PL05	m	34	Przasnysz	no

North-west (Poznań), and two from the North-east (Warsaw and Przasnysz). Ethical clearance was obtained for this study from the University of Manchester (REF 2016-0099-76). The participants received a small monetary compensation.

2.2. Equipment set-up

An Articulate Instruments Ltd™ set-up was used for this study (Figure 1). The ultrasonic data was collected through a TELEMED Echo Blaster 128 unit with a TELEMED C3.5/20/128Z-3 ultrasonic transducer (20mm radius, 2-4 MHz). A synchronisation unit (P-Stretch) was plugged into the Echo Blaster unit and used for automatic audio/ultrasound synchronisation. A FocusRight pre-amplifier and a Movo LV4-O2 Lavalier microphone were used for audio recording. The acquisition of the ultrasonic and audio signals was achieved with the software Articulate Assistant Advanced (AAA, v2.17.2) running on a Hawlett-Packard ProBook 6750b laptop with Microsoft Windows 7. Stabilisation of the ultrasonic transducer was ensured by using a stabilisation headset produced by Articulate Instruments Ltd™ (Articulate Instruments Ltd™ 2008).

2.3. Materials

Disyllabic words of the form $C_1V_1C_2V_2$ were used as targets, 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.), yielding a total of 12 target words. A labial stop was chosen as the first consonant to reduce possible coarticulation with the following vowel (although see Vazquez-Alvarez and Hewlett 2007). Only coronal and velar stops were used as target consonants since labial consonants cannot be imaged with ultrasonography. The target words were embedded in a frame sentence. Prosodically similar sentences were used to ensure comparability between languages. The frame sentence was *Dico X lentamente* 'I say X slowly' for Italian, and *Mówię X teraz* 'I say X now' for Polish.

2.4. Procedure

The sentences with the target words were randomised for each participant, although the order was kept the same between repetitions within participant due to software constraints. Each participant repeated the list of randomised stimuli six times. A grand total of 576 tokens (288 per language) was recorded. The participant's occlusal plane was obtained using a bite plate (Scobbie et al. 2011), and the hard palate was imaged by asking the participant to swallow water (Epstein and Stone 2005). The frame rate of the acquisition of the ultrasonic data ranged between 43 and 68 frames per second. Other settings values varied depending on the frame rate. The ranges of the settings in this study were: scanlines = 88-114, pixel per scanline = 980-988, field of view = 71-93, pixel offset = 109-263, depth (mm) = 75-180. The audio signal was recorded at 22050 MHz (16-bit).

2.5. Data processing and analysis

Synchronisation of the ultrasonic and audio signal was achieved in post-processing, using a built-in procedure of AAA. The audio data were subject to force alignment using the SPPAS force aligner (Bigi 2015). The outcome of the automatic annotation was manually corrected after the force alignment, according to the criteria in Table 2. The onset of the target consonant burst (C2 burst) was detected automatically in Praat (Boersma and Weenink 2016) by means of the algorithm described in Ananthapadmanabha et al.

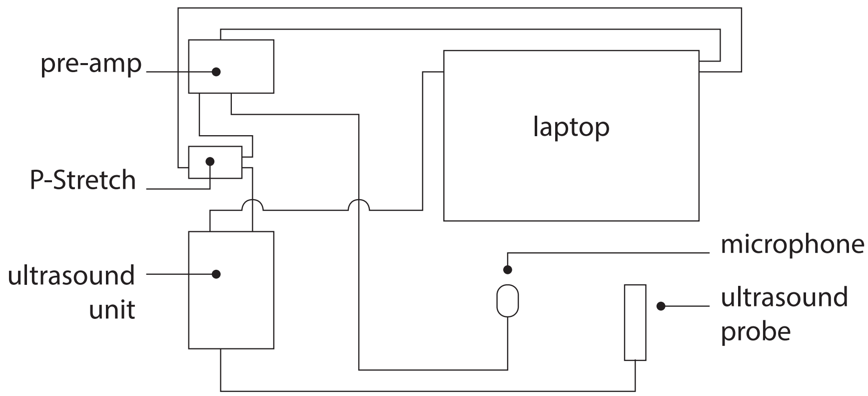

Figure 1: Schematic representation of the equipment setup (Articulate Instruments Ltd™ 2011). The system is described in details in Section 2.2.

