AUTHOR QUERY FORM

AIP Publishing	**Journal: J. Acoust. Soc. Am.** **Article Number: JASA-04467**	**Please provide your responses and any corrections by annotating this PDF and uploading it according to the instructions provided in the proof notification email.**

Dear Author,

Below are the queries associated with your article; please answer all of these queries before sending the proof back to AIP. Please list figures that are to appear as color in print ____________________________ (a fee of $325 per figure will apply). If you do not wish to have color page charges, the figures will be color online only.

Article checklist: In order to ensure greater accuracy, please check the following and make all necessary corrections before returning your proof.

1. Is the title of your article accurate and spelled correctly?
2. Please check affiliations including spelling, completeness, and correct linking to authors.
3. Did you remember to include acknowledgment of funding, if required, and is it accurate?

Location in article	Query / Remark: click on the Q link to navigate to the appropriate spot in the proof. There, insert your comments as a PDF annotation.
AQ1	Payment of page charges for papers over 12 pages is mandatory. If you will not honor the page charges then you must reduce your paper to 12 pages.
AQ2	Please check that the author names are in the proper order and spelled correctly. Also, please ensure that each author's given and surnames have been correctly identified (given names are highlighted in red and surnames appear in blue).
AQ3	Please check edit to read "Rothenberg (1967, p. 99) hypothesised" in the sentence "Rothenberg (1967, p. 99) hypothesised, after an informal investigation, that a maximal ballistic expansion…."
AQ4	Please define "CVCV" in the sentence "Coretta (2019) argues, based on the acoustics of the same data reported here, that…."
AQ5	Please check inserted definition for "SE" in the sentence "A linear model, however, suggests a difference of 16 ms [standard error (SE) = 4.4] in both…." Please check throughout the text for insert of acronym "SE" following this definition..
AQ6	Please define "VC" in the sentence "Within this interval, the timing of VC boundary (the vowel…."
AQ7	Please specify the sections to which you refer with "in the following sections" in the sentence "The results from the articulatory data of this study, which will be discussed in…."
AQ8	Please check inserted company location for Articulate Instruments Ltd^{TM}, Movo, FocusRight, Hewlett-Packard, and Microsoft.
AQ9	Please provide the company location for TELEMED.
AQ10	Please check edit to read "post-aspiration" in the sentence "The burst and any eventual voiceless post-aspiration of C1 are not included…."
AQ11	Please check inserted definition for "SD" in the sentence "The mean pixel size as used by the automatic tracker was 0.47 mm…."
AQ12	Please define "ATR" in the sentence "Figure 1 in Kirkham and Nance (2017) suggests that the tongue root…."
AQ13	Please check edit from "the previous section" to "Sec. III" in the sentence "As mentioned in Sec. III, the correlation between tongue root position and vowel duration…."
AQ14	Please check edit from "The following section" to "Section IV C" in the sentence "Section IV C discusses this point."
AQ15	Please provide the day the website was last accessed in footnote 2.
AQ16	Please provide the day the website was last accessed in Ahn (2015), *R* Core Team (2018), Coretta (2018a,b), Crüwell *et al.* (2018), Keating (1984), Lehiste (1970b), Lindblom (1967), Maddieson and Gandour (1976), Malisz and Klessa (2008), Ohala (2011), Sprouse *et al.* (2008), Strycharczuk and Scobbie (2015).
AQ17	Please provide the page range in Kuznetsova *et al.* (2017).

AQ18	Please verify the single-page article in Lehiste (1970a); otherwise, please provide the full page range.
AQ19	Please provide the country in the location of the publisher in Machač and Skarnitzl (2009).
AQ20	Please provide the English translation for the article title in Magno Caldognetto *et al*. (1979).
AQ21	Please provide the publisher and location in van Rij *et al*. (2017).
AQ22	Please provide the location of the publisher in Wood (2017).
AQ23	Please define “CI” and “df” in Table II. Also, what is the significance of the last column in Tables II–V? Please define “<” and “*.”
AQ24	Please provide a digital object identifier (doi) for Bigi (2015), Vaux (1996), Yanagihara and Hyde (1966). For additional information on doi’s please select this link: http://www.doi.org/. If a doi is not available, no other information is needed from you.
AQ25	Please provide some descriptive text for the citations for Tables II–VI.

Thank you for your assistance.

J_ID: JASMAN DOI: 10.1121/10.0000556 Date: 14-January-20 Stage: Page: 1 Total Pages: 17

PROOF COPY [JASA-04467]

JASA ARTICLE

Longer vowel duration correlates with greater tongue root advancement at vowel offset: Acoustic and articulatory data from Italian and Polish

Stefano Coretta[a)]

Linguistics and English Language, University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom

ABSTRACT:

Voiced stops tend to be preceded by longer vowels and produced with a more advanced tongue root than voiceless stops. The duration of a vowel is affected by the voicing of the stop that follows, and in many languages vowels are longer when followed by voiced stops. Tongue root advancement is known to be an articulatory mechanism, which ensures the right pressure conditions for the maintenance of voicing during closure as dictated by the aerodynamic voicing constraint. In this paper, it is argued that vowel duration and tongue root advancement have a direct statistical relationship. Drawing from acoustic and ultrasound tongue imaging data from 17 speakers of Italian and Polish in total, it is proposed that the comparatively later closure onset of voiced stops is responsible for both greater root advancement and shorter closure durations of voiced stops. It is further shown that tongue root advancement is initiated during the vowel, and vowel duration and tongue root position at vowel offset are positively correlated so that longer vowel durations correspond to greater tongue root advancement. © *2020 Acoustical Society of America*.

https://doi.org/10.1121/10.0000556

(Received 30 May 2019; revised 7 December 2019; accepted 11 December 2019; published online xx xx xxxx)

[Editor: Anders Lofqvist] Pages: 1–15

I. INTRODUCTION

It is well known that voiced stops are almost universally associated with two phonetic correlates: advanced tongue root and increased duration of the preceding vowel. While a lot of work has been done on each of these aspects separately (tongue root, Ahn, 2018; Kent and Moll, 1969; Perkell, 1969; Rothenberg, 1967; Westbury, 1983, vowel duration, Chen, 1970; Farnetani and Kori, 1986; Fowler, 1992; House and Fairbanks, 1953; Klatt, 1973; Lisker, 1974; Peterson and Lehiste, 1960), less is known about their relationship. In this paper, I propose a link between the position of the tongue root at the onset of a post-vocalic stop and the duration of the vowel preceding that stop. In this exploratory study of the articulatory correlates of stop voicing, it was found that tongue root advancement—a mechanism known to facilitate voicing during stop closure—is initiated during the production of the vowel preceding the stop. This replicates previous work on tongue root position (Ahn, 2018; Kent and Moll, 1969; Perkell, 1969; Rothenberg, 1967; Westbury, 1983). Furthermore, the results of this study indicate that a comparatively later closure onset for voiced consonants, resulting in a longer preceding vowel duration, correlates with greater tongue root advancement at closure onset. Both the shorter closed phase of the voiced consonant and the more advanced tongue root, which expands the supraglottal cavity, have the potential to maintain voicing throughout C2 and preserve the voicing contrast.

A. Tongue root position and voicing

The initiation and maintenance of vocal fold vibration (i.e., voicing) during a stop closure requires a difference in air pressure between the cavities below and above the glottis. Specifically, the subglottal pressure needs to be higher than the supraglottal pressure. In other words, there must be a positive trans-glottal air pressure differential (Rothenberg, 1967; van den Berg, 1958). 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 supraglottal pressure quickly increases due to the incoming airstream from the lungs. Such a pressure increase can hinder the ability to sustain vocal fold vibration during closure to the point that voicing ceases.

A number of solutions can be used to counterbalance this pressure increase. For example, a cross-linguistically common difference between voiceless and voiced stops concerns their respective closure durations. The closure of English voiced stops is generally shorter than that of voiceless stops (Davis and Summers, 1989; de Jong, 1991; Lisker, 1957; Summers, 1987; Umeda, 1977). A shorter closure favours maintenance of vocal fold vibration by ensuring that the pressure buildup in the oral cavity does not equalise the subglottal and supraglottal pressures (at which point voicing would stop). Other articulatory solutions which can help sustaining voicing during closure rather

[a)]Electronic mail: stefano.coretta@manchester.ac.uk, ORCID: 0000-0001-9627-5532.

https://doi.org/10.1121/10.0000556

concern enlargement of the oral cavity. Among these solutions there are tongue root advancement (Ahn, 2018; Kent and Moll, 1969; Perkell, 1969; Rothenberg, 1967; Westbury, 1983), larynx lowering (Riordan, 1980), opening of the velopharyngeal port (Yanagihara and Hyde, 1966), and producing a retroflex occlusion (Sprouse *et al.*, 2008).

This study focuses on tongue root advancement as one of the articulatory adjustments implemented in voiced stops to expand the oral cavity and comply with the aerodynamic voicing constraint. In the context of articulatory adjustments, a distinction between passive and active gestures is generally drawn (see, for example, Rothenberg, 1967). A passive enlargement of the oral cavity is the product of the incoming airflow, the pressure of which expands the pliable soft tissues of the cavity walls. On the other hand, active expansion is achieved by muscular activity, which can in turn be purposive (produced with the goal of cavity expansion) or non-purposive. While Rothenberg (1967) recognises that the distinction between purposive and non-purposive active gestures can be at times blurry, it is nonetheless important to note that the qualification of a gesture as active does not automatically imply a speaker's intention to produce the obtained result.

