Vowel duration and tongue root advancement: Results from an exploratory study of the relation between voicing and vowel duration

Stefano Coretta

1 Introduction

This paper reports a correlation between vowel duration and degree of tongue root advancement. In an exploratory study of the articulatory correlates of stop voicing, it has been found that tongue root advancement—a known mechanism that facilitates voicing during stop closure—can be implemented not only during the closure of a stop, but even during the production of the vowel preceding a stop. Moreover, vowel acoustic duration turned out to be linearly correlated with degree of tongue root advancement, such that longer vowels show greater tongue root advancement.

It is well known that voiced stops (broadly defined) are almost universally accompanied by two phonetic correlates: advanced tongue root and preceding longer vowel durations.

1.1 Tongue root position and voicing

One of the differences in supra-glottal articulation between voiced and voiceless stops concerns the position of the tongue root relative to the antero-posterior axis of the oral tract. The realisation of vocal fold vibration (i.e. voicing) requires a difference in air pressure between the cavities below and above the glottis. Specifically, the sub-glottal pressure needs to be higher than the supra-glottal pressure for voicing to be maintained. This property of voicing is formally known as the Aerodynamic Voicing Constraint (Ohala 2011). When the oral tract is completely occluded during the production of a stop closure, the supra-glottal pressure quickly increases, due to the incoming airstream from the lungs. Such pressure increase can hinder the ability to sustain vocal fold vibration during closure, to the point voicing ceases.

An articulatory solution to counterbalance the increased pressure is to enlarge the supra-glottal cavity by advancing the root of the tongue. It has been repeatedly observed that the tongue root is in a more front position in voiced stops compared to voiceless stops (Kent & Moll 1969; Perkell 1969; Westbury 1983). Rothenberg (1967) calculates that the walls of the supraglottal cavity can absorb the incoming airflow for 20 to 30 ms by passive expansion, after which the sub- and supra-glottal pressures would equalise and voicing cease. Rothenberg (1967) thus argues that a passive expansion of the pharyngeal walls is not sufficient.

According to Rothenberg (1967), a complete ballistic forward gesture of the tongue root has a time constant of 70 to 90 ms. Given that stop closures are generally shorter than that, it is natural that advancement is initiated during the vowel, so that an appreciable amount of advancement is obtained when closure is achieved. Furthermore, Westbury (1983) finds

that tongue root advancement is initiated before achievement of full closure and that there is a forward movement even in the context of voiceless stops.

However, the relationship between tongue root advancement and voicing is a complex one. First, tongue root advancement is not the only mechanism for sustaining voicing during a stop (Rothenberg 1967; Westbury 1983; Ohala 2011) and it has a certain level of idiosyncrasy (Ahn & Davidson 2016). Other solutions include expansion of the lateral walls of the pharynx [], larynx lowering (Riordan 1980), opening of the velopharyngeal port (Yanagihara & Hyde 1966), producing a retroflex occlusion (Sprouse et al. 2008). Second, implementation of tongue root advancement can be decoupled from the presence of actual vocal fold vibration. In Westbury (1983), advancement of the tongue root is found even in the context of voiceless stops, which is counterintuitive given that tongue root advancement is generally considered to be a feature of voiced stops. Similarly, Ahn (2015); Ahn & Davidson (2016); Ahn (2018) find that the tongue root is more advanced in the phonologically voiced stops independent of whether they actually show vocal fold vibration or not.

1.2   Vowel duration and voicing

A great number of studies showed that, cross-linguistically, vowels tend to be longer when followed by voiced obstruents than when they are followed by voiceless obstruents (House & Fairbanks 1953; Peterson & Lehiste 1960; Chen 1970; Klatt 1973; Lisker 1974; Farnetani & Kori 1986; Fowler 1992; Hussein 1994; Esposito 2002; Lampp & Reklis 2004; Durvasula & Luo 2012). This so-called 'voicing effect' has been reported in a variety of languages, including (but not limited to) English, German, Hindi, Russian, Arabic, and Korean, Italian, and Polish (see Maddieson & Gandour 1976 and Beguš 2017 for a more comprehensive list).[1] The existence of the voicing effect is supported by abundant empirical evidence, although no agreement has been reached regarding its causes (Durvasula & Luo 2012; Sóskuthy 2013).

Coretta (2018) presents results on vowel durations in Italian and Polish based on the acoustic data of the study discussed here. The data indicates that the raw mean difference in vowel duration before voiceless vs. voiced stops in Italian and Polish is about 11.5 and 7.5 ms respectively. Linear mixed modelling suggests an effect of 16 ms (SE = 4.4) in both languages (see Coretta 2018 for details).

1.3   This study

To summarise, tongue root advancement and longer vowel durations are two correlates of voicing (broadly defined). Previous studies have shown that voicing can be maintained by advancing the tongue root during the production of voiced stops and that vowels followed by voiced stops tend to be longer than vowels followed by voiceless stops. The results from this exploratory study of Italian and Polish show that tongue root advancement and longer vowel durations are also directly linked in different ways.

---

[1]While Keating (1984) finds no statistical difference in vowel durations before voiceless vs. voiced stops, Nowak (2006) reports a 4.5 ms effect and the data in Malisz & Klessa (2008) suggest an effect of 3.5 ms. Beguš (2017) argues that the null finding in Keating (1984) could be due to low statistical power.

