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

It is known that the root of the tongue can play a role in maintaining voicing during the closure of voiced stop consonants (Halle and Stevens 1967, Kent and Moll 1969, Perkell 1969, Westbury 1983). The production of vocal fold vibration requires a pressure differential between the sub-glottal and the supra-glottal cavities (with lower pressure in the supra-glottal cavity). During the production of voiced obstruents, the pressure in the supra-glottal cavity quickly increases, due to the additional air injected from the lungs in the supra-glottal cavity, which is completely sealed in stops. Such pressure increase can hinder the ability to maintain voicing during closure, at the point that voicing can stop if the lowest threshold of pressure differential is reached and surpassed.

Westbury (1983) argues that one way to counterbalance the pressure increase in the supra-glottal cavity is to enlarge the cavity through the advancement of the tongue root. Drawing from ultrasound tongue imaging, Ahn and Davidson (2016) demonstrate that the root of the tongue is advanced during the articulation of voiced consonants in American English. They also showed that tongue root advancement is present even when vocal fold vibration is not implemented during closure in underlyingly voiced stops. An interesting question arising from the connection between voicing and tongue root is weather the advancement of the root is correlated with other phonetic characteristics, like the duration of vowels preceding stops. Such hypothesis will be expanded on in this section, after a brief overview of work on the effect of consonant voicing on vowel duration.

An extensive pool of studies shows that vowels tend to be longer when followed by voiced obstruents and shorter when followed by voiceless obstruents (House and Fairbanks 1953, Chen 1970, Klatt 1973, Lisker 1973:just to mention a few). Most of the literature on this phenomenon, known as the "voicing effect", suggests that different languages show different magnitudes of such durational differential, and that in some other languages the duration of vowels is not affected by the voicing of the following obstruent.[1] Although several attempts have been put forward to explain the effect of voicing on vowel durations, no consensus has been reached to date. Nonetheless, a recurrent theme focusses on the differences that characterise the gestural implementation of voiced and voiceless stops.[2]

One of the earliest articulatory accounts of the voicing effect attributed the difference in vowel duration to the divergent configuration of the vocal folds between voicing in sonorants and in obstruents (Halle and Stevens 1967; reiterated in Chomsky and Halle 1968). According to Halle and Stevens (1967), voicing in obstruents is produced with a state of the glottis that is different from the configuration necessary to produce vocal fold vibration in sonorants like vowels. On the contrary, they claim that voiceless stops do not require any specific glottal configuration and thus the voicing of the preceding vowel can just naturally cease at closure (or a few milliseconds after it). Halle and Stevens (1967) thus hypothesise that, to allow the glottal state to change from the sonorant voicing

*Corresponding author

[1] For a different opinion on the first matter, see Laeufer (1992).

[2] However, see Javkin (1976) and Kluender et al. (1988) for two perceptually inclined proposals.

of the preceding vowel to obstruent voicing, the vowel is lengthened so that enough time is available for such adjustments to happen.

Although such account seemed promising at the time it was proposed, later studies failed to demonstrate that obstruent voicing is any different from sonorant voicing []. Given the established connection between voicing and tongue root advancement, the hypothesis follows that tongue root advancement could also be linked to vowel duration. If this were the case, a language in which vowels have different durations depending on the voicing of the following consonant should also show tongue root advancement in voiced stops, while tongue root advancement should not be employed in those languages in which vowel durations are not affected by voicing. Building on the hypothesis in Halle and Stevens (1967), I put forward an account in which a relatively more complex tongue gesture in voiced consonants requires a longer time to be achieved (Section 4).

In this paper, I focus on two languages, Italian and Polish, that have been reported to show and lack the voicing effect respectively. In a study assessing general properties on segmental durations of spoken Italian, Farnetani and Kori (1986) shows that the first vowel in /lada/ is on average 35 milliseconds longer than the vowel in /lata/ (/lata/ 223 msec, sd = 18; /lada/ 258 msec, sd = 13, p. 26). Esposito (2002) extends Farnetanis's research to all vowels and stops and demonstrates that vowels are longer when followed by a voiced stop, with an estimate similar to what reported in Farnetani and Kori (1986). On the other hand, Keating (1984) reports that vowels in Polish are not affected by the voicing of the following consonant. The average duration of the first vowel in /rata/ is 167.5 milliseconds, while the preconsonantal vowel in /rada/ is just two milliseconds longer. Based on the hypothesis put forward that there is a link between tongue root advancement and the voicing effect, it is expected that Italian will show tongue root advancement, while Polish will lack such articulatory gesture in the implementation of voiced stops.