(2014). The durations of the following intervals were extracted from the annotated acoustic landmarks using an automated procedure in Praat: vowel duration (V1 onset to V1 offset), consonant duration (V1 offset to V2 onset), and closure duration (V1 offset to C2 burst).

Spline curves were automatically fitted to the visible contours using the AAA batch tracking function. Manual correction was applied in those cases that showed clear tracking errors. The time of maximum tongue displacement within consonant closure was then calculated in AAA following the method in Strycharczuk and Scobbie (2015). A fan-like frame consisting of 42 equidistant radial lines was used as the coordinate system. The origin of the 42 fan-lines coincides with the centre of the ultrasonic probe, such that each fan-line is parallel to the direction of the ultrasonic signal. Tongue displacement was thus calculated as the displacement of the fitted splines along the fan-line vectors. The time of maximum tongue displacement was the time of greater displacement along the vector that showed the greatest standard deviation. The vector search area was restricted to the portion of the splines corresponding to the tongue tip for coronal consonants, and to the portion corresponding to the tongue dorsum for velar consonants.

The cartesian coordinates of the tongue contours were exported from two time points: the onset of C2 closure, and the time maximum tongue displacement (which is always within C2 closure). The contours were normalised within speaker by applying offsetting and rotation relative to the participant's occlusal plane (Scobbie et al. 2011). The durational data were analysed with linear mixed effects models using `lme4` in R (R Core Team 2017, Bates et al. 2015). P-values were obtained with likelihood ratio tests comparing the full model with a nested model without the tested predictor, and with `lmerTest` (Kuznetsova et al. 2016), which employs the Satterthwaite approximation to degrees of freedom. Generalised additive mixed models (GAMMs, Wood 2006, Zuur 2012) were used for the statistical analysis of tongue contour data. Significance testing in GAMMs was achieved through model comparison and visual inspection of the difference smooth, as suggested in (Sóskuthy 2017).

3. Results

The following sections report the results of the durational data (Section 3.1), and of the ultrasonic data (Section 3.2). Given the poor quality of the ultrasonic data for /u/, this vowel was not included in the statistical analysis of tongue contours, but the durational data of /u/ were kept in the linear regression models.

Table 2: 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 burst onset	(C2 burst)	automatic detection (Ananthapadmanabha et al. 2014)

The results in Section 3.2 thus refer only to /a/ and /o/.

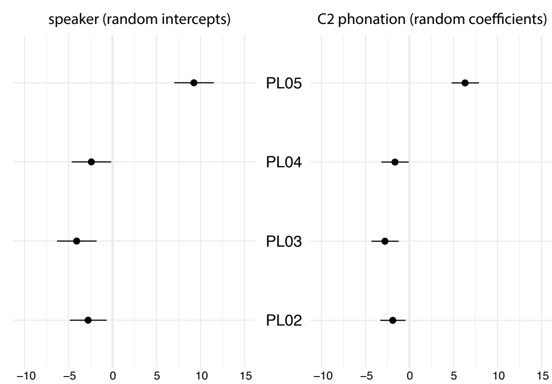

Figure 2: By-speaker random intercepts (left) and coefficients (right) for the effect of C2 voicing on vowel duration in Polish. The higher coefficient estimate (+6.3 ms) for PL05 indicates a relatively stronger effect of voicing for this participant.

3.1. Vowel duration and voicing

A linear mixed effects regression model was fitted to the Italian vowel duration data with DURATION as the outcome variable; VOWEL IDENTITY (/a, o, u/), VOICING and PLACE OF ARTICULATION of the following consonant, SENTENCE DURATION as fixed effects; random intercepts by speaker and word, and by-speaker random slopes for voicing. An interaction between voicing and vowel quality was also included in the final model, since it significantly improved the model. According to the full model as specified above, Italian vowels are 22 milliseconds longer (se = 6) if followed by a voiced stop ($\chi 2 = 16.61$; $p < 0.001$; df = 3). The following terms and interactions were also significant: place of articulation C2, vowel identity, sentence duration, and the interaction between C2 voicing and vowel identity.

For Polish, the same model structure was used, to the exclusion of the voicing/vowel-identity interaction (which was not significant). Contrary to previous findings, the model reveals a significant 8 milliseconds effect (se = 3) of consonantal voicing on the preceding vowel ($\chi 2 = 5.4$; $p < 0.05$, df = 1). Vowel identity and sentence duration were also significant. The place of C2 significantly improved the model ($\chi 2 = 6.1$; $p < 0.05$; df = 1), so it was included in the full model even though it is deemed as non significant according to the single predictor p-value (t = -2.3; ns; df = 7). The inspection of the model residuals confirmed the assumptions of normality and homoscedasticity.