2 Methodology

2.1 Participants

Participants were recruited in Manchester (UK), and Verbania (Italy) Eleven native speakers of Italian (5 females, 6 males) and 6 native speakers of Polish (3 females, 3 males) participated in this study. Most speakers of Italian are originally from the North of Italy, while 3 are from Central Italy. The Polish speakers came from different parts of Poland (2 from the west, 3 from the centre, and 1 from the east). This study has been approved by the SALC Ethics committee of the University of Manchester (REF 2016-0099-76). The participants signed a written consent and received a monetary compensation of £10.

2.2 Equipment

Simultaneous recordings of audio and ultrasound tongue imaging were obtained in the Phonetics Laboratory at the University of Manchester (UK) and in a quiet room in Verbania (Italy). An Articulate Instruments Ltd™ system was used for this study. The system is made of a TELEMED Echo Blaster 128 unit, an Articulate Instruments Ltd™ P-Stretch synchronisation unit, and a FocusRight Scarlett Solo pre-amplifier (see **??**). A TELEMED C3.5/20/128Z-3 ultrasonic transducer (20mm radius, 2-4 MHz) and a Movo LV4-O2 Lavalier microphone were used respectively for the acquisition of ultrasonic and audio data. The ultrasonic probe was placed in contact with the sub-mental triangle, aligned with the mid-sagittal plane. A metallic headset designed by Articulate Instruments Ltd™ (2008) was used to hold the probe in a fixed position and inclination relative to the head. The acquisition of the mid-sagittal 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. The synchronisation of the ultrasonic and audio signals was performed by AAA after recording by means of a synchronisation signal produced by the P-Stretch unit. The ranges of the ultrasonic settings were: 43-68 frames per second, 88-114 number of scan lines, 980-988 pixel per scan line, field of view 71-93°, pixel offset 109-263, depth 75-180 mm. The audio signal was sampled at 22050 Hz (16-bit).

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.), giving a total of 12 target words, used both for Italian and Polish.[2] The resulting words are nonce words, with a few exceptions, and they were presented in the languages' respective writing conventions (see Appendix A). A labial stop was chosen as the first consonant to reduce possible coarticulation with the following vowel.[3] Central/back vowels only were included in the target words for

---

[2]Note that stressed vowels in open syllables in Italian are long (Renwick & Ladd 2016). Moreover, /o/ is used here for typographical simplicity to indicate the mid-back vowels of Italian and Polish, although they do differ in quality. See Krämer (2009); Renwick & Ladd (2016); Gussmann (2007).

[3]However, note that Westbury (1983) and Vazquez-Alvarez & Hewlett (2007) report tongue body lowering in the context of labial stops.

two reasons. First, high and mid front vowels tend to be difficult to image with ultrasound, given their greater distance from the ultrasonic probe when compared with back vowels. Second, high and mid front vowels usually produce less tongue displacement from and to a stop consonant. This characteristic can make it more difficult to identify gestural landmarks using the methodology discussed in Section 2.5. Since the focus of the study was to explore differences in the closing gesture of voiceless and voiced stops, only lingual consonants have been included, since of course the closure of labial stops cannot be imaged with ultrasound. The sentence *Dico X lentamente* 'I say X slowly' in Italian, and *Mówię X teraz* 'I say X now' for Polish functioned as frames for the test words.[4] Speakers were instructed to read the sentences without pauses and to speak at a comfortable pace.

2.4 Procedure

The participants familiarised themselves with the sentence stimuli at the beginning of the session. Headset and probe were then fitted on the participant's head. The participant read the sentence stimuli, which were presented on the computer screen in a random order, while the audio and ultrasonic signals were acquired simultaneously. The random list of sentences was read 6 times consecutively (with the exception of IT02, who repeated the sentences 5 times only). Due to software constraints, the order of the sentences within participant was kept the same for each of the six repetitions. The participant could optionally take breaks between one repetition and the other. Sentences with hesitations or speech errors were immediately discarded and re-recorded. A total of 1212 tokens (792 from Italian, 420 from Polish) were obtained.

2.5 Data processing and statistical analysis

The audio data was subject to force alignment using the SPeech Phonetisation Alignment and Syllabification software (SPPAS, Bigi 2015). The outcome of the automatic alignment was then manually corrected, according to the recommendations in Machač & Skarnitzl (2009). The onset and offset of V1 in the $C_1V_1C_2V_2$ test words were respectively placed in correspondence of the appearance and disappearance of higher formant structure in the spectrogram. Vowel duration was calculated as the duration of the V1 onset to V1 offset interval. Speech rate is the number of syllables in the sentence (8 in Italian and 6 in Polish) divided by the duration of the sentence in seconds.

The displacement of the tongue root was obtained from the ultrasonic data according to the procedure used in Kirkham & Nance (2017). Smoothing splines were automatically fitted to the visible tongue contours in AAA. Manual correction was then applied in cases of clear tracking errors. A fan-like frame consisting of 42 equidistant radial lines superimposed on the ultrasonic image was used as the coordinate system. The origin of the 42 fan-lines coincides with the (virtual) origin of the ultrasonic beams, such that each fan-line is parallel to the direction of the ultrasonic scan lines. Tongue root displacement was thus calculated as the displacement of the fitted spline along a selected vector (Strycharczuk & Scobbie 2015). For

---

[4]Due to software constraints, the Polish frame sentence was presented on screen without diacritics. The speakers read them as if spelled correctly.

each participant, the fan-line with the highest standard deviation of displacement within the area corresponding to the speaker's tongue root was chosen as the tongue root displacement vector. A Savitzky–Golay smoothing filter (second-order, frame length 75 ms) was applied to the raw displacement. Displacement values for analysis are taken from the smoothed displacement signal. Tongue root displacement was obtained from a static time point (the onset of the closure of C2) and along the duration of the vowel. The displacement values along the vowel duration were extracted at time points corresponding to real ultrasonic video frames. Given the average frame rate is 55 frames per second, values are sampled about every 20 ms.