Rothenberg (1967, p. 99) hypothesised, after an informal investigation, that a maximal ballistic expansion movement of the tongue root to increase the size of the lower pharynx AQ3 would take 70–90 ms. Based on these estimates, a passive expansion of the pharyngeal walls is thus not generally sufficient to maintain voicing during the closure of a lingual stop. Given that voiced stop closures are, on average, shorter than that (the mean duration is about 64 ms in Luce and Charles-Luce, 1985), it is expected that the movement could be initiated during the production of 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 the achievement of full closure and there is a forward movement even in some cases of voiceless stops, although the rate and magnitude of the advancement are consistently higher in voiced stops. Finally, tongue root adjustments seem to target more specifically lingual consonants, while tongue body lowering is more involved in labials (Perkell, 1969; Vazquez-Alvarez and Hewlett, 2007; Westbury, 1983).

However, the relation 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 (Ohala, 2011; Rothenberg, 1967; Westbury, 1983), and it has a certain degree of idiosyncrasy (Ahn, 2018). Moreover, Ahn (2018) finds that not all the speakers she surveyed did show tongue root advancement, and a few had rather the reverse pattern. Second, implementation of tongue root advancement can be decoupled from the actual presence of vocal fold vibration. In Westbury (1983), advancement of the tongue root is found in some productions of voiceless stops. This is counterintuitive, since tongue root advancement is generally considered to be a feature of voiced stops, which require voicing-related pressure adjustments. Moreover, Ahn (2015, 2018) looked at utterance-initial stops and found that the tongue root is more advanced in the phonologically voiced stops independent of whether they are implemented with vocal fold vibration or not.

To summarise, tongue root advancement is a common articulatory solution employed to counterbalance the increase in supraglottal pressure and maintaining voicing during the production of at least lingual voiced stops. While this gesture is not exclusive to voiced stops and it is sometimes implemented even in the absence of vocal fold vibration, tongue root advancement is strongly associated with (phonological) voicing.

B. Vowel duration and voicing

The results discussed here are part of a larger study that focuses on the effect of consonant voicing on preceding vowel durations. 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 ones (see, for example, Chen, 1970; Fowler, 1992; House and Fairbanks, 1953; Klatt, 1973; Lisker, 1974; Peterson and Lehiste, 1960, for English; Esposito, 2002; Farnetani and Kori, 1986, for Italian; Durvasula and Luo, 2012, for Hindi; Hussein, 1994, for Arabic). This so-called "voicing effect" has been reported in a variety of languages, including (but not limited to) English, German, Hindi, Russian, Arabic, Korean, Italian, and Polish (see Maddieson and Gandour, 1976; Beguš, 2017, for a more comprehensive list).

Italian and Polish offer an opportunity to study the articulatory aspects of the voicing effect, given that the former has been consistently reported as a voicing-effect language (Esposito, 2002; Farnetani and Kori, 1986; Magno Caldognetto *et al.*, 1979), while the voicing effect in the latter is more complex with some studies finding an effect (Coretta, 2019; Malisz and Klessa, 2008; Nowak, 2006; Slowiaczek and Dinnsen, 1985) and others not (Jassem and Richter, 1989; Keating, 1984). Moreover, the segmental phonologies of these languages facilitate the design of sufficiently comparable experimental material (see Coretta, 2019, for a more thorough discussion).[1]

Coretta (2019) argues, based on the acoustics of the same data reported here, that the first (stressed) vowel of disyllabic (~~CVCV~~) words is 11.5 ms longer in Italian and AQ4 7.55 ms longer in Polish when followed by a voiced stop. A linear model, however, suggests a difference of 16 ms [standard error (SE) = 4.4] in both languages, and language was AQ5 not a significant parameter. Moreover, the high degree of inter-speaker variation, backed up by statistical modelling, also indicates that these languages possibly behave similarly in regard to the voicing effect. More specifically, speakers of both Italian and Polish show a range of magnitudes of the voicing effect, and no particular language-specific patterns can be discerned. Independent of language, some speakers have a greater effect (of following consonant

JASA
https://doi.org/10.1121/10.0000556

voicing on vowel duration) and others a small or negligible effect (see Coretta, 2019, for details).

Finally, the temporal distance between two consecutive stop releases in CVCV words is not affected by the voicing of the second consonant. According to a Bayes factor analysis (Coretta, 2019), the duration of the release to release interval is not affected by the voicing of C2. Within this interval, the ~~timing of VC boundary (the vowel offset/onset of stop closure)~~ produces differences in the respective durations of vowel and closure, following a mechanism of temporal compensation (Lehiste, 1970a,b; Lindblom, 1967; Slis and Cohen, 1969a,b). A later closure onset results in a long vowel and a short closure, while an earlier closure onset corresponds to a short vowel and a long closure. Since the closure of voiceless stops is longer than that of voiced stops, it follows that vowels are shorter when followed by the former than when followed by the latter.

As more thoroughly discussed in Coretta (2019), the release to release interval *per se* does not have a special status, but rather it has been used as a proxy to acoustic temporal stability more in general. The aspects of the compensatory account proposed in Coretta ~~(2018b)~~ that are relevant to the current study are (1) C2 voicing does not affect the interval that includes V1 and C2, and (2) the placement of the VC boundary determines the duration of both V1 and C2 closure. Future studies are warranted to investigate the production and/or perceptual reasons behind both (1) and (2).

C. Rationale for the current study

Previous research has established that longer pre-consonantal vowel durations (Chen, 1970; House and Fairbanks, 1953; Peterson and Lehiste, 1960) and greater tongue root advancement (Kent and Moll, 1969; Perkell, 1969; Westbury, 1983) are associated with voicing in post-vocalic plosives. We know that voicing during plosive closure can be sustained by advancing the tongue root during the production of voiced plosives, and tongue root advancement probably begins before the closure onset (i.e., during the preceding vowel). We also know that vowels followed by voiced plosives tend to be longer than vowels followed by voiceless plosives. The acoustic analysis of the current dataset suggests an apparent compensatory relationship between the duration of the plosive closure and the duration of the preceding vowel (Coretta, 2019); the shorter the plosive closure, the longer the preceding vowel.

The results from the articulatory data of this study, which will be discussed in ~~the following sections,~~ offer new insights into the link between closure duration and vowel duration. We will see that the relative timing of the closure also covaries with the degree of tongue root advancement found at closure onset, resulting in a three-way relationship between stop consonant duration, vowel duration, and tongue root advancement. More specifically, the timing of the closure onset within the release to release interval determines not only the duration of the vowel and that of the closure (as discussed in Coretta, 2019), but also the degree of tongue root advancement at V1 offset/C2 onset. Finally, it will be argued that a later closure onset, as in the case of voiced stops, has the double advantage of producing both a short closure duration and greater tongue root advancement, features both known to comply with the aerodynamic voicing constraint.

II. METHODOLOGY

Following recent practices, which encourage scientific transparency and data attribution (Berez-Kroeker *et al.*, 2018; Crüwell *et al.*, 2018; Roettger, 2019), data (Coretta, 2018a) and the analysis code are available on the Open Science Framework. The analysis code can be found at ~~a temporary link for peer-review.[2] A public link will be generated in the case of acceptance.~~

A. Participants

Participants were recruited in Manchester (UK), and Verbania (Italy). Participants in this study included 11 native speakers of Italian (5 females, 6 males) and 6 native speakers of Polish (3 females, 3 males). Most speakers of Italian are originally from the North of Italy, while three are from Central Italy. The Polish speakers came from different parts of Poland (two from the west, three from the centre, and one from the east). This study has been approved by the School of Arts, Languages, and Culture Ethics committee of the University of Manchester (REF 2016-0099-76). The participants signed a written consent and received a monetary compensation of £ 10.

B. Equipment

Simultaneous recordings of audio and ultrasound tongue imaging were obtained in the Phonetics Laboratory at the University of Manchester (UK) or in a quiet room in Verbania (Italy). The possible influence of English on the speakers was reduced by talking to them and giving them instructions in their native language prior to and during the experiment. See Coretta (2019) for a thorough discussion. An Articulate Instruments Ltd™ system (Edinburgh, UK) was used for this study. The system is made of a TELEMED Echo Blaster 128 unit with a TELEMED C3.5/20/128Z-3 ultrasonic transducer (20 mm radius, 2–4 MHz; ■), and an Articulate Instruments Ltd™ *P*-stretch synchronisation unit. A Movo LV4-O2 Lavalier microphone (■) with a FocusRight Scarlett Solo pre-amplifier (■) were used for the acquisition of audio data. The ultrasonic probe was placed in contact with the flat area below the chin, aligned along the participant's mid-sagittal plane so that the mid-sagittal profile of the tongue could be imaged. 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 Hewlett-Packard ProBook 6750b laptop (■) with Microsoft Windows 7 (■). The synchronisation of the

https://doi.org/10.1121/10.0000556
JASA

ultrasonic and audio signals was performed by AAA after recording by means of a synchronisation signal produced by the ultrasound unit and amplified by the *P*-stretch unit. The ultrasonic settings were adjusted on a speaker basis to accommodate the scan area to the speaker's anatomy, and their ranges 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 22 050 Hz (16-bit).

C. 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.[3] The resulting words are nonce words, with a few exceptions, and they were presented in the languages' respective writing conventions (see Table I). A labial stop was chosen as the first consonant to reduce possible coarticulation with the following vowel. However, note that Westbury (1983) and Vazquez-Alvarez and Hewlett (2007) report tongue body lowering in the context of labial stops. Central/back vowels only were included in the target words for 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 Sec. II E. Since the focus of the study was to explore timing and articulatory differences in the closing gesture of voiceless and voiced stops, only lingual consonants have been included (the closure of labial stops cannot, of course, 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. Speakers were instructed to read the sentences without pauses and speak at a comfortable pace.

