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

Stefano Coretta

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. The production of vocal fold vibration requires a pressure differential between the subglottal 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) argued that one way to counterbalance the pressure increase in the supra-glottal cavity is to enlarge the cavity through expansion of the pharyngeal walls. One way to achieve this is to advance the root of the tongue. Ahn and Davidson (2016) have recently demonstrated, drawing from ultrasound tongue imaging, 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.

An extensive pool of studies showed 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 the topic 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 in sonorant and obstruent voicing (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 perpetuated during the vowel can just naturally ceases at closure (or a few milliseconds after it). The authors thus hypothesise that, to allow the glottal state to change from sonorant voicing to obstruent voicing, the vowel is lengthen so that enough time is available for the 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 hypothe-

[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.

sis 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. On the same line of the hypothesis in Halle and Stevens (1967), I put forward an account in which a more complex tongue gesture in voiced consonants requires a longer time to be achieved (Section 4).

In a study assessing general properties on segmental durations of spoken Italian, Farnetani and Kori (1986) found that the first vowel in /lada/ was on average 35 msec longer than the vowel in /lata/ (/lata/ 223 msec, sd = 18; /lada/ 258 msec, sd = 13, p. 26). Esposito (2002) extended Farnetanis's research to all vowels and stops and found that vowels were longer when followed by a voiced stop, with an estimate similar to what reported in Farnetani and Kori (1986). Vowels in Polish, on the other hand, are not affected by the voicing of the following consonant, according to Keating (1984). For these reasons, Italian and Polish have been chosen as the two test languages for this study.

2. Methodology

2.1. Participants

Eight native speakers of Italian (2 females, 2 males) and Polish (2 females, 2 males) have been 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 Poznań, one from Przasnysz, and one from Warsaw. This research has obtained ethic clearance from the University of Manchester (REF 2016-0099-76). The participants received a small monetary compensation.

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

2.2. Equipment set-up

An Articulate Instruments Inc. set-up was used for this study (Figure 1). This was constituted by 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 Inc. (not shown in the 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 cf. Vazquez-Alvarez and Hewlett 2007). Only coronal and velar stops were used as target consonants since

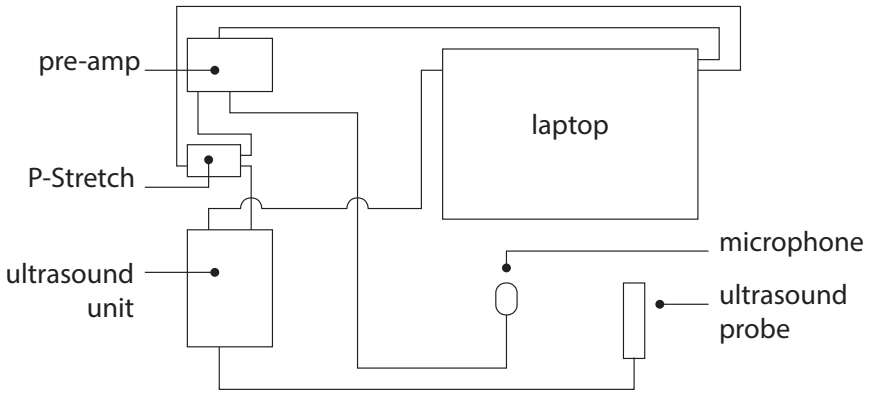

Figure 1: Schematic representation of the equipment setup (Articulate Instruments Ltd 2011, see text for details).

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, and the hard palate was imaged by asking the participant to swallow water (Scobbie et al. 2011). The frame rate of the acquisition of the ultrasonic data varied between 55 and 65 frames per second (one frame every 15-18 milliseconds). The audio signal was recorded at 22050 MHz (16-bit).

2.5. Data processing

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 then 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).

Tongue contours were extracted from the ultrasonic data using AAA. Spline curves were first 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). Fan line selection in this study was achieved by finding the fan line within the relevant area of the tongue (tongue tip for coronal consonants and tongue dorsum for velar consonants) with the highest standard deviation of displacement.

2.6. Analysis

The tongue contours coordinates were exported at two time points: (1) at the onset of C2 closure, and

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)

(2) at maximum tongue displacement (within C2 closure). The contours were normalised by applying offsetting and rotation relative to the participant's occlusal plane (Scobbie et al. 2011). Generalised additive mixed effects regression models (GAMMs, Wood 2006) were used for the statistical analysis of tongue contour data in R (R Core Team 2017). Duration measurements were subject to linear mixed effects models using `lme4` in R (Bates et al. 2015).

3. Results

3.1. Vowel duration and voicing

A linear mixed effects regression model was fitted on 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 through likelihood ratio tests comparing the full model including voicing with a null model without voicing as a predictor. According to the full model, Italian vowels are 19.5 milliseconds (±5.5 standard errors) longer if followed by a voiced stop ($\chi^2(3) = 18.5$, p = 0.000337).