Statistical analysis was performed in R v3.5.2 (R Core Team 2018). Linear mixed-effects models were fitted with lme4 v1.1-19 (Bates et al. 2015). Factor terms were coded with treatment contrasts (the reference level is the first listed for each factor): C2 voicing (voiceless, voiced), vowel (/a/, /o/, /u/). Speech rate was centred for inclusion in the statistical models, by subtracting the mean speech rate across all speakers from the calculated speech rate values. Centring ensures the intercepts are interpretable. t-tests with Satterthwaite's approximation to degrees of freedom on the individual terms were used to obtain p-values using lmerTest v3.0-1 (Kuznetsova et al. 2017; Luke 2017). An effect is considered significant if the p-value is below the alpha level ($\alpha = 0.05$). Generalised additive mixed models were fitted with mgcv v1.8-26 (Wood 2011, 2017). The smooths used thin plate regression splines as basis (Wood 2003). The ordered factor difference smooths method described in Sóskuthy (2017); Wieling (2018) was used to model the effect of factor terms in GAMs. The models were fitted by maximum likelihood (ML) and autoregression in the residuals was controlled with a first-order autoregressive model. Significance testing of the relevant predictors was achieved by comparing the ML score of the full model with the score of a null model (in which the relevant predictor is dropped), using the `compareML()` function of the itsadug package (van Rij et al. 2017). A preliminary analysis indicated that including either language or C2 place of articulation as predictors produced respective p-values above the alpha level, without affecting the estimates of the other terms. For this reason, these variables were not included in the models reported here and will not be discussed. Future research is warranted to ascertain language-related differences and possible effects of place of articulation.

2.6 Open Science statement

3 Results

3.1 Tongue root position at C2 closure onset

Figure 1 shows raw data points and boxplots of the position of the tongue root at C2 closure onset when C2 is voiceless (left) and voiced (right). Since the actual value of tongue root posision (in mm) is not informative (it depends on the speaker's anatomy and on the probe location), scaled tongue root position is used here (note though that the raw data is used in statistical modelling). The trend is that, not surprisingly, the position of the tongue root is more advanced if C2 is voiced compared to its position when C2 is voiceless. A linear mixed-effects model with tongue root position as the outcome variable was fitted with the following predictors: fixed effects for C2 voicing (voiceless, voiced), centred speech rate (as
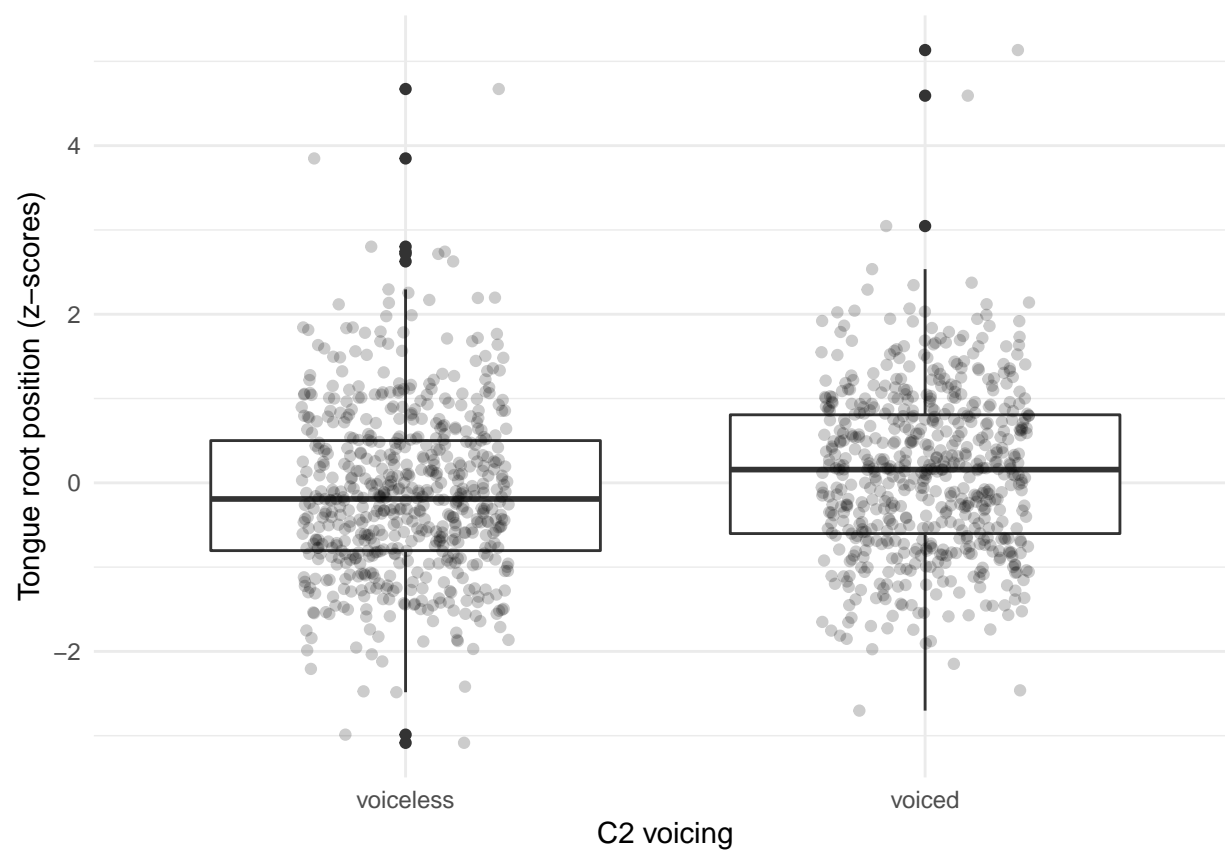
Figure 1: Fig caption