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

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 North-west and one from Northeast). The Polish group was more heterogeneous, with two speakers from the North-west (Poznań), and two from the North-east (Warsaw and Przasnysz). Ethical clearance was obtained for this work 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

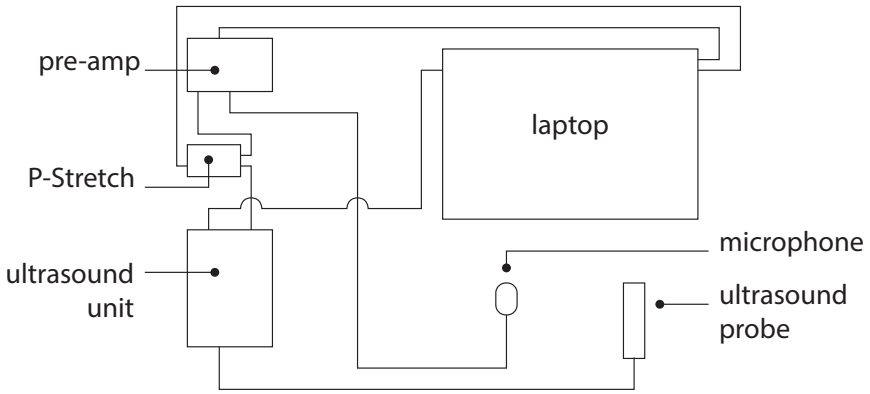

Figure 1: Schematic representation of the equipment setup (Articulate Instruments Ltd 2011, see text for details).

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, not shown in figure).

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 influence on 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. 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 varied depending on the frame rate (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 data were then subject to force alignment using the SPPAS force aligner (Bigi 2015). The outcome of the automatic annotation was then manually corrected, 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), employing a implementation of the algorithm described in Ananthapadmanabha et al. (2014). The durations of the following intervals were extracted from the 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).

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)

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 by applying offsetting and rotation relative to the participant's occlusal plane (Scobbie et al. 2011). Durational measurements were analysed with linear mixed effects models using `lme4` in R (R Core Team 2017, Bates et al. 2015). Generalised additive mixed effects regression models (GAMMs, Wood 2006) were used for the statistical analysis of tongue contour data.

3. Results

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 QUALITY (/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. P-values were obtained with likelihood ratio tests comparing the full model with a nested model without the tested predictor, and with `lmerTest` [], which employs the Satterthwaite approximation to degrees of freedom. According to the full model as specified above, Italian vowels are 22 milliseconds (SE = 6) longer if followed by a voiced stop ($\chi 2(3) = 16.61$; $p < 0.001$). 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, excluding the voicing-vowel 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 ($\chi 2(1) = 5.4$, $p < 0.05$). Vowel identity and sentence duration were also significant. The place of C2 significantly improved the model ($\chi 2(1) = 6.1$, $p < 0.05$), so it was included in the full model even though $p > 0.05$ according to the single predictor p-value. The inspection of the model residuals confirmed the assumptions of normality and homoscedasticity. The exploration of the random slopes for each speaker indicated that PL05 showed a particularly higher slope for voicing, meaning that the effect of voicing was stronger in his data, but removing this speaker from the model doesn't remove the effect. The estimated effect of voicing on vowel duration for PL05 was 14 milliseconds. These observations will become relevant when discussing about the results of the tongue contour data.

3.2. Tongue contours

The tongue contour data were analysed with GAMMs (Wood 2006). Individual GAMMs were fitted for each speaker: 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 QUALITY, 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 residuals. Significance testing in GAMMs was achieved through model comparison and visual inspection of the difference smooth, as suggested in (Sóskuthy 2017). Given the poor quality of the ultrasonic data for /u/, this vowel was not included in the statistical analysis of tongue contours, hence the results reported in this section only refer to /a/ and /o/.

The analysis of the Italian ultrasonic data shows that voiced stops are produced with advancement of the root of the tongue, as expected based on previous research on English. Figure 2 (top half) shows the predicted tongue contours in voiceless (dashed line) and voiced stops (thick line) at maximum tongue displacement for each Italian speaker. Below each tongue contour panel, the difference smooth for voicing is also shown (black line, confidence interval in grey). Tongue contours are significantly different in those point 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 grey area.