The exploration of the random coefficients for each speaker indicated that PL05 has a particularly higher coefficient for voicing, meaning that the effect of voicing is stronger in his data (Figure 2). Note that the exclusion of this speaker from the model doesn't remove the effect (β = 6 ms; sd = 2; $\chi 2 = 8.28$; $p < 0.01$, df = 1). The estimated effect of voicing on vowel duration for PL05 was about 14 milliseconds (vowels followed by voiced stops are on average 14 ms longer in PL05). These observations will become relevant in the next section, in which the results of the tongue contour data will be discussed.

3.2. Tongue contours

GAMMs were fitted to the data of each speaker individually: the Y-COORDINATES of the contours were included in the model as the outcome variable; the X-COORDINATES as the only parametric term. The following smooths were specified: a reference smooth term for the x-coordinates, three difference smooths for the x-coordinates by VOICING, VOWEL IDENTITY, and PLACE of articulation of the following consonant respectively, and BY-WORD random smooths. A first-order autoregressive model was included to correct for the high autocorrelation of the residuals.

The analysis of the Italian ultrasonic data shows that voiced stops are produced with advancement of the root of the tongue, similarly to what found in previous research on English (Ahn and Davidson 2016). Figure 3 (top half) shows for each Italian speaker the estimated tongue contours in voiceless (dashed lines) and voiced stops (continuous lines) at the time of maximum tongue displacement. Below each tongue contour panel, the by-voicing difference smooth is also shown (confidence interval in grey). Tongue contours are significantly different in those points in which the confidence interval of the difference smooth does not include 0 on the ordinate axis. The significantly different portions of the contours are also indicated in the figures by a shaded rectangular shape.

In two participants out of four (IT01, IT02), the root was significantly more front in voiced stops in both vocalic contexts (/a, o/). On the other hand, one participant (IT03) had significant tongue root advancement only following /a/, while the fourth participant (IT04) didn't show advancement at all. For Polish (bottom half of Figure 3), three out of four speakers (PL02, PL03, PL04) did not have tongue root advancement, while the fourth speaker (PL05) had significant advancement in voiced stops in both vocalic contexts.

An additional contour analysis was carried out at C2 closure onset for the Italian and Polish speakers showing advancement. The tongue root at closure onset was found to be advanced in the voiced consonants of all these speakers. Comparisons of tongue contours at C2 onset and at the time of maximum tongue displacement in voiced consonants further indicated that the degree of root advancement was larger at maximum displacement for the Italian speakers (IT01, IT02, IT03), but not for the Polish speaker (PL05). Figure 4 exemplifies the results with IT01 and PL05.

The ultrasonic data also showed that the tongue body was raised in speakers with tongue root advancement. The presence of this gesture in addition to root advancement makes sense from an anatomical point of view. Raising the tongue body could be a way to counterbalance the compression of the tongue mass caused by the advancement of the root (Perkell 1969, Jackson 1988, Kingston et al. 1997, Fulop et al. 1998). It is not thus surprising to observe a raised tongue body in voiced stops accompanying root advancement.

4. Discussion

Based on the previously established link between tongue root advancement and voicing, it was proposed at the beginning of this paper that the presence of the voicing effect in a language should be correlated with the presence of tongue root advancement in voiced stops. To test the correlation between tongue root advancement and vowel durations, ultrasonic data were collected from two languages that differ in the presence/absence of the voicing effect, Italian and Polish respectively. It was predicted that voiced consonants should be produced with an advanced tongue root in Italian, but not in Polish. The results indicate that tongue root advancement in voiced stops indeed correlates with the presence of the VE, although the details of such correlation differ from what was expected. This section will discuss … and the implications of the findings reported in this paper with regard to the source of the voicing effect.