number of syllables per second, centred), vowel (/a/, /o/, /u/); by-speaker and by-word random intercepts (a by-speaker random coefficient for C2 voicing led to singular fit, so was not included in the final model). The effects of C2 voicing and vowel are significant according to t-tests with Satterthwaite's approximation to degrees of freedom. The tongue root at C2 closure onset is 0.77 mm (SE = 0.35) more front when C2 is voiced, and it is 1.87 mm (SE = 0.42) more retracted if V1 is /o/.

3.2 Tongue root position during V1

The position of the tongue root during the articulation of V1 was assessed with generalised additive mixed models (GAMM). A GAMM was fitted to tongue root position with the following terms: C2 voicing as a parametric term; a smooth term over centred speech rate, a smooth term over V1 proportion with a by-C2 voicing difference smooth, a tensor product interaction over V1 proportion and centred speech rate; a factor random smooth over V1 proportion by speaker (penalty order = 1). A chi-squared test on the ML scores of the full model and model excluding C2 voicing indicates C2 voicing significantly improves fit ($\chi(3)$ = 7.758, p = 0.001). **??** shows that the root advances during the production of the vowel, relative to its position at V1 onset. Such forward movement can be seen both in the context of a following voiced stop and in the one of a voiceless stop. However, the magnitude of the movement is greater in the former. At V1 offset, there is a difference in tongue root position of about 1 mm.

3.3 Correlation between tongue root position and V1 duration

A second linear mixed regression was fitted to tongue root position to assess the effect of V1 duration on root position. The following terms were included: centred V1 duration (in milliseconds), centred speech rate (as number of syllables per second), vowel (/a/, /o/, /u/), C2 place of articulation (coronal, velar); an interaction between centred V1 duration and vowel; by-speaker and by-word random intercept and a by-speaker random coefficient for V1 duration. All predictors and the V1 duration/vowel interaction are significant. V1 duration and tongue root position are positively correlated: The longer the vowel, the more advanced the tongue root is at V1 offset ($\hat{\beta}$ = 0.056 mm, SE = 0.015). The effect is stronger with /a/ than with /o/ and /u/ (see Figure 3).