D. 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 six times consecutively

TABLE I. 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

(with the exception of IT02, who repeated the sentences five times only). Due to software constraints, the order of the sentences within a 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.

E. Data processing and statistical analysis

The audio data were 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 and 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 formants structure in the spectrogram (F2–F4, as per Machač and Skarnitzl, 2009). The burst and any eventual voiceless post-aspiration of C1 are not included in the duration of V1. See Fig. 1 for a segmentation example. AQ10 Vowel duration was calculated as the duration of the V1 onset to V1 offset interval. Speech rate was measured as the number of syllables in the sentence (eight in Italian and six 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 and Nance (2017). Note that, while the data were recorded without taking care that the tongue root was visible, the back part of the tongue just above the hyoid bone shadow (roughly corresponding to the uppermost part of the tongue root) was always imaged. Smoothing splines were automatically fitted to the visible tongue contours in AAA. The mean pixel size as used by the automatic tracker was 0.47 mm [standard deviation (SD) = 0.16] so that differ- AQ11 ences in tongue position smaller than that would not be captured. Manual correction was then applied to the automatically fitted tongue contours 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 and Scobbie, 2015); see Fig. 2. For each participant, the fan-line with the highest SD of displacement within the area corresponding to the speaker's tongue root was chosen as the tongue root displacement vector. The chosen fan-lines across all speakers range between fan-lines 25 and 34 (a higher number indicates a more posterior position), and these are always backer in the vocal tract than the fan-lines along which velar closure is articulated by the respective speaker. A Savitzky–Golay smoothing filter (second-order, frame length 75 ms) was

ID: aipepub3b2server Time: 02:39 I Path: D:/AIP/Support/XML_Signal_Tmp/AI-JAS#190902

JASA
https://doi.org/10.1121/10.0000556

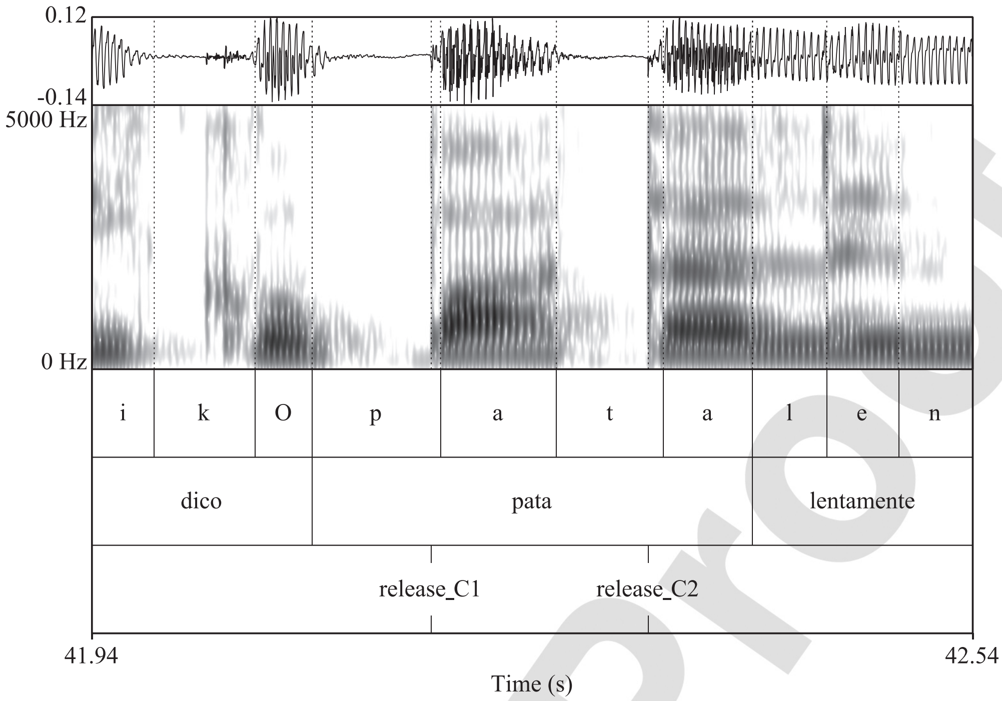

FIG. 1. Segmentation example of the words *pata* uttered by the Italian speaker IT09 (the times on the x axis refer to the times in the original audio file).

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 offset of V1/onset of the closure of C2) and along the duration of V1. The displacement values along the vowel duration were extracted at time points corresponding to ultrasonic video frames. Given the average frame rate is 55 frames per second, values are sampled about every 20 ms. The frame rate is adjusted by the system depending on other settings, so there is no standard frame rate. To

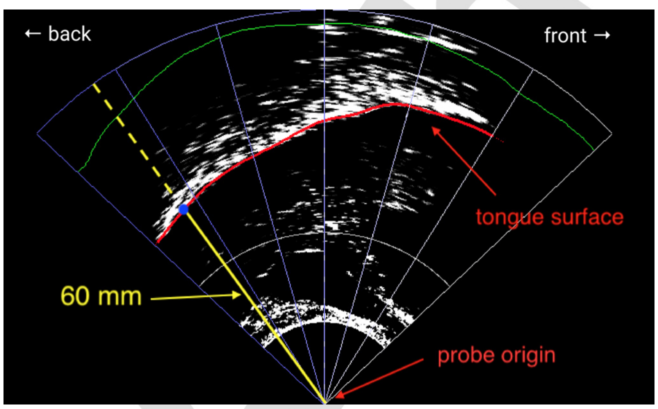

FIG. 2. (Color online) Schematics of the operationalisation of tongue root position based on Kirkham and Nance (2017). The tongue root surface corresponds to the lower edge of the white band in the image. The tongue tip is on the right side. The outline of the fan-like coordinate systems is shown. The yellow line starting from the probe origin is the selected fan-line from which the tongue root position is calculated (see the text for the method of fan-line selection). Tongue root position thus corresponds to the distance (in mm) between the probe origin and the intersecting point of the tongue surface with the selected fan-line (after z-scoring normalization, the sign is flipped so that greater values indicate greater tongue root advancement).

facilitate interpretation of the displacement values, the sign of these was flipped so that higher values indicate a more advanced tongue root (greater tongue root advancement) after Kirkham and Nance (2017).

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 (speech rate = number of syllables in the sentence/sentence duration). 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 (GAMMs) were fitted with mgcv v1.8–26 (Wood, 2011, 2017). The smooths used thin plate regression splines as a basis (Wood, 2003). The ordered factor difference smooths method described in Sóskuthy (2017) and Wieling (2018) was used to model the effect of factor terms in GAMMs. 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 in GAMMs 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

https://doi.org/10.1121/10.0000556
JASA

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. Speakers of both Italian and Polish were analysed together (models were fitted to a single dataset with data from both languages). Section IV C further discusses the idiosyncratic behaviour of the tongue root observed between speakers, which does not seem to pattern in any way with their native language. For these reasons, 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.

III. RESULTS

A. Tongue root position at C2 closure onset

Figure 3 shows raw data points and boxplots of the position of the tongue root at V1 offset/C2 closure onset when C2 is voiceless (left) and voiced (right). Since the position of the tongue root in mm depends on the speaker's anatomy and the probe location, scaled (z-scored) tongue root position is used in this plot (note, though, that the unscaled data are used in statistical modelling). As a trend, 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 (Table II): fixed effects for C2 voicing (voiceless, voiced), centred speech rate (as 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 it 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/.

B. Tongue root position during V1

The position of the tongue root during the articulation of V1 was assessed with GAMMs. A GAMM was fitted to tongue root position with the following terms (between parenthesis an explanation of how the term contributes to the model fit; Table III): C2 voicing as a parametric term (average root position difference between the voiceless and voiced contexts); a smooth term over centred speech rate (non-linear effect of speech rate on average tongue root position); a smooth term over V1 proportion (tongue root position along the duration of V1) with a by-C2 voicing difference smooth (difference in tongue root position along V1 in voiceless vs voiced contexts); a tensor product interaction over V1 proportion and centred speech rate (to model differences in tongue root position along V1 among different speech rates); a factor random smooth over V1 proportion by speaker (penalty order = 1, to model inter-speaker variation).

A chi-squared test on the ML scores of the full model and a model excluding the terms with C2 voicing (C2 voicing parametric term and by-C2 voicing difference smooth) indicates that C2 voicing significantly improves fit [$\chi(3) = 7.758$, $p = 0.001$]. Figure 4 shows the predicted tongue root position along the duration of V1 before voiceless

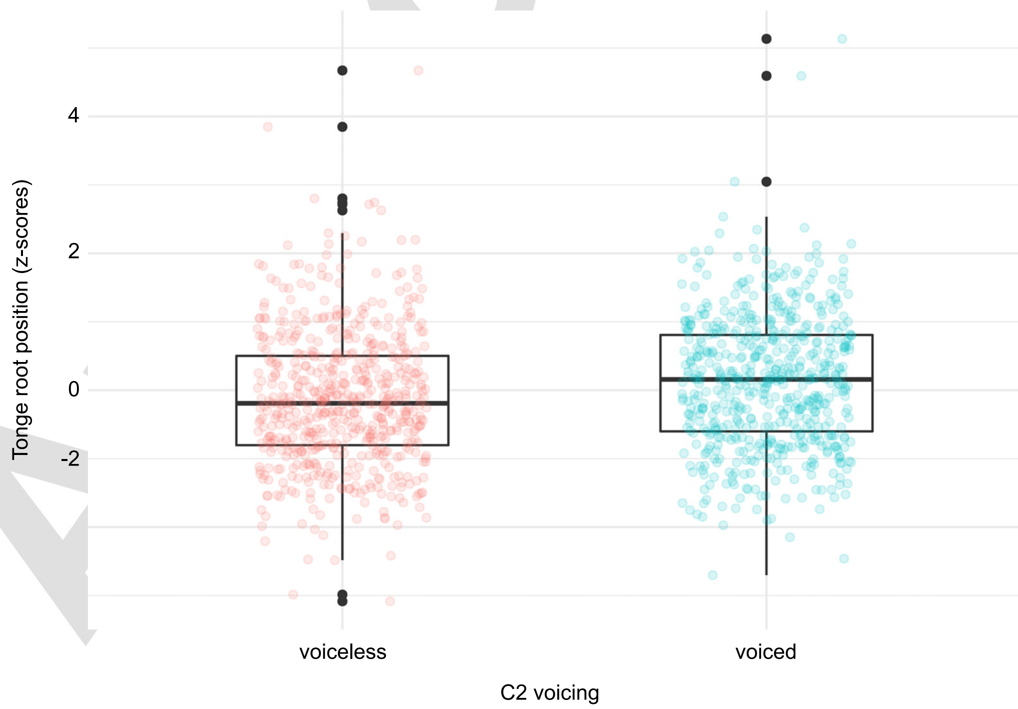

FIG. 3. (Color online) Raw data (z-scores) and boxplots of tongue root position in voiceless and voiced stops at closure onset. Higher values indicate advancement.