For Polish, the same model structure was used, excluding the voicing-vowel interaction (which was not significant). Surprisingly, the model reported a partially significant 8 milliseconds (±3 standard errors) effect of consonantal voicing on the preceding vowel ($\chi^2(1) = 5.4$, p = 0.02). 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. The estimated effect of voicing on vowel duration for PL05 was 14 milliseconds. This observation will come handy when discussing about the results of the tongue contour data.

3.2. Tongue contours

Given the poor quality of the ultrasonic data for /u/, this vowel was not included in the statistical analysis. Significant testing in GAMMs was achieved through model comparison and visual inspection of the difference smooth, as suggested in (Sóskuthy 2017). The analysis of the Italian ultrasonic data showed that voiced stops are produced with advancement of the root of the tongue (), as expected based on previous research on English. 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,

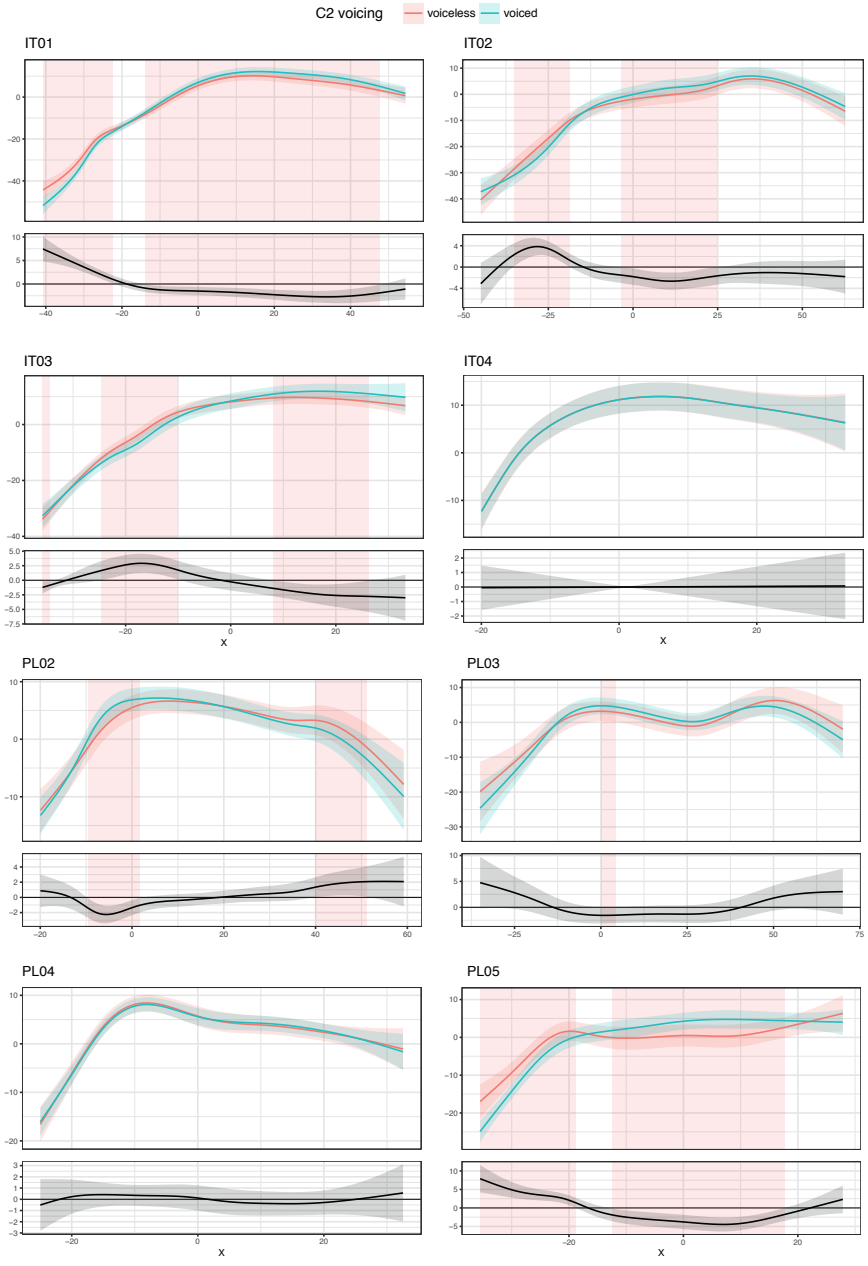

Figure 2: Comparison of tongue contours at maximum tongue displacement (within C2 closure) in Italian (top half) and Polish (bottom half). The plotted contours are reference contours for the coronal consonants preceded by /a/. See text for more details.

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.

Figure 2 (top half) shows the predicted tongue contours in voiceless (red) and voiced stops (blue) 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 red 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, and on the account by Halle and Stevens (1967), 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, if the latter has a link with the durational differences found in vowels before voiced stops. The hypothesis was that the additional time required for the tongue to reach an advanced position in voiced stops could be compensated for during the preceding vowel, thus lengthening in comparison to vowels followed by voiceless stops. To test the correlation between tongue root advancement and vowel durations, ultrasonic data were collected from two languages with and without the voicing effect, Italian and Polish respectively.