Keating (1984) investigated the effect of consonant voicing on the duration of preceding vowels and reported that vowels followed by voiceless and voiced

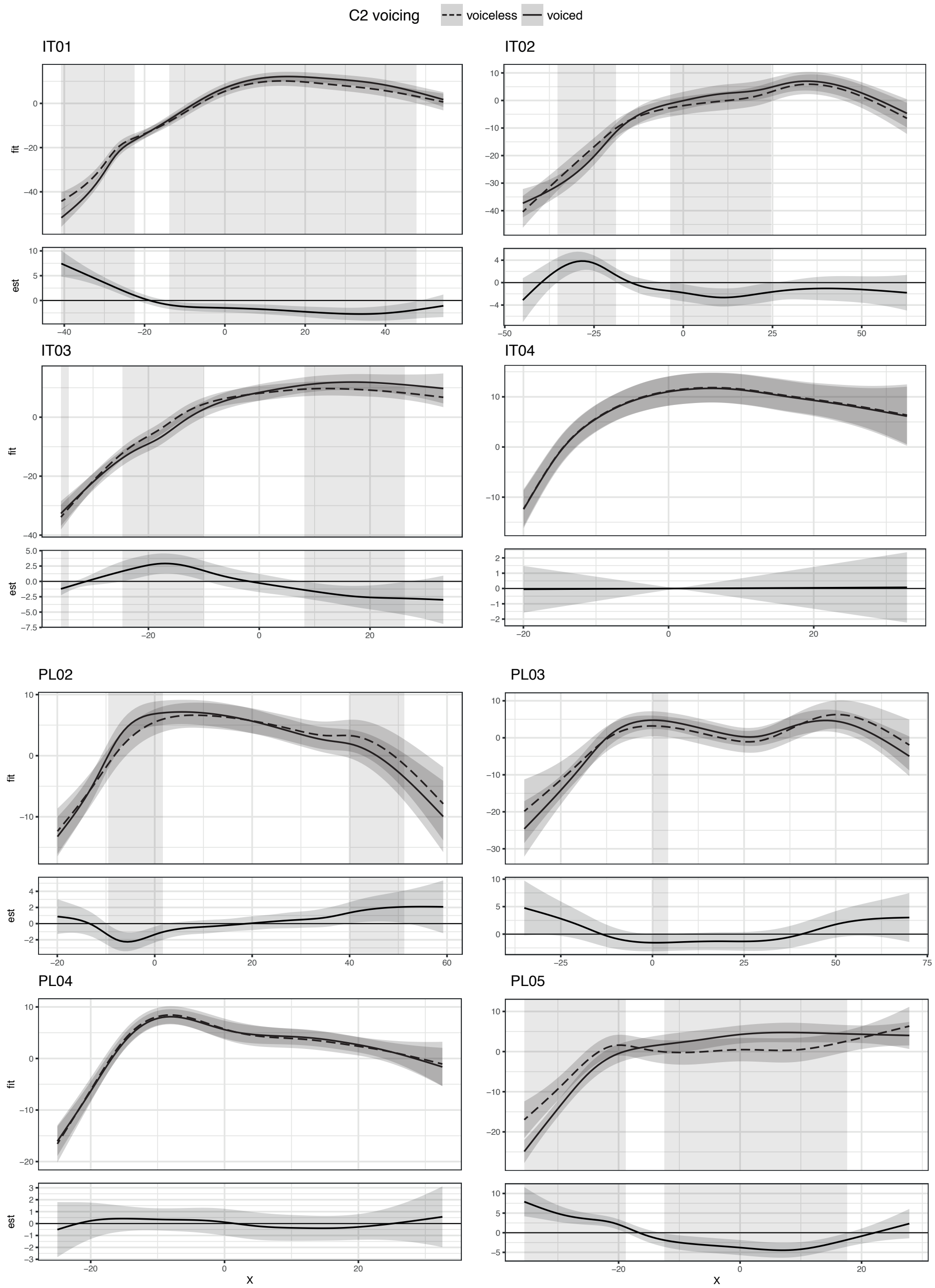

Figure 3: Comparison of tongue contours at the time of maximum tongue displacement (within C2 closure) in Italian (top half) and Polish (bottom half). The plotted contours are the estimated curves in the context of the vowel /a/ followed by coronal consonants. By-voicing difference smooths are shown below the estimated curves. See Section 3.2 for more details.