JASA
https://doi.org/10.1121/10.0000556

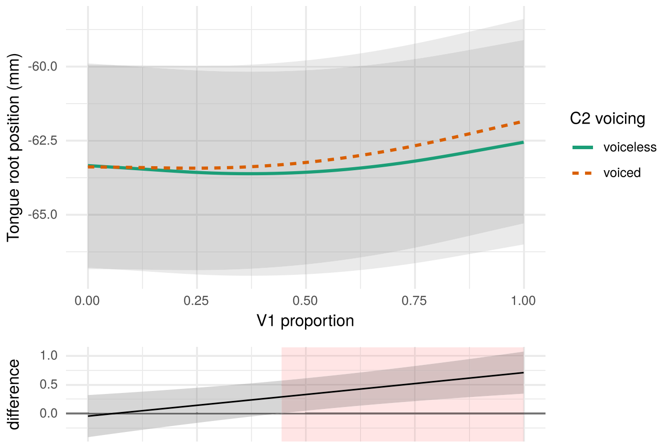

FIG. 4. (Color online) Predicted tongue root position (top) during vowels preceding voiceless (green solid line) and voiced stops (orange dashed line) with 95% confidence intervals, and difference smooth (bottom). Higher values of tongue root position indicate a more advanced root. The shaded red area in the difference smooth indicates where the two curves are different. Predictions from a GAMM (see Sec. III B).

(green solid line) and voiced stops (orange dashed line). Figure 4 indicates that the root advances during the production of the vowel relative to its position at V1 onset. This forward movement (increasing values of tongue root position in Fig. 4) is observed both in the context of a following voiced stop and that of a following voiceless stop. However, the magnitude of the movement is greater in the former. At V1 offset (= C2 closure onset), the graph suggests a difference in tongue root position of about 1 mm.

C. Correlation between tongue root position and V1 duration

A second linear mixed regression was fitted to tongue root position at V1-offset/C2-onset to assess the effect of V1 duration on root position (Table IV). 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 (a by-speaker random coefficient for V1 duration led to non-convergence, so it was not included in the final model). A separate model, which also included C2 voicing and its interaction with vowel duration, indicated that both terms are not significant so they were dropped in the model above. All other predictors and the V1 duration/vowel interaction are significant. V1 duration and tongue root position at V1 offset/C2 onset are positively correlated. The longer the vowel, the more advanced the tongue root is at V1 offset/C2 onset ($\hat{\beta} = 0.065$ mm, SE $= 0.007$). The effect is stronger with /a/ than with /o/ and /u/ (see Fig. 5).

D. Tongue root position during V1 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 (Table V): tongue root position as the outcome variable; smooth terms over V1 duration (non-linear effect of V1 duration on tongue root position) and V1 proportion

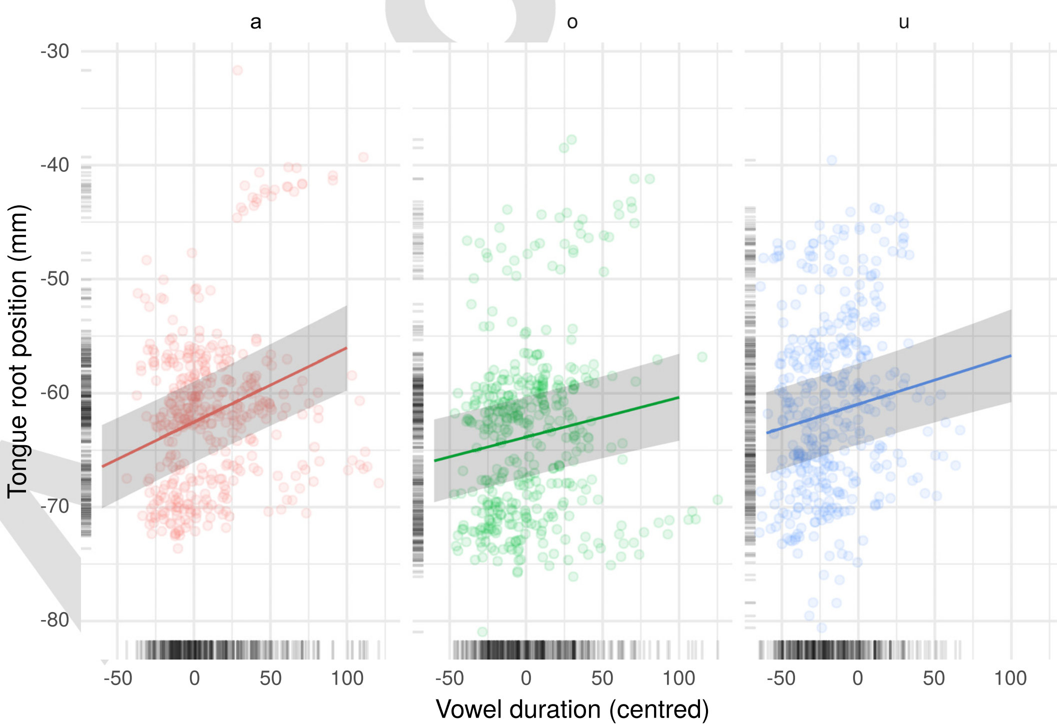

FIG. 5. (Color online) Raw data, regression lines, and 95% confidence intervals of the correlation between vowel duration and tongue root position for each vowel (/a/, /o/, and /u/). The regression line and confidence intervals are from a mixed-effects model (see Sec. III C).

https://doi.org/10.1121/10.0000556
JASA

(non-linear effect of V1 proportion); a tensor product interaction over V1 proportion and V1 duration (to model differences in tongue root position along V1 among different vowel durations); a factor random smooth over V1 proportion by speaker (penalty order = 1, to model inter-speaker variation). 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 [$\chi(3) = 12.609$, $p < 0.001$]. Figure 6 shows the estimated root position during vowels at four values of vowel duration. The general trend is that the total amount of the root advancement during the vowel is greater the longer the duration of the vowel (Fig. 6) and greater advancement at V1 offset/C2 onset is achieved the longer the vowel.

IV. DISCUSSION

A. Voicing, tongue root position, and vowel duration

The results of this study of voicing and vowel duration in Italian and Polish revealed a few patterns in the relation between consonant voicing, tongue root position, and vowel duration. Unsurprisingly, the position of the tongue root at vowel offset is 0.77 mm (SE = 0.35) more front when the following stop is voiced than when the following stop is voiceless in both surveyed languages (see Sec. IV B for a discussion about the magnitude of the difference and potential errors related to spline fitting). This finding aligns with the results of previous work on English (Ahn, 2018; Kent and Moll, 1969; Perkell, 1969; Westbury, 1983). When looking at the position of the tongue root during the vowel, it was found that the root starts advancing during the articulation of the vowel. Westbury (1983) found the same pattern in English. Moreover, similarly to the results in Westbury (1983), some tongue root advancement during the production of the vowel is found even when C2 is voiceless.

A possible reason for the presence of advancement in voiceless lingual stops is offered by arguments in relation to the absence of advancement in labials (voiced or voiceless).

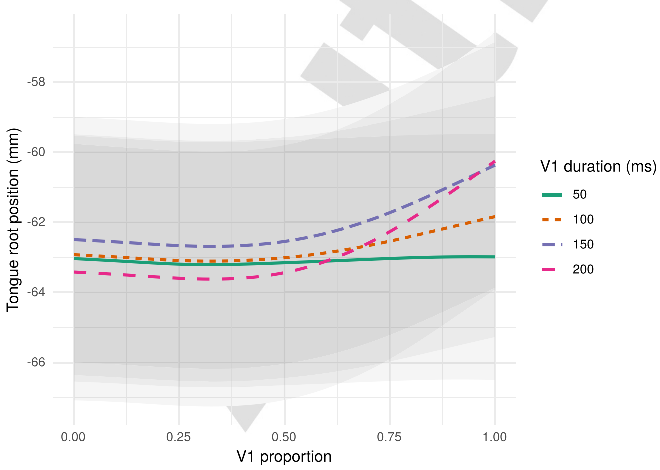

FIG. 6. (Color online) Predicted tongue root position during vowels at four exemplifying values of vowel duration with 95% confidence intervals. Predictions from a GAMM (see Sec. III D).