The generalised presence of tongue root advancement in Italian but not in Polish, as found in the data from this study, provides initial support to the proposed link between longer vowels following voiced stops and tongue root advancement. A plausible cause for the longer duration of vowels before voiced consonants with tongue root advancement (as in Italian) is that a more complex tongue gesture—a gesture that requires adjustments of both the tongue root and the tongue dorsum/tip—requires longer time to be achieved. Protracting the preceding vocalic gesture in time could then be a solution to allow for the required additional time. 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. However, two complications derive from the data presented in this study. First, tongue root advancement was found in one of the Polish speakers (PL05) on one hand and it was absent from one of the Italian speakers (IT04) on the other. Second, vowels followed by voiced stops in Polish were 8 milliseconds longer in Polish, contrary to what argued in Keating (1984).

The first issue can be overcome by the observation that, as mentioned above, PL05 had a strikingly higher slope estimate for the effect of voicing on

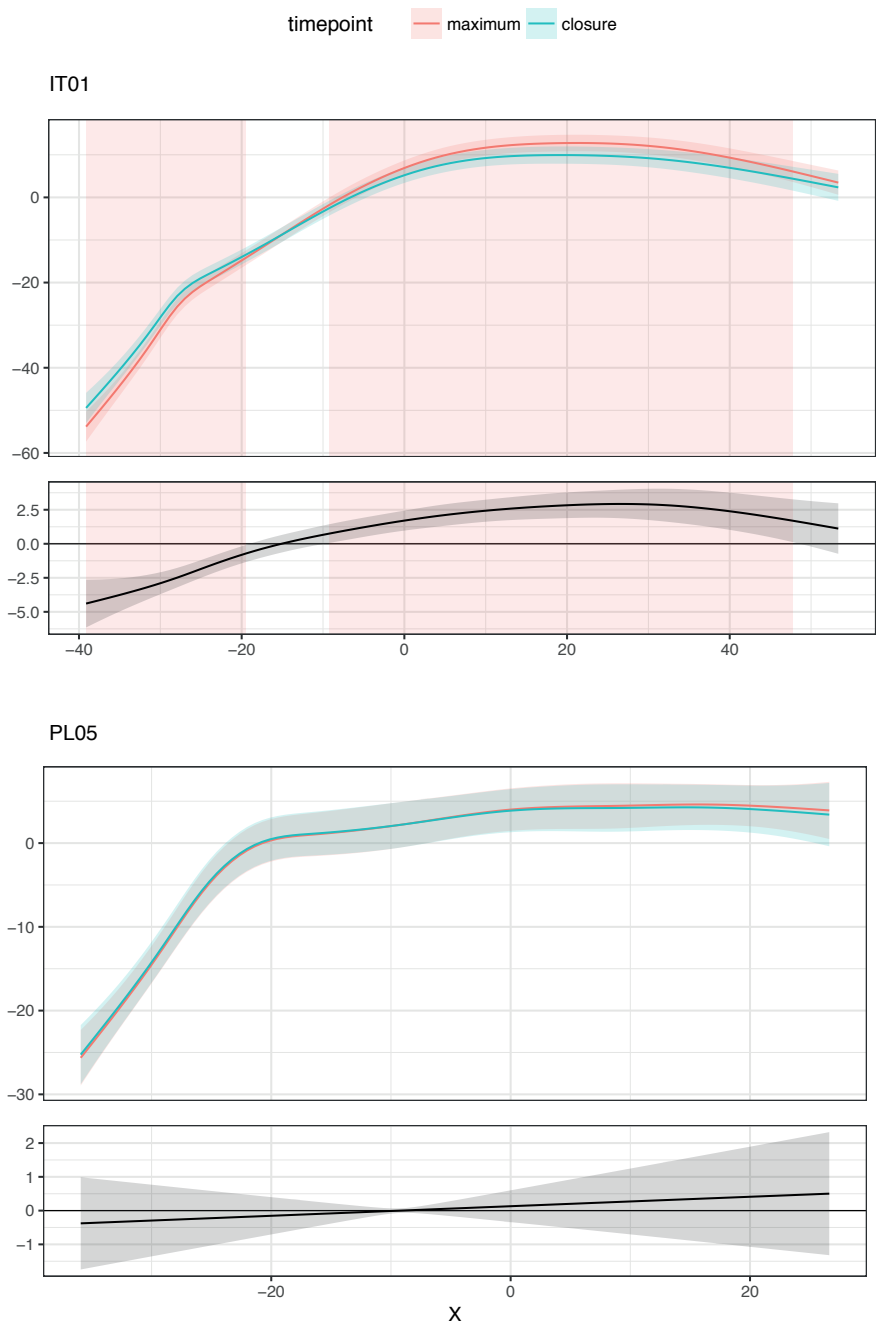

Figure 3: Comparison of tongue contours of voiced consonants at C2 closure onset and maximum tongue displacement (within C2 closure) in IT01 (Italian) and PL05 (Polish). See text for more details.

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 gradual correlation between vowel duration and amount of tongue root advancement.

Finally, the ultrasonic data showed raising of the tongue dorsum concomitant to root advancement. 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, 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 that of multiple tongue gestures.

[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 []. 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.

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. 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.

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.

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.

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.