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 2), 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.

Further 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 in advanced in voiced consonants (). 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 3 shows the results for IT01 as a representative example and for PL05.

4. Discussion

Based on the established link between tongue root and voicing, it was proposed at the beginning of the 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

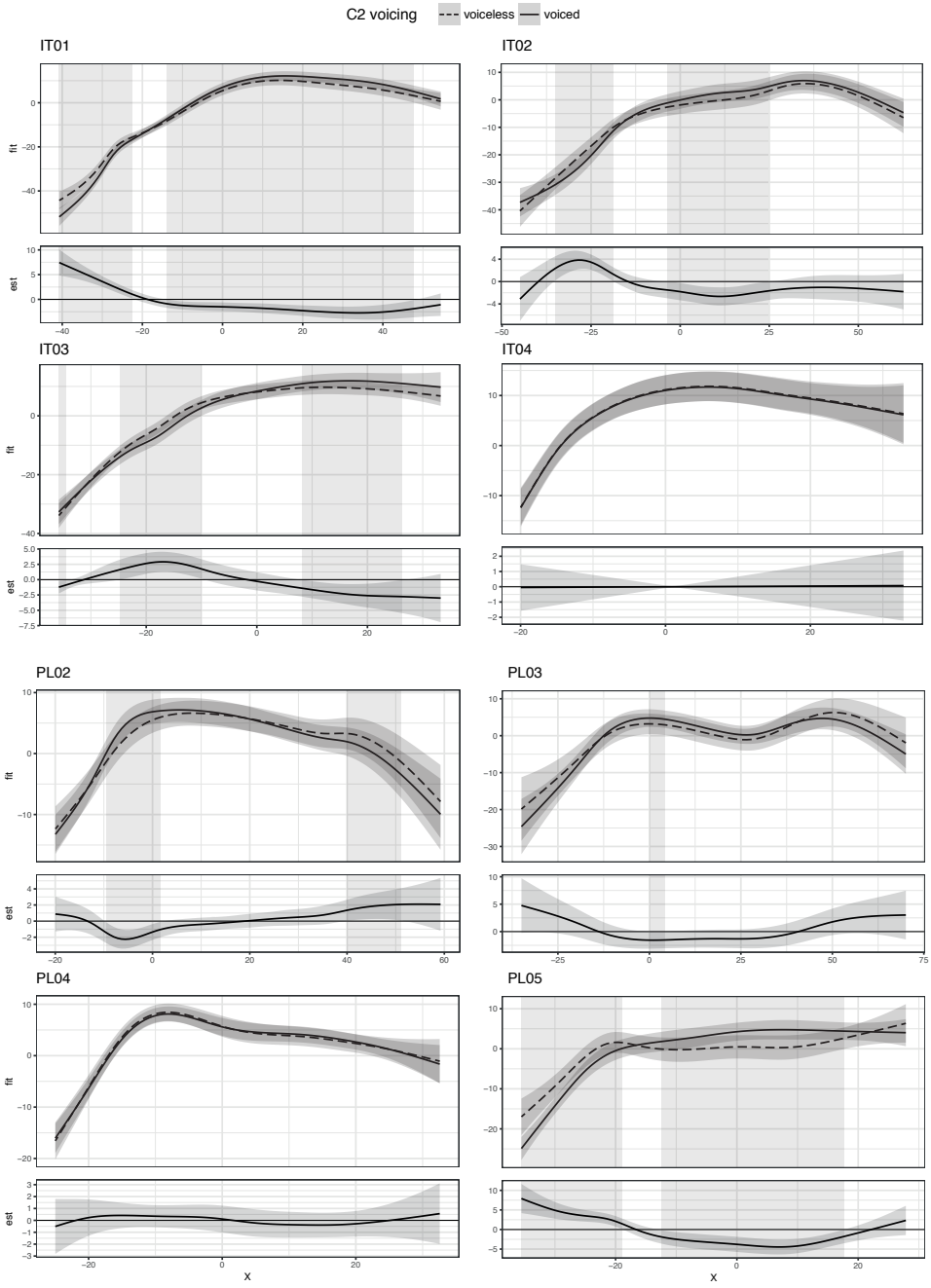

Figure 2: Comparison of tongue contours at maximum tongue displacement (within C2 closure) in Italian (top half) and Polish (bottom half). The plotted contours are reference contours for the coronal consonants preceded by /a/. See Section 3.2 for more details.