Westbury (1983) proposes that the articulation of the closure of lingual stops mechanically involves movements of the tongue root, so that, in order to keep a constant oral cavity volume, the root moves forward while the tongue body moves upward. On the other hand, the tongue can move freely in labial stops since their closure involves the lips. This idea is supported by the "trough effect" (Vazquez-Alvarez and Hewlett, 2007), i.e., VCV sequences involving a labial stop show tongue body lowering, and by the fact that voiced labials tend to resort to tongue body lowering rather than tongue root advancement as a mechanism for voicing maintenance (Ahn, 2018; Perkell, 1969; Westbury, 1983). The small degree of advancement in voiceless lingual stops could then as well be a mechanic consequence of the tongue moving upward for producing the stop closure.

The data discussed here also suggest that tongue root position at V1 offset/C2 onset is positively correlated with vowel duration such that longer vowels show a more advanced tongue root at V1 offset/C2 onset than shorter vowels. Said correlation exists independent of the voicing status of the consonant following the vowel. In other words, the position of the tongue at V1 offset/C2 onset is correlated with vowel duration both when the vowel is followed by a voiceless and a voiced stop. This finding is compatible with the finding that the tongue root advances during the production of vowels even when the following stop is voiceless (although it reaches less advancement than when the vowel is followed by a voiced stop).

The correlation between tongue root at V1 offset/C2 onset and vowel duration could indicate that the onset of the forward gesture of the root is timed not relative to the stop closure, but rather relative to an acoustic/articulatory event preceding the closure (what this event might be should be further investigated). Under this scenario, schematically represented in Fig. 7, the delay between the beginning of the tongue root advancement gesture of C2 and, for example, the release of C1 would not be affected by the voicing of C2. The timing of the tongue root advancement gesture would thus be independent of the time of stop closure onset and, hence, independent of the total duration of the vowel. Finally, the timing of full closure during the root advancement movement would sanction the degree of advancement found at closure onset (the later the closure onset relative to the onset of the advancement gesture, the greater advancement at closure onset).

The dynamic data of tongue root advancement during the articulation of the vowel (Sec. III B) indicate that vowels followed by voiced stops have greater tongue root advancement at V1 offset than vowels followed by voiceless stops, in accordance with the results from the static analysis at V1 offset. Moreover, a significant interaction was found between vowel duration and overall degree of advancement during the vowel (Sec. III D). Shorter vowels have overall less root advancement, while longer vowels have overall greater root advancement. This pattern could simply be a consequence of the fact that the tongue root has more time to advance the longer the duration of the vowel. I

JASA
https://doi.org/10.1121/10.0000556

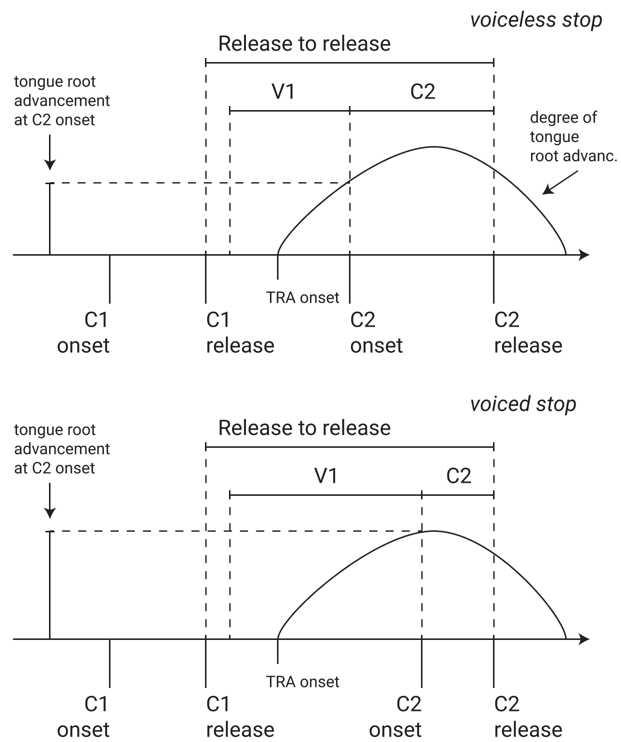

FIG. 7. Schematic representation of the proposed relationship between the acoustic vowel and closure durations and tongue root advancement. The x axis represents time, while the y axis denotes the degree of tongue root advancement (higher means more advancement). The articulatory onset of the tongue root advancement gesture (labelled as TRA onset) is at a stable distance from the preceding consonant release and vowel onset. The different timing of the VC boundary determines both the degree of tongue root advancement at vowel offset/consonant closure onset and the duration of the consonant closure.

have no explanation for why the degree of root advancement at *V1 onset* seemingly increases with increasing vowel duration except when the duration goes from 150 to 200 ms, and future work is necessary to shed light on this pattern.

The articulatory patterns observed in this paper contribute to the understanding of the acoustic patterns discussed in Sec. I. 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. Other things being equal, an increase in cavity volume increases the time required to reach trans-glottal pressure equalisation, which would cause cessation of voicing. This mechanism thus contributes to the maintenance of voicing during the stop closure.

The closure of voiced stops is achieved later (relative to the preceding consonant release) compared to the closure of voiceless stops. Moreover, the temporal distance between the releases of the two consecutive stops in CVCV words is not affected by the voicing category of the second stop (Coretta, 2019). Given the stability of the release to release interval duration, the delay in producing a full closure seen in the context of voiced stops has therefore a double advantage: (1) A greater degree of tongue root advancement is achieved at vowel offset/closure onset, and (2) the stop closure is shorter. Both of these articulatory features are compliant with the requirements dictated by the aerodynamic voicing constraint. A more advanced tongue root ensures that the trans-glottal pressure differential is sufficient for voicing to be sustained, and a shorter closure reduces the pressure buildup during the stop closure. To conclude, it is proposed that the combined action of a temporally stable release to release interval and the differential timing of the VC boundary in the context of voiceless vs voiced stops contribute to both the acoustic patterns of vowel and closure duration and the articulatory patterns of tongue root position.

B. Estimates of tongue root displacement

It is worth briefly discussing the estimated difference in tongue root position between voiceless and voiced stops and its significance. The estimated magnitude of such a difference is 0.77 mm (SE = 0.35). The 95% confidence interval for the difference is approximately within the range 0–1.5 mm. Rothenberg (1967) argues that the anterior wall of the lower pharynx (corresponding to the tongue root) can move by 5 mm along the antero-posterior axis. Figure 1 in Kirkham and Nance (2017) suggests that the tongue root of one of the Twi speakers recorded is about 4 mm more front in /e/ (a [+~~ATR~~] vowel) than in /ɛ/ (a [-ATR] vowel). Given that the articulatory space AQ12 within which the tongue can move is generally more constrained in stops than in vowels, and given that Kirkham and Nance (2017) find a difference of 4 mm in tongue root position in vowels, it makes sense to expect that differences in tongue root position as driven by consonantal factors should be of some magnitude smaller, like the ones found in this study. Moreover, the data presented here indicate that for every millisecond increase in vowel duration there is a 0.065 mm increase in tongue root advancement (see Sec. III C). If a maximal ballistic forward movement of the tongue root takes between 70 and 90 ms as suggested by the informal investigation by Rothenberg (1967), we can calculate the maximum displacement plausible to be between 4.55 and 5.85 mm (0.065 mm times 70–90 ms). These values are in agreement with the maximum root displacement of 5 mm estimated by Rothenberg. A note of caution is due since the actual error rate of the automatic tracker used for spline fitting is not known, and manual correction might have affected the splines (although a relatively small number of tokens had to be manually corrected).

The results of this study also shed some light on timing aspects of tongue root advancement. As mentioned in ~~Sec. III,~~ the correlation between tongue root position and vowel duration could be a consequence of the timing of the advancement gesture. In order to obtain such a correlation, AQ13 the articulatory onset of the advancement gesture (during the articulation of the vowel) should be at a stable distance from an earlier reference point (like the vowel onset or the preceding consonant offset) such that the timing of consonant closure will create the correlation seen in the data. Although ideally the timing of the onset of the advancing gesture relative to a preceding articulatory landmark should not be affected by the voicing of C2, the velocity of the

https://doi.org/10.1121/10.0000556
JASA

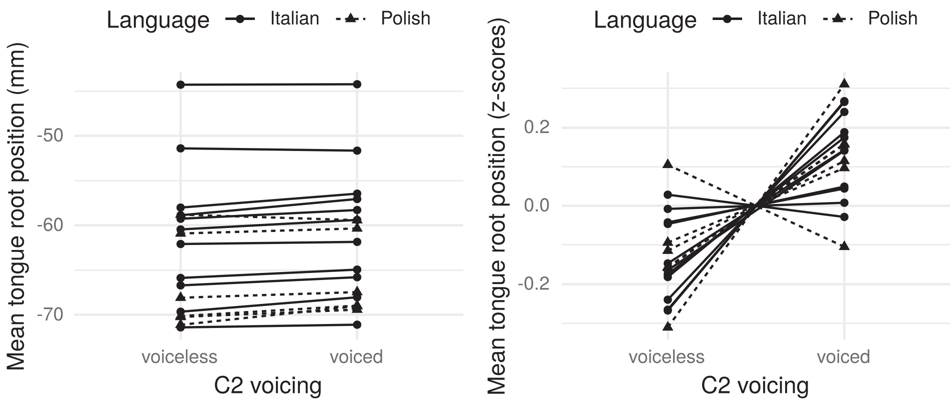

FIG. 8. Slope plots of mean tongue root position in voiceless and voiced stops at closure onset, by speaker. The plot on the left has raw position values in mm, while the plot on the right shows standardised values (*z*-scores) by speaker. See the text for details.

gesture itself could be different depending on the voicing of the following consonant. It is possible that the velocity will be greater in the context of voiced stops, especially if the advancing gesture in this context is executed with greater muscular force. Indeed, movement speed and amplitude tend to be correlated both in non-speech and speech movements (see Löfqvist and Gracco, 1994, Kuberski and Gafos, 2019, and references therein). Unfortunately, a preliminary screening of the current data was inconclusive as to whether timing and velocity are similar or different in the voiceless and voiced contexts, due to the difficulty in identifying the onset of the advancing gesture. Further data should be collected with the aim of testing the hypothesis that the timing of the gesture onset (for example, relative to the onset of the preceding vowel gestural onset) is the same in voiceless and voiced contexts, while the velocity of the gesture should differ.