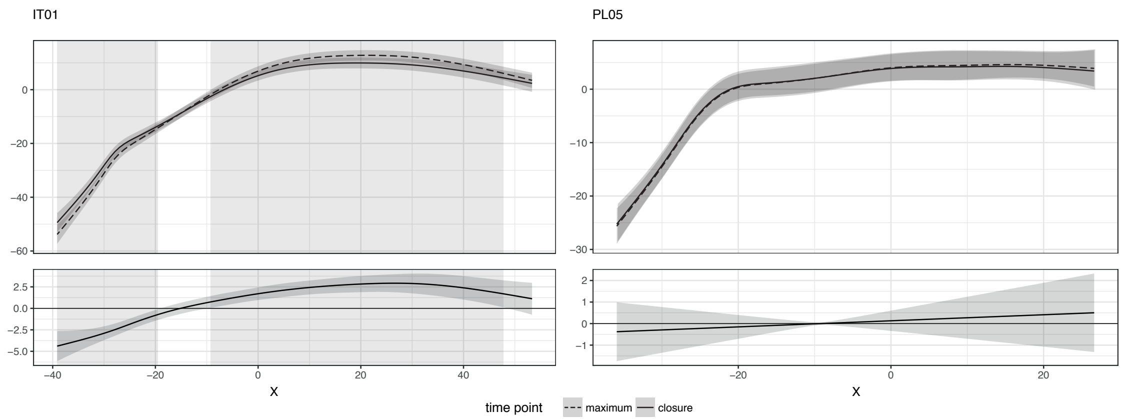

Figure 4: Comparison of tongue contours of voiced consonants at C2 closure onset and maximum tongue displacement (within C2 closure) in IT01 (Italian) and PL05 (Polish). See Section 3.2 for more details.

stops do not differ in duration. Nevertheless, the analysis of the durational data discussed above indicates that an effect of voicing exists in the Polish speakers of this study, such that vowels followed by voiced stops are on average 8 milliseconds longer. The hypothesis that voiced consonants should not have tongue root advancement in Polish rests on the notion that Polish has been reported as a non-voicing-effect language, which was not corroborated by our data. Crucially, the Polish speaker with the strongest effect of voicing (PL05, cf. Section 3.1) is also the only Polish speaker who produced voiced consonants with an advanced tongue root, both at C2 closure onset and at the time of maximum tongue displacement.

This pattern is similar to that of the Italian speakers, who have both a relatively strong voicing effect and tongue root advancement. This is true for all the speakers of Italian in this sample, except IT04. The vowels in this speaker are 22 milliseconds longer when followed by voiced stops, but her tongue root position does not differ in voiced and voiceless stops. The interpretation of IT04's different behaviour as an age effect suggests itself given the older age of this speaker (54 year old vs. the average age of the other Italian speaker, which was 26).

Overall, the speakers who produced voiced consonants with an advanced tongue root had a relatively strong voicing effect, with estimates of the durational differences ranging between 15 and 30 milliseconds. Bear in mind that the inverse is not true: IT04 had a strong voicing effect, without accompanying tongue root advancement. Nonetheless, the generalisation that tongue root advancement is correlated with a relatively strong voicing effect holds independently of the speaker's language: speakers of both Italian and Polish pattern according to this principle. All things considered, the following implicational statement can be inferred: if a speaker realises voiced consonants with concomitant advancement of the tongue root, other things being equal, then that same same speaker will also show a considerable durational difference in vowels, with longer vowels preceding voiced consonants.

The fact that the degree of tongue advancement at C2 closure onset and at maximum displacement does not differ in PL05's voiced consonants could be linked to the relative weaker effect of voicing in this speaker compared to the Italian speakers (14 vs 22 milliseconds). The weaker voicing effect in PL05 suggests that the relationship between tongue root advancement and vowel duration might be gradient rather then categorical (presence vs. absence). If this is the case, the magnitude of the voicing effect should correlate with the degree of tongue root advancement. More specifically, vowel duration is predicted to be directly correlated with the degree of advancement. Future work will set out to investigate the putative gradient effect of tongue root advancement on vowel duration.

5. Conclusion

Suzy Ahn and Lisa Davidson. Tongue root positioning in English voiced obstruents: Effects of manner and vowel context. *The Journal of the Acoustical Society of America*, 140(4):3221–3221, 2016.

T. V. Ananthapadmanabha, A. P. Prathosh, and A. G. Ramakrishnan. 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, 2014.

Articulate Instruments Ltd™. Ultrasound stabilisation headset users manual: Revision 1.4. Edinburgh, UK: Articulate Instruments Ltd, 2008.

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

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1): 1–48, 2015.

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

Paul Boersma and David Weenink. Praat: doing phonetics by computer [Computer program]. Version 6.0.23, 2016. URL `http://www.praat.org/`.

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

Noam Chomsky and Morris Halle. *The sound pattern of English*. New York, Evanston, and London: Harper & Row, 1968.