3.4 Tongue root position during V2 as a function of V1 duration

The effect of V1 duration on tongue root position during V1 was modelled by fitting a GAMM with the following terms: tongue root position as the outcome variable, smooth terms over V1 duration and V1 proportion, a tensor product interaction over V1 proportion and V1 duration; a factor random smooth over V1 proportion by speaker (penalty order = 1). The full model with the tensor product interaction over V1 proportion and V1 duration has better fit according to model comparison with a model without the interaction ($\chi(3)$ = 12.559, p < 0.001). The general trend is that the forward movement of the root during the vowel is greater the longer the duration of the vowel (Figure 4). Moreover, the trajectory
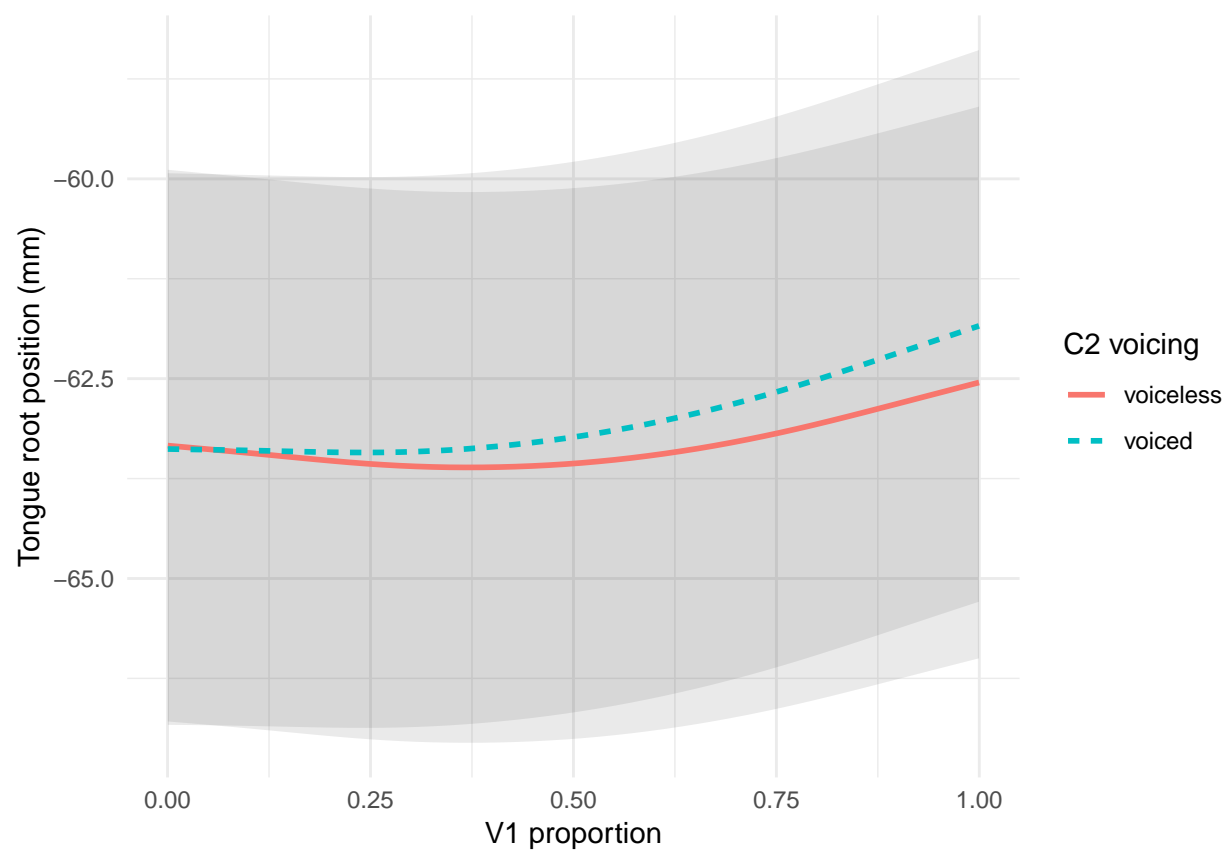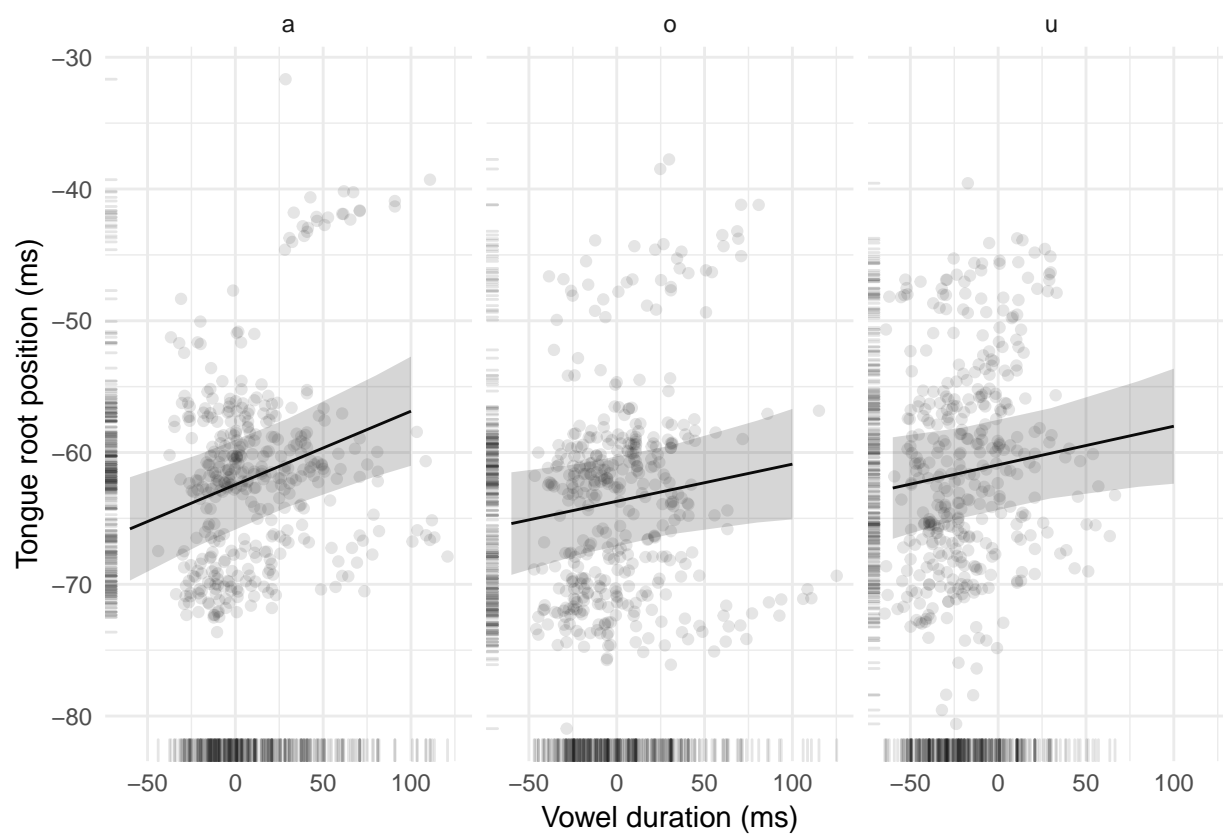
Figure 2: Fig caption
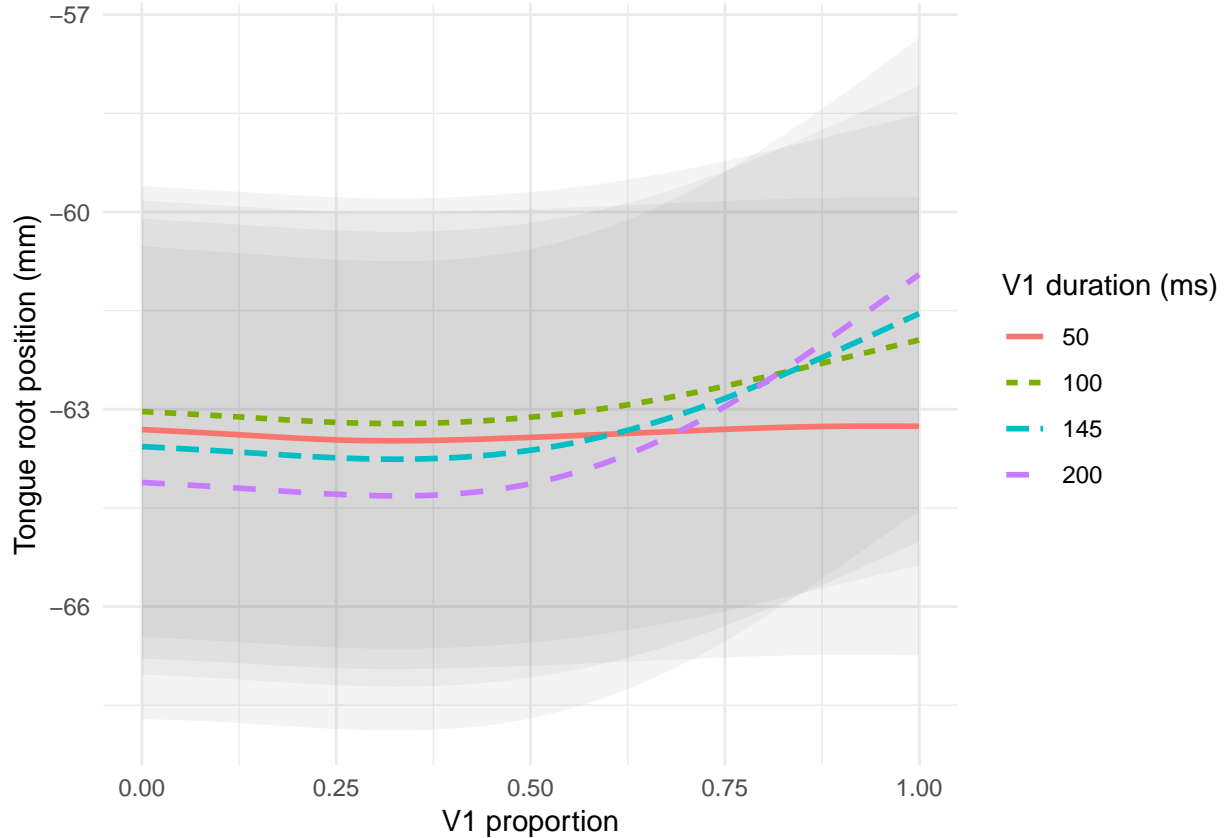
Figure 3: Fig caption