Although the results of this study are in agreement with previous work, the correlation between tongue root position and vowel duration needs to be replicated by expanding the enquired contexts to other types of consonants and vowels, and with other languages. Investigating the relative phasing of tongue root and body gestures in lingual and labial consonants is also necessary to clarify the mechanisms that could underlie the gestural timing of stop closure and tongue root advancement. Moreover, while the paper so far has focussed on group-level trends, it should be noted that, as found in other studies on the tongue root, individual speakers show a somewhat high degree of variability. Section IV C discusses this point.

C. Individual differences

The results presented in Sec. III and discussed in Sec. IV are group-level patterns of the population sampled in the present study. However, the data are characterised by a certain degree of individual-level differences. Figure 8 shows two slope plots of mean tongue root position depending on C2 voicing for each speaker. In each plot, the two means of each speaker are linked by a line that shows the difference (or lack thereof) in means. Solid lines are Italian speakers, while dashed lines are Polish speakers. The *y* axis of the left plot is the raw mean position in mm, while that of the right plot is the standardised values (*z*-scores) of the mean position. An upward-slanted slope line indicates that the mean tongue root position in the voiced condition is higher, while a downward-slanted slope is interpreted as a decrease in mean root position. A flat slope suggests there is no difference in means between the voiceless and voiced condition.

These plots show that all three possibilities of slope direction are found in the data. The mean value of tongue root position of a voiced C2 relative to that of a voiceless stop is greater in some speakers, smaller in others, and similar in yet other speakers. Moreover, no discernible pattern can be found between speakers of Italian and Polish. Speakers of both languages show more or less the same range of variation. However, as we have seen in Sec. III, the estimated overall effect of C2 voicing is robust and it implies a more advanced tongue root in voiced stops. The right plot of Fig. 8 confirms this point visually. Two speakers show a declining slope (one is Italian and the other Polish), one speaker has a virtually flat slope, while all the others have an increasing slope at varying degrees. Note that the individual variation across speakers found in this data is qualitatively comparable to that in Ahn (2018).

The mean difference in tongue root position at the onset of voiceless vs voiced stops has been calculated for each speaker from the raw data. Figure 9 plots the speakers'

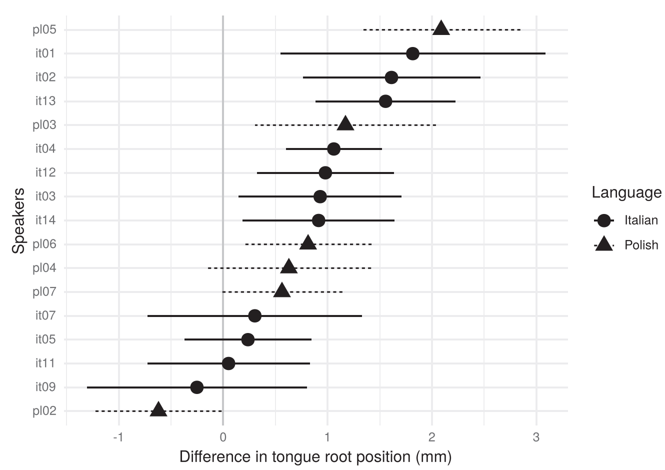

FIG. 9. By-speaker raw mean difference in tongue root position between voiceless and voiced stops at closure onset (in mm). The horizontal segments are the SEs of the mean differences.

JASA
https://doi.org/10.1121/10.0000556

mean differences with the respective SE bars. Overall, the means of the top 14 speakers indicate that these speakers have a more advanced tongue root in the context of voiced stops, while the bottom 3 speakers have means that indicate no difference or greater advancement in voiceless stops. As for the uncertainty of the estimates, the top ten speakers have a robust positive difference. The bottom seven speakers show either a weak negative difference (the tongue root is slightly more advanced in voiceless stops) or a weak positive difference with wide SEs. Finally, speakers of each language do not cluster together, reiterating the observation made above that language does not seem to be an informative parameter.

Finally, interesting individual patterns can also be seen in the trajectories of tongue root position. Figure 10 shows

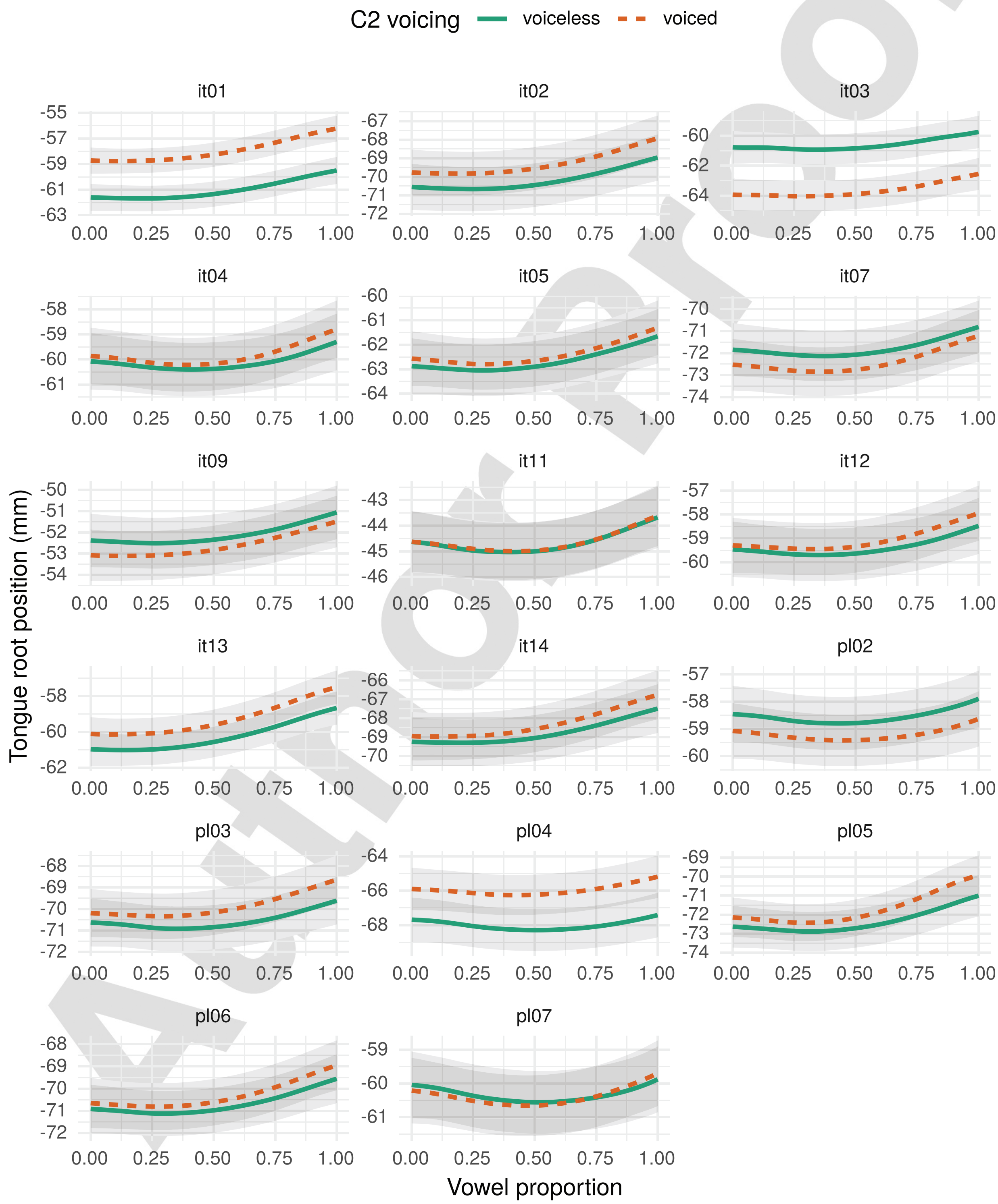

FIG. 10. Predicted tongue root position during vowels followed by voiceless and voiced stops for each speaker. Predicted from a GAMM (see the text). Note the different scales on the *y* axis.

https://doi.org/10.1121/10.0000556

these trajectories for all the speakers (note that the y axis of each plot is on a different scale, so magnitude comparisons should not be made visually). Speakers IT01, IT03, and PL04, in particular, have a somewhat categorical distinction in tongue root position during vowels followed by voiceless vs voiced stops. Such tongue root distinction is implemented across the total duration of the vowel, rather than toward the end (as suggested by the results from the aggregated data; see Sec. III B). The phonological literature reports cases in which the difference in tongue root position in vowels is enhanced, leading to phonological alternations or diachronic loss of the voicing distinction with maintenance of the tongue root distinction (see Vaux, 1996, and references therein). The ultrasound data from this study offer articulatory evidence for a possible precursor of said phonological patterns.[4]

D. A note on speech rate and vowel duration

When comparing the effects of vowel duration and speech rate on tongue root position, we are faced with a paradox. Both variables have a positive effect on tongue root position so that longer vowels and higher speech rates imply a more advanced tongue root at V1 offset/C1 onset. On the other hand, 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 holds for these data; see Coretta, 2019). If higher speech rates mean shorter vowels and shorter vowels imply a less advanced root, we should also find less advancement with higher speech rates.