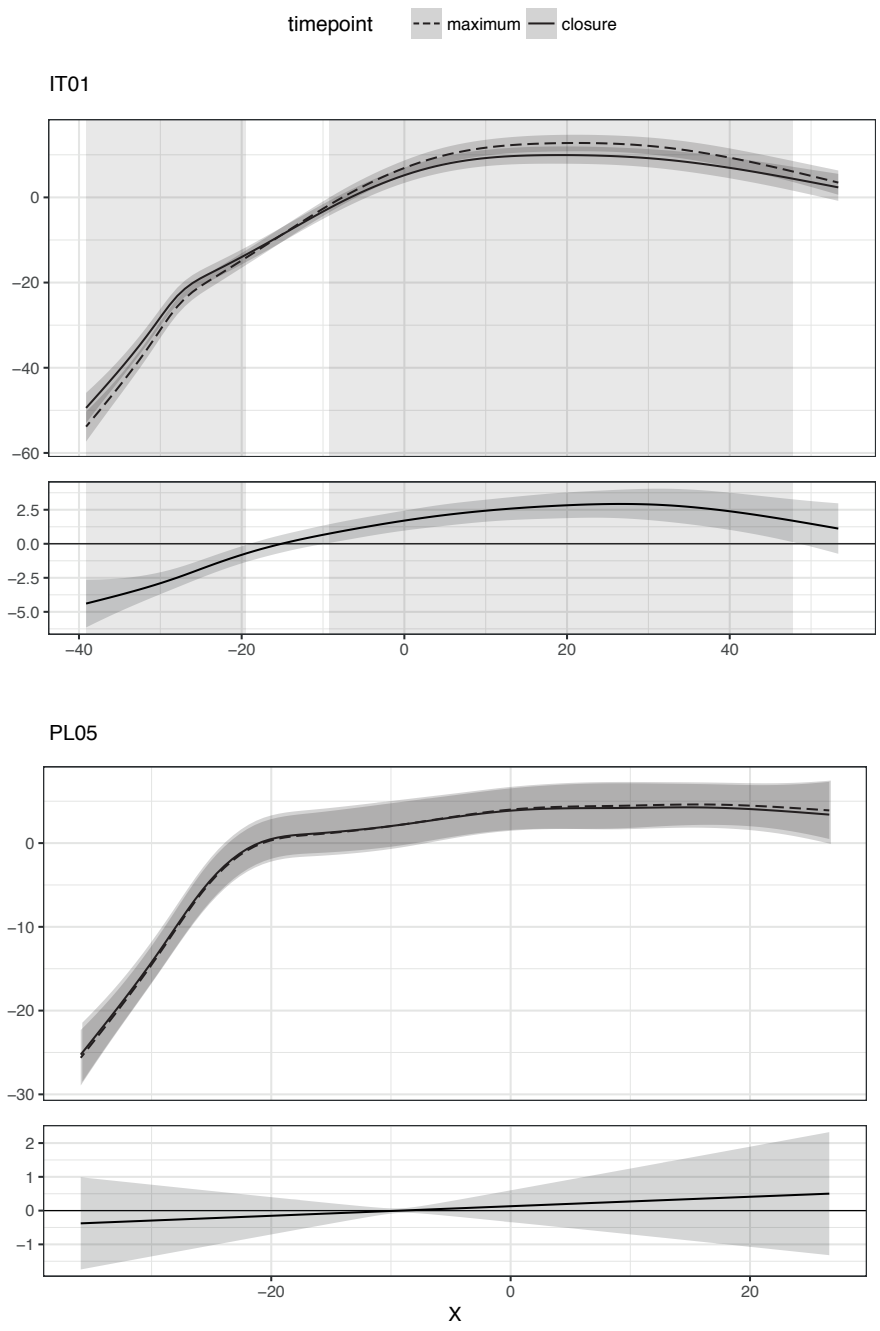

Figure 3: Comparison of tongue contours of voiced consonants at C2 closure onset and maximum tongue displacement (within C2 closure) in IT01 (Italian) and PL05 (Polish). See text for more details.

two languages with and without the voicing effect, Italian and Polish respectively. The data from most speakers pattern according to the hypothesis that a link subsists between longer vowels following voiced stops and tongue root advancement in that type of consonants.