Figure 4: Fig caption

curvature increases with vowel duration: Shorter vowels have a flatter trajectory of tongue root advancement.

4 Discussion

The results from this exploratory study of voicing and vowel duration in Italian and Polish suggest there is a more direct relationship between tongue root and vowel duration. Unsurprisingly, the position of the tongue root at vowel offset if the following stop is voiced is more front than its position if the following stop is voiceless. This result is a replication of replicates the results in ... However, the data discussed here also suggest that tongue root position is positively correlated with vowel duration, such that longer vowels show a more advanced tongue root at vowel offset than shorter vowels. Said correlation exists independent of the voicing of the consonant following the vowel. When looking at the position of the tongue root during the vowel, it was found that the tongue root starts advancing during the articulation of the vowel. Vowels followed by voiced stops showed greater tongue root advancement at vowel offset then vowels followed by voiceless stops, in accordance with the static results. Moreover, a significant interaction between vowel duration and the trajectory shape was found: Shorter vowels have a flatter trajectory.

If we take the release of the consonant preceding the vowel as a reference point, a delayed

consonant closure could ensure that, by the time closure is made, an appreciable amount of tongue root advancement is achieved, so that the threshold for voicing cessation (such that transglottal pressure equalisation is reached) is raised (so that more time would be required for reaching it). This mechanism would ensure that voicing can be maintained during the stop closure.

Coretta (2018) argues that the temporal distance between two consecutive stop releases in CVCV words of Italian and Polish is not affected by the voicing of consonant following the stressed vowel (the first vowel). The duration of the so-called release to release interval is stable across voicing contexts. Within this interval, the timing of the onset of the stop closure onset produces differences in the respective durations of vowel and stop closure. A later closure onset implies a longer vowel and a shorter closure, while vice versa an earlier closure onset generates a shorter vowel and a longer closure. It is well known that the closures of voiced stops are longer that those of voiceless stops. As seen in this paper, the closure of voiced stops is initiated later (relative to the preceding consonant release) compared to the closure of voiceless stops. Given the temporal stability of the release to release interval, the delay in producing a full closure seen in the context of voiced stops has thus a double advantage: (1) A greater degree of tongue root advancement is achieved at closure onset, and (2) the stop closure is shorter. These two features cooperate in accordance with the aerodynamic voicing constraint.

When comparing the effects of vowel duration and speech rate on tongue rot position, we face a paradox. Both variables have a positive effect on tongue root position: Longer vowels and higher speech rates imply a more advanced root. However, speech rate has a negative effect on vowel duration (and segments duration in general), such that higher speech rates are correlated with shorter vowel durations (this of course holds for this data too, see Coretta 2018). If higher speech rates mean shorter vowels and shorter vowels mean less advanced root, we should also see less advancement with higher speech rates. But what we see is the opposite: With higher speech rates there is more root advancement. A way out would be for the tongue root to be already in a more advanced position at *vowel onset* when the speech rate is higher. Indeed, this is what a regression model on the vowel onset data suggests: Speech rate is positively correlated with tongue root position at vowel onset.