However, the results of this study indicate the opposite, namely that higher speech rates are correlated with more root advancement. A linear regression model fitted to the position of the tongue root at *V1 onset* indicates that speech rate is positively correlated with tongue root position at vowel onset. The greater the speech rate, the greater the advancement of the tongue root at V1 onset. This means that the tongue root is already in a more advanced position at V1 onset when the speech rate is high so that if vowel duration is held constant, more advancement is expected at vowel offset with higher speech rates even when higher speech rate has a negative effect on vowel duration.

V. CONCLUSION

The maintenance of voicing during the closure of stops can be achieved through a variety of articulatory mechanisms. Among these, shorter closure durations (Davis and Summers, 1989) and cavity expansion by tongue root advancement (Westbury, 1983) are commonly observed solutions. Another robust correlate of consonant voicing is longer preceding vowel duration. This paper discussed articulatory data from an exploratory study of the effect of voicing on vowel duration first introduced in Coretta (2019). Similarly to what was previously found for English (for example, Ahn, 2018; Westbury, 1983), the tongue root at stop closure onset is more advanced in voiced than in voiceless stops in Italian and Polish. The average difference in tongue root position is 0.77 mm (SE = 0.35). By modelling the trajectory of the tongue root during the production of vowels preceding stops, it was found that the root starts advancing during the vowel, both preceding voiceless and voiced stops. The magnitude of the advancing gesture was, however, greater in the voiced context. Moreover, tongue root position at vowel offset and vowel duration was found to be positively correlated. Longer vowel durations correspond to greater tongue root advancement at vowel offset.

It was argued that the combined action of two factors contributes to the patterns observed: (1) the duration of the interval between two consecutive releases, and (2) the timing of the C2 closure onset within such interval. The release to release interval duration has been found not to be affected by the voicing of the second consonant. The later closure onset of voiced stops within the release to release interval (compared to voiceless stops) has the double advantage of producing a shorter closure duration and ensuring that enough tongue root advancement is reached by the time the stop closure is achieved. Both of these aspects comply with the aerodynamic voicing constraint (Ohala, 2011) by delaying trans-glottal pressure equalisation (which would prevent vocal fold vibration). Future studies will need to test whether these findings are replicable in Italian and Polish, and if they extend to other languages and contexts. In particular, further work on the relative differences in timing and velocity of the closing gesture and the root advancement gesture will be necessary to obtain a more in-depth understanding of the relation between consonant voicing, tongue root position, and vowel duration.

APPENDIX A: OUTPUT OF STATISTICAL MODELS

~~See Tables II, III, IV, and V.~~ AQ25

TABLE II. Summary of the linear mixed-effects model fitted to tongue root position at vowel offset (see Sec. III A).

AQ23

Predictor	Estimate	SE	CI low	CI up	df	t-value	p-value	
Intercept	−62.1396	1.8113	−65.6898	−58.5895	17.1188	−34.3058	0.0000	*
Voicing = voiced	0.7689	0.3473	0.0881	1.4497	19.3947	2.2137	0.0390	*
Speech rate (centred)	0.4114	0.2793	−0.1360	0.9588	1168.1100	1.4732	0.1410	
Vowel = /o/	−1.8742	0.4249	−2.7069	−1.0414	19.2874	−4.4112	0.0003	*
Vowel = /u/	0.0865	0.4270	−0.7503	0.9233	19.6974	0.2027	0.8415	

JASA
https://doi.org/10.1121/10.0000556

TABLE III. Summary of the GAM model fitted to tongue root position during V1 (see Sec. III B).

Predictor	Estimate	SE	EDF	Ref.DF	Statistic	*p*-value	<
Intercept	−63.3328	1.7562			−36.0623	0.0000	*
Voicing = voiced	0.3311	0.1432			2.3122	0.0208	*
s(Speech rate (centred))			7.5310	8.5159	4.4781	0.0000	*
s(Proportion)			3.6906	4.3631	10.4450	0.0000	*
s(Proportion): voiced			1.0121	1.0233	9.8423	0.0015	*
ti(Proportion, speech rate (c.))			2.1298	2.7632	2.9030	0.0429	*
s(Proportion, speaker)			62.2802	152.0000	57.3447	0.0000	*

TABLE IV. Summary of the linear mixed-effects model for testing the correlation between tongue root position and V1 duration (see Sec. III C).

Predictor	Estimate	SE	CI low	CI up	df	*t*-value	*p*-value	<
Intercept	−62.5793	1.7818	−66.0716	−59.0870	17.0874	−35.1212	0.0000	*
V1 duration (centred)	0.0651	0.0073	0.0507	0.0795	955.6436	8.8558	0.0000	*
Speech rate (centred)	1.2412	0.2903	0.6722	1.8102	1169.6885	4.2755	0.0000	*
Vowel = /o/	−1.3031	0.4597	−2.2040	−0.4021	18.3761	−2.8348	0.0108	*
Vowel = /u/	1.5863	0.5049	0.5967	2.5759	25.8255	3.1419	0.0042	*
V1 duration × /o/	−0.0303	0.0079	−0.0457	−0.0149	736.2314	−3.8504	0.0001	*
V1 duration × /u/	−0.0227	0.0090	−0.0403	−0.0052	751.2493	−2.5345	0.0115	*

TABLE V. Summary of the GAM model fitted to tongue root position during V1 as a function of V1 duration (see Sec. III D).

Predictor	Estimate	SE	EDF	Ref.DF	Statistic	*p*-value	<
Intercept	−63.0612	1.7406			−36.2285	0	*
s(V1 duration)			12.8981	15.3759	5.6011	0	*
s(Proportion)			3.9643	4.7060	18.0074	0	*
ti(Proportion,V1 duration)			2.8798	3.3636	8.9103	0	*
s(Proportion,speaker)			60.0873	152.0000	65.7194	0	*

APPENDIX B: ACOUSTIC DURATION MEASURES FOR EACH SPEAKER

~~See Table VI.~~

TABLE VI. By-speaker raw means in milliseconds of vowel duration, speech rate (number of syllables per second), and closure to closure interval. Each mean is followed by its SD.

Speaker	V1 duration	SD	Speech rate	SD	Closure to closure	SD
it01	126.34	15.09	5.06	0.11	230.26	13.43
it02	163.99	32.16	4.30	0.27	290.97	32.34
it03	123.27	32.38	4.82	0.45	268.38	36.64
it04	130.00	26.00	4.81	0.21	254.22	26.23
it05	92.93	18.32	5.65	0.30	199.32	21.16
it07	100.08	15.13	5.54	0.16	199.82	23.14
it09	76.77	18.56	6.43	0.24	172.39	18.60
it11	132.59	29.20	4.61	0.21	272.22	26.07
it12	92.94	14.13	5.98	0.27	188.22	16.36
it13	102.53	15.20	5.50	0.40	201.39	18.53
it14	102.52	21.22	6.78	0.49	207.27	18.74
pl02	78.03	17.01	5.98	0.36	224.23	27.41
pl03	77.32	13.88	5.88	0.35	192.82	19.35
pl04	72.22	16.69	6.51	0.40	178.91	14.29
pl05	98.02	19.28	5.68	0.21	225.92	22.69
pl06	72.56	12.17	4.92	0.40	181.97	21.72
pl07	77.34	18.10	5.27	0.23	213.28	21.04

[1]Simultaneous electroglottographic data (not discussed here) were also collected during the experiment. These data indicate that virtually all tokens of voiced stops were uttered with vocal fold vibration with just a few exceptions (four tokens were voiceless in the speaker PL02).

[2]~~https://osf.io/d245b/?view_only=c7ec58d937454de8b7ad9212c261776b (Last viewed ■).~~ AQ15

[3]Note that stressed vowels in open syllables in Italian are long (Renwick and 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 and Ladd, 2016; Gussmann, 2007).

[4]All the examples in Vaux (1996) are on vowels *following* voiceless vs voiced stops, rather than preceding, as in the current study. While beyond the scope of this paper, whether this is a systematic gap or not and how this relates to the present findings should be examined in future work.

Ahn, S. (**2015**). "The role of the tongue root in phonation of American English stops," paper presented at Ultrafest VII, available at http://www.ultrafest2015.hku.hk/docs/S_Ahn_ultrafest.pdf (Last viewed ■). AQ16

Ahn, S. (**2018**). "The role of tongue position in laryngeal contrasts: An ultrasound study of English and Brazilian Portuguese," J. Phon. **71**, 451–467.

Articulate Instruments Ltd[TM] (**2008**). *Ultrasound Stabilisation Headset Users Manual: Revision 1.4* (Articulate Instruments Ltd, Edinburgh, UK).

Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Software **67**(1), 1–48.

Beguš, G. (**2017**). "Effects of ejective stops on preceding vowel duration," J. Acoust. Soc. Am. **142**(4), 2168–2184.

Berez-Kroeker, A. L., Gawne, L., Kung, S. S., Kelly, B. F., Heston, T., Holton, G., Pulsifer, P., Beaver, D. I., Chelliah, S., and Dubinsky, S.

https://doi.org/10.1121/10.0000556
JASA

(**2018**). "Reproducible research in linguistics: A position statement on data citation and attribution in our field," Linguistics **56**(1), 1–18.

Bigi, B. (**2015**). "SPPAS—Multi-lingual approaches to the automatic annotation of speech," The Phonetician **111-112**, 54–69.