Following the reasoning of the proposal by Halle and Stevens (1967), a plausible cause for the longer duration of vowels before voiced consonants with tongue root advancement (as in Italian) could therefore be the additional time required to achieve a more complex consonantal gesture—a gesture that requires adjustments of both the tongue root and the tongue dorsum/tip. The reported absence of the voicing effect in Polish could then be ascribed to the absence of tongue root advancement in the production of voiced consonants in this language. Complications arise from the fact that we do find a small but significant vowel duration difference in Polish, and therefore cannot claim that the voicing effect is entirely absent from Polish, in contrast to previous findings by Keating (1984). Moreover, tongue root advancement was found in one of the Polish speakers on one hand (PL05), and it was absent from one of the Italian speakers on the other (IT04).

It is noteworthy that, as mentioned above, PL05 had a strikingly higher slope estimate for the effect of voicing on vowel duration, compared to the other Polish speakers, meaning that the voicing effect in his data was stronger.[3] Incidentally, PL05 is also the only Polish speaker who produced voiced consonants with an advanced tongue root. Furthermore, a model comparing tongue root at closure onset versus maximum tongue displacement in PL05 indicated that there was no difference in tongue root advancement at these two time points (). Assuming the effect in the other Polish speakers is small enough to be discarded as an artefact (see footnote 3), it follows that, independently of the language, the presence of tongue root advancement in voiced stops correlates with a concomitant increased duration in vowels preceding voiced consonants.

Given the smaller effect of voicing in PL05 compared to the effect in Italian (14 vs 19.5 milliseconds), a possible hypothesis could be that the correlation between tongue root advancement and vowel duration is gradual, rather then categorical. In this case, the magnitude of the voicing effect should correlate with the amount of tongue root advancement even *within speaker*, or at least across speakers independently of their language. Since the durational difference in the Polish speaker PL05 was quite small, one expects the magnitude of the advancement of the root to be proportionately smaller for this speaker. Future work will set out to investigate the hypothetically gradient correlation between vowel duration and amount of tongue root advancement.

In addition to tongue root advancement, the ultrasonic data also showed raising of the tongue dorsum. The presence of such gesture, although not expected, makes sense from an anatomical point of view. Raising of the tongue body could be implemented as a way to counterbalance the compression of the tongue mass caused by the advancement of the root. It is not thus surprising to observe a raised tongue body in voiced stops accompanying root advancement. An alternative account could ascribe tongue body raising to aerodynamic properties of voiced stops. Since the intra-oral pressure is higher in voiced stops due to the amount of air needed to maintain voicing, a firmer seal at the point of oral constriction could be used to compensate for the increasing pressure. Expanding the area of contact by raising the tongue body would provide for such a firmer constriction.

A possible critique to the account proposed here is that, if an active gesture for maintaining voicing is required during the closure of voiced stops, then it is not clear how the Polish speakers can maintain voicing without tongue root advancement. However,

[3] Note that excluding PL05 from the durational data produced an estimated difference of vowel duration of 6 milliseconds, which still calls for an explanation. Given the small magnitude of the effect, however, it is likely that such effect is an artefact of the difficulty of segmenting vowel to consonant transitions when the consonant is voiced [@allen1978]. Moreover, such downside would not apply to the data in PL05 given the larger estimates for the effect of voicing and the random slope, as discussed above.

root advancement is not the only solution: manipulations of the larynx or of the velopharyngeal port, rather than the tongue, can also counterbalance the increased intra-oral pressure []. The gestural timing of the larynx and the velopharyngeal port are (at least partially) anatomically independent from the timing of tongue gestures [], and would ideally not require a more complex planning as with an articulatory implementation of multiple tongue gestures.

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.

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.

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.

Hector R Javkin. The perceptual basis of vowel duration differences associated with the voiced/voiceless distinction. *Report of the Phonology Laboratory, UC Berkeley*, 1:78–92, 1976.

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.

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.

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

Leigh Lisker. On "explaining" vowel duration variation. In *Proceedings of the Linguistic Society of America*, pages 225–232, 1973.

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.