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

Karthik Durvasula and Qian Luo. Voicing, aspiration, and vowel duration in Hindi. *Proceedings of Meetings on Acoustics*, 18:1–10, 2012.

Melissa A. Epstein and Maureen Stone. The tongue stops here: Ultrasound imaging of the palate. *The Journal of the Acoustical Society of America*, 118 (4):2128–2131, 2005.

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

Edda Farnetani and Shiro Kori. Effects of syllable and word structure on segmental durations in spoken Italian. *Speech communication*, 5(1):17–34, 1986.

Carol A. Fowler. Vowel duration and closure duration in voiced and unvoiced stops: There are no contrast effects here. *Journal of Phonetics*, 20(1):143–165, 1992.

Sean A. Fulop, Ethelbert Kari, and Peter Ladefoged. An acoustic study of the tongue root contrast in degema vowels. *Phonetica*, 55(1-2):80–98, 1998.

Morris Halle and Kenneth Stevens. Mechanism of glottal vibration for vowels and consonants. *The Journal of the Acoustical Society of America*, 41 (6):1613–1613, 1967.

Arthur S. House and Grant Fairbanks. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, 25(1):105–113, 1953.

Lutfi Hussein. *Voicing-dependent vowel duration in Standard Arabic and its acquisition by adult American students*. PhD thesis, The Ohio State University, 1994.

Michel T. T. Jackson. Phonetic theory and cross-linguistic variation in vowel articulation. In *Working Papers in Phonetics*, volume 71. University of California: Los Angeles, 1988.

Ryohei Kagaya and Hajime Hirose. Fiberoptic electromyographic and acoustic analyses of hindi stop consonants. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics*, 9:27–46, 1975.

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

Raymond D. Kent and Kenneth L. Moll. Vocal-tract characteristics of the stop cognates. *Journal of the Acoustical Society of America*, 46(6B):1549–1555, 1969.

John Kingston, Neil A. Macmillan, Laura Walsh Dickey, Rachel Thorburn, and Christine Bartels. Integrality in the perception of tongue root position and voice quality in vowels. *The Journal of the Acoustical Society of America*, 101(3):1696–1709, 1997.

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

Keith R. Kluender, Randy L. Diehl, and Beverly A. Wright. Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics*, 16:153–169, 1988.

Alexandra Kuznetsova, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen. `lmerTest`: Tests in linear mixed effects models. `https://CRAN.R-project.org/package=lmerTest`, 2016. R package version 2.0-33.

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

Claire Lampp and Heidi Reklis. Effects of coda voicing and aspiration on Hindi vowels. *The Journal of the Acoustical Society of America*, 115(5):2540–2540, 2004.

Leigh Lisker. On “explaining” vowel duration variation. In *Proceedings of the Linguistic Society of America*, pages 225–232, 1974.

Ian Maddieson and Jack Gandour. Vowel length before aspirated consonants. In *UCLA Working papers in Phonetics*, volume 31, pages 46–52, 1976.

Joseph S. Perkell. *Physiology of Speech production: Results and implication of quantitative cineradiographic study*. Cambridge, MA: MIT Press, 1969.

R Core Team. R: A language and environment for statistical computing, 2017. URL `https://www.R-project.org/`.

James M. Scobbie, Eleanor Lawson, Steve Cowen, Joanne Cleland, and Alan A. Wrench. A common co-ordinate system for mid-sagittal articulatory measurement. In *QMU CASL Working Papers*, pages 1–4, 2011.

Márton Sóskuthy. Generalised additive mixed models for dynamic analysis in linguistics: a practical introduction. arXiv preprint arXiv:1703.05339, 2017.

Patrycja Strycharczuk and James M. Scobbie. Velocity measures in ultrasound data. Gestural timing of post-vocalic /l/ in English. In *Proceedings of the 18th International Congress of Phonetic Sciences*, pages 1–5, 2015.

Yolanda Vazquez-Alvarez and Nigel Hewlett. The ‘trough effect’: an ultrasound study. *Phonetica*, 64 (2-3):105–121, 2007.

John R. Westbury. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *The Journal of the Acoustical Society of America*, 73(4):1322–1336, 1983.

Simon Wood. *Generalized additive models: An introduction with R.* CRC Press, 2006. ISBN 1584884746.

Alain F. Zuur. *A beginner’s guide to generalized additive models with R.* Highland Statistics Limited: Newburgh, 2012. ISBN 0957174128.