4.1   Individual differences

The results presented in Section 3 and discussed in Section 4 are general patterns that can be attributed to a population as inferred from the present sample. However, the data shows a great amount of individual differences.
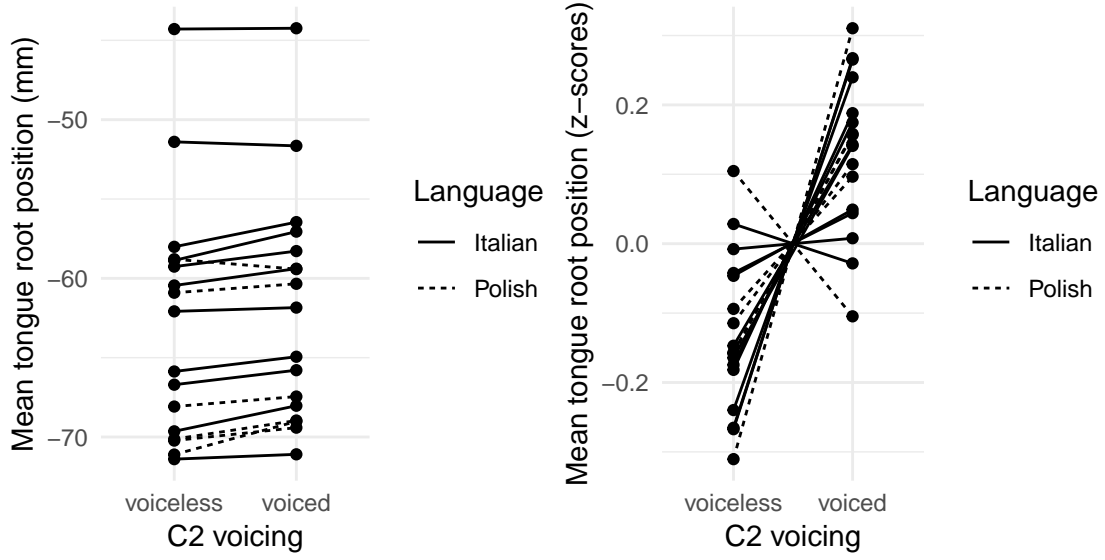
A   Target words

See Table 1.

Figure 5: Fig caption

Table 1: The list of Italian and Polish target words. An asterisk indicates a real word.

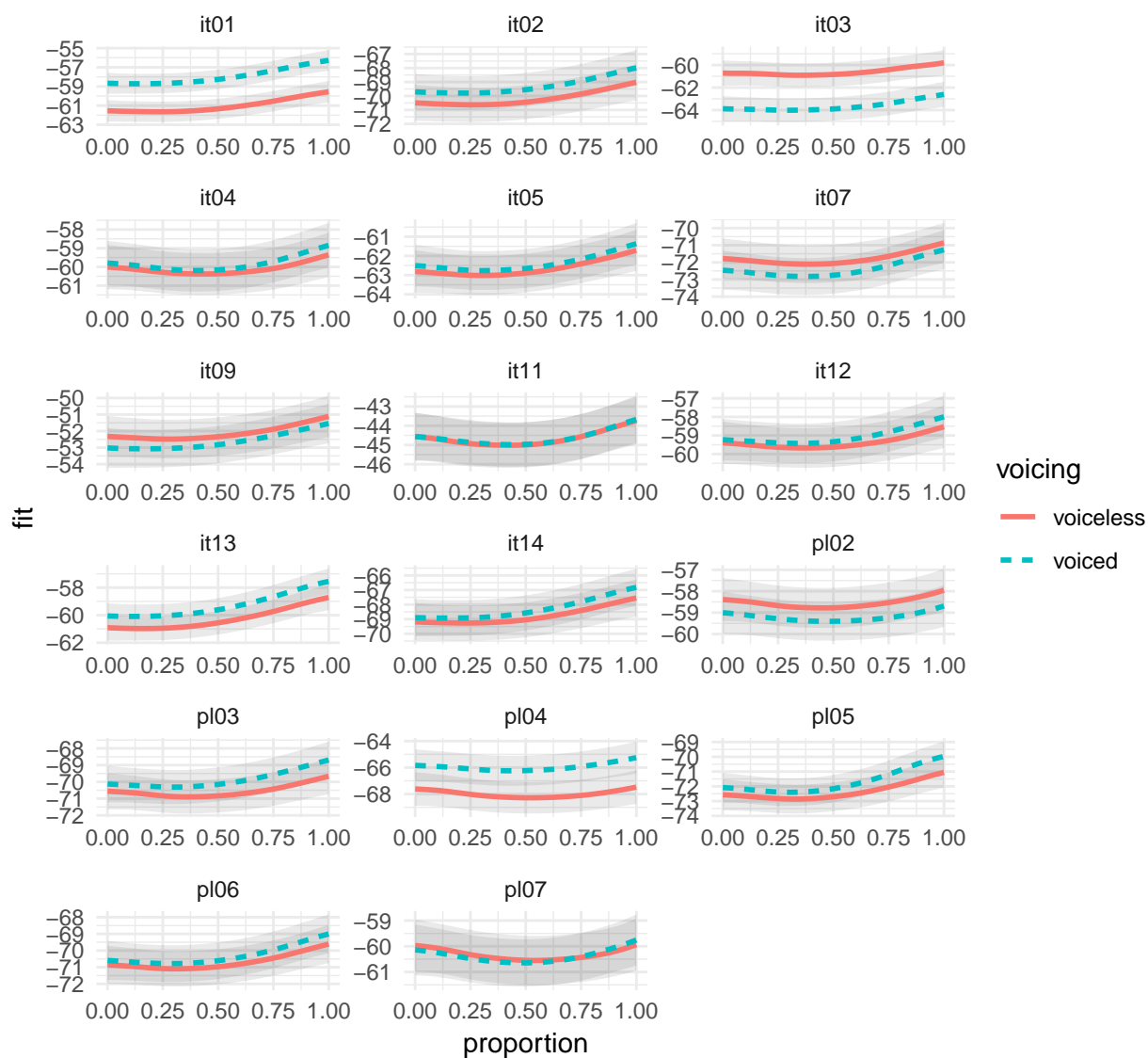| Italian				Polish		
pata	poto*	putu		pata	poto	putu
pada	podo	pudu		pada*	podo	pudu
paca*	poco*	pucu		paka*	poko	puku
paga*	pogo	pugu		paga	pogo	pugu

Figure 6: Fig caption

References

Ahn, Suzy. 2015. The role of the tongue root in phonation of American English stops. Paper presented at Ultrafest VII.

Ahn, Suzy. 2018. The role of tongue position in laryngeal contrasts: An ultrasound study of english and brazilian portuguese. *Journal of Phonetics* 71. 451–467.

Ahn, Suzy & Lisa Davidson. 2016. 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.

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

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. doi:10.18637/jss.v067.i01.

Beguš, Gašper. 2017. Effects of ejective stops on preceding vowel duration. *The Journal of the Acoustical Society of America* 142(4). 2168–2184. doi:10.1121/1.5007728.

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

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

Coretta, Stefano. 2018. An exploratory study of voicing-related differences in vowel duration as compensatory temporal adjustment in Italian and Polish. Submitted.

Durvasula, Karthik & Qian Luo. 2012. Voicing, aspiration, and vowel duration in Hindi. *Proceedings of Meetings on Acoustics* 18. 1–10. doi:10.1121/1.4895027.

Esposito, Anna. 2002. On vowel height and consonantal voicing effects: Data from Italian. *Phonetica* 59(4). 197–231. doi:10.1159/000068347.

Farnetani, Edda & Shiro Kori. 1986. Effects of syllable and word structure on segmental durations in spoken Italian. *Speech communication* 5(1). 17–34. doi:10.1016/0167-6393(86)90027-0.

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

Gussmann, Edmund. 2007. *The phonology of Polish.* Oxford University Press.

House, Arthur S. & Grant Fairbanks. 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. doi:10.1121/1.1906982.

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

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

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

Kirkham, Sam & Claire Nance. 2017. An acoustic-articulatory study of bilingual vowel production: Advanced tongue root vowels in Twi and tense/lax vowels in Ghanaian english. *Journal of Phonetics* 62. 65–81. doi:10.1016/j.wocn.2017.03.004.

Klatt, Dennis H. 1973. Interaction between two factors that influence vowel duration. *The Journal of the Acoustical Society of America* 54(4). 1102–1104. doi:10.1121/1.1914322.

Krämer, Martin. 2009. *The phonology of Italian.* Oxford: Oxford University Press.

Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen. 2017. `lmerTest` package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). doi:10.18637/jss.v082.i13.

Lampp, Claire & Heidi Reklis. 2004. Effects of coda voicing and aspiration on Hindi vowels. *The Journal of the Acoustical Society of America* 115(5). 2540–2540. doi:10.1121/1.4783577.

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

Luke, Steven G. 2017. Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods* 49(4). 1494–1502. doi:10.3758/s13428-016-0809-y.

Machač, Pavel & Radek Skarnitzl. 2009. *Principles of phonetic segmentation.* Epocha.

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

Malisz, Zofia & Katarzyna Klessa. 2008. A preliminary study of temporal adaptation in Polish VC groups. In *Proceedings of speech prosody*, 383–386.

Nowak, Pawel. 2006. *Vowel reduction in Polish*: University of California, Berkeley dissertation.

Ohala, John J. 2011. Accommodation to the aerodynamic voicing constraint and its phonological relevance. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 64–67.

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

Peterson, Gordon E. & Ilse Lehiste. 1960. Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America* 32(6). 693–703. doi:10.1121/1.1908183.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Renwick, Margaret & Robert D. Ladd. 2016. Phonetic distinctiveness vs. lexical contrastiveness in non-robust phonemic contrasts. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1). 1–29. doi:10.5334/labphon.17.

Riordan, Carol J. 1980. Larynx height during english stop consonants. *Journal of Phonetics* 8. 353–360.

Rothenberg, Martin. 1967. *The breath-stream dynamics of simple-released-plosive production*, vol. 6. Basel: Biblioteca Phonetica.

Sóskuthy, Márton. 2013. *Phonetic biases and systemic effects in the actuation of sound change*: University of Edinburgh dissertation.

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

Sprouse, Ronald L., Maria-Josep Solé & John J. Ohala. 2008. Oral cavity enlargement in retroflex stops. *Proceedings of the 8th International Seminar on Speech Production, Strasbourg* 425–428.

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

van Rij, Jacolien, Martijn Wieling, R. Harald Baayen & Hedderik van Rijn. 2017. itsadug: Interpreting time series and autocorrelated data using GAMMs. R package version 2.3.

Vazquez-Alvarez, Yolanda & Nigel Hewlett. 2007. The 'trough effect': an ultrasound study. *Phonetica* 64(2-3). 105–121. doi:10.1159/000107912.

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

Wieling, Martijn. 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* 70. 86–116. doi:10.1016/j.wocn.2018.03.002.

Wood, Simon. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)* 73(1). 3–36.

Wood, Simon. 2017. *Generalized additive models: An introduction with R*. Chapman and Hall/CRC 2nd edn.

Wood, Simon N. 2003. Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65(1). 95–114.

Yanagihara, Naoaki & Charlene Hyde. 1966. An aerodynamic study of the articulatory mechanism in the production of bilabial stop consonants. *Studia Phonologica* 4. 70–80.