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

Coretta, S. (**2018a**). "An exploratory study of the voicing effect in Italian and Polish, [data]," Open Science Framework, https://osf.io/8zhku/ (Last viewed ■).

~~Coretta, S. (2018b). "Quantifying vocal fold activity: Two new methods for analysing electroglottographic data, [Research compendium]," Open Science Framework, https://osf.io/pj8rt/ (Last viewed ■).~~

Coretta, S. (**2019**). "An exploratory study of voicing-related differences in vowel duration as compensatory temporal adjustment in Italian and Polish," Glossa: J. Gen. Linguist. **4**(1), 1–25.

Crüwell, S., van Doorn, J., Etz, A., Makel, M., Moshontz, H., Niebaum, J., Orben, A., Parsons, S., and Schulte-Mecklenbeck, M. (**2018**). "8 easy steps to open science: An annotated reading list," PsyArXiv, doi■ (Last viewed ■).

Davis, S., and Summers, W. V. (**1989**). "Vowel length and closure duration in word-medial VC sequences," J. Phon. **17**, 339–353.

de Jong, K. (**1991**). "An articulatory study of consonant-induced vowel duration changes in English," Phonetica **48**(1), 1–17.

Durvasula, K., and Luo, Q. (**2012**). "Voicing, aspiration, and vowel duration in Hindi," Proc. Meet. Acoust. **18**, 1–10.

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

Farnetani, E., and Kori, S. (**1986**). "Effects of syllable and word structure on segmental durations in spoken Italian," Speech Commun. **5**(1), 17–34.

Fowler, C. A. (**1992**). "Vowel duration and closure duration in voiced and unvoiced stops: There are no contrast effects here," J. Phon. **20**(1), 143–165.

Gussmann, E. (**2007**). *The Phonology of Polish* (Oxford University Press, Oxford, UK).

House, A. S., and Fairbanks, G. (**1953**). "The influence of consonant environment upon the secondary acoustical characteristics of vowels," J. Acoust. Soc. Am. **25**(1), 105–113.

Hussein, L. (**1994**). "Voicing-dependent vowel duration in Standard Arabic and its acquisition by adult American students," Ph.D. thesis, The Ohio State University, Columbus, OH.

Jassem, W., and Richter, L. (**1989**). "Neutralization of voicing in Polish obstruents," J. Phon. **17**(4), 317–325.

Keating, P. A. (**1984**). "Universal phonetics and the organization of grammars," in *UCLA Working Papers in Phonetics*, Vol. 59, pp. 35–49, available at https://escholarship.org/uc/item/2497n8jq (Last viewed ■).

Kent, R. D., and Moll, K. L. (**1969**). "Vocal-tract characteristics of the stop cognates," J. Acoust. Soc. Am. **46**(6B), 1549–1555.

Kirkham, S., and Nance, C. (**2017**). "An acoustic-articulatory study of bilingual vowel production: Advanced tongue root vowels in Twi and tense/lax vowels in Ghanaian English," J. Phon. **62**, 65–81.

Klatt, D. H. (**1973**). "Interaction between two factors that influence vowel duration," J. Acoust. Soc. Am. **54**(4), 1102–1104.

Krämer, M. (**2009**). *The Phonology of Italian* (Oxford University Press, Oxford, UK).

Kuberski, S. R., and Gafos, A. I. (**2019**). "The speed-curvature power law in tongue movements of repetitive speech," PLoS One **14**(3), e0213851.

Kuznetsova, A., Bruun Brockhoff, P., and Haubo Bojesen Christensen, R. (**2017**). "lmerTest package: Tests in linear mixed effects models," J. Stat.
AQ17 Software **82**(13), ■.

Lehiste, I. (**1970a**). "Temporal organization of higher-level linguistic
AQ18 units," J. Acoust. Soc. Am. **48**(1A), 111.

Lehiste, I. (**1970b**). "Temporal organization of spoken language," in *OSU Working Papers in Linguistics* (The Ohio State University, ■), Vol. 4, pp. 96–114, available at https://linguistics.osu.edu/sites/linguistics.osu.edu/files/workingpapers/osu_wpl_04.pdf (Last viewed ■).

Lindblom, B. (**1967**). "Vowel duration and a model of lip mandible coordination," Speech Transmission Laboratory Quarterly Progress Status Report Vol. 4, pp. 1–29, available at http://www.speech.kth.se/prod/publications/files/qpsr/1967/1967_8_4_001-029.pdf (Last viewed ■).

Lisker, L. (**1957**). "Closure duration and the intervocalic voiced-voiceless distinction in English," Language **33**(1), 42–49.

Lisker, L. (**1974**). "On 'explaining' vowel duration variation," in *Proceedings of the Linguistic Society of America*, pp. 225–232.

Löfqvist, A., and Gracco, V. L. (**1994**). "Tongue body kinematics in velar stop production: Influences of consonant voicing and vowel context," Phonetica **51**(1–3), 52–67.

Luce, P. A., and Charles-Luce, J. (**1985**). "Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production," J. Acoust. Soc. Am. **78**(6), 1949–1957.

Luke, S. G. (**2017**). "Evaluating significance in linear mixed-effects models in *R*," Behav. Res. Meth. **49**(4), 1494–1502.

Machač, P., and Skarnitzl, R. (**2009**). *Principles of Phonetic Segmentation* (Epocha, Praha, ■). AQ19

Maddieson, I., and Gandour, J. (**1976**). "Vowel length before aspirated consonants," in *UCLA Working Papers in Phonetics*, Vol. 31, pp. 46–52, available at https://escholarship.org/uc/item/31f5j8m7 (Last viewed ■).

Magno Caldognetto, E., Ferrero, F., Vagges, K., and Bagno, M. (**1979**). "Indici acustici e indici percettivi nel riconoscimento dei suoni linguistici (con applicazione alle consonanti occlusive dell'italiano)" ("■"), Acta Phoniatr. Lat. **2**, 219–246. AQ20

Malisz, Z., and Klessa, K. (**2008**). "A preliminary study of temporal adaptation in Polish VC groups," in *Proceedings of Speech Prosody*, pp. 383–386, available at ~~http://www.isle.illinois.edu/sprosig/sp2008/papers/id182.pdf~~ (Last viewed ■).

Nowak, P. (**2006**). "Vowel reduction in Polish," Ph.D. thesis, University of California, Berkeley.

Ohala, J. J. (**2011**). "Accommodation to the aerodynamic voicing constraint and its phonological relevance," in *Proceedings of the 17th International Congress of Phonetic Sciences*, pp. 64–67, available at https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/SpecialSession/Session1/Ohala/Ohala.pdf (Last viewed ■).

Perkell, J. S. (**1969**). *Physiology of Speech Production: Results and Implication of Quantitative Cineradiographic Study* (MIT Press, Cambridge, MA).

Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**(6), 693–703.

R Core Team (**2018**). "*R*: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, available at https://www.R-project.org (Last viewed ■).

Renwick, M., and Ladd, R. D. (**2016**). "Phonetic distinctiveness vs. lexical contrastiveness in non-robust phonemic contrasts," Lab. Phonol. J. Assoc. Lab. Phonol. **7**(1), 1–29.

Riordan, C. J. (**1980**). "Larynx height during English stop consonants," J. Phon. **8**, 353–360.

Roettger, T. B. (**2019**). "Researcher degrees of freedom in phonetic sciences," Lab. Phonol. J. Assoc. Lab. Phonol. **10**(1), 1–27.

Rothenberg, M. (**1967**). *The Breath-Stream Dynamics of Simple-Released-Plosive Production* (Biblioteca Phonetica, Basel, Switzerland), Vol. 6.

Slis, I. H., and Cohen, A. (**1969a**). "On the complex regulating the voiced-voiceless distinction I," Lang. Speech **12**(2), 80–102.

Slis, I. H., and Cohen, A. (**1969b**). "On the complex regulating the voiced-voiceless distinction II," Lang. Speech **12**(3), 137–155.

Slowiaczek, L. M., and Dinnsen, D. A. (**1985**). "On the neutralizing status of Polish word-final devoicing," J. Phon. **13**(3), 325–341.

Sóskuthy, M. (**2017**). "Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction," arXiv:1703.05339.

Sprouse, R. L., Solé, M.-J., and Ohala, J. J. (**2008**). "Oral cavity enlargement in retroflex stops," in *Proceedings of the 8th International Seminar on Speech Production*, Strasbourg, pp. 425–428, available at https://issp2008.loria.fr/Proceedings/PDF/issp2008-101.pdf (Last viewed ■).

Strycharczuk, P., and Scobbie, J. M. (**2015**). "Velocity measures in ultrasound data. Gestural timing of post-vocalic /l/ in English," in *Proceedings of the 18th International Congress of Phonetic Sciences*, pp. 1–5, available at https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0309.pdf (Last viewed ■).

Summers, W. V. (**1987**). "Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses," J. Acoust. Soc. Am. **82**(3), 847–863.

Umeda, N. (**1977**). "Consonant duration in American English," J. Acoust. Soc. Am. **61**(3), 846–858.

van den Berg, J. (**1958**). "Myoelastic-aerodynamic theory of voice production," J. Speech Hear. Res. **1**(3), 227–244.

van Rij, J., Wieling, M., Baayen, R. H., and van Rijn, H. (**2017**). "itsadug: Interpreting time series and autocorrelated data using GAMMs," R package version 2.3 (■, ■). AQ21

Vaux, B. (**1996**). "The status of ATR in feature geometry," Linguist. Inquiry **27**(1), 175–182.

Vazquez-Alvarez, Y., and Hewlett, N. (**2007**). "The 'trough effect': An ultrasound study," Phonetica **64**, 105–121.

Westbury, J. R. (**1983**). "Enlargement of the supraglottal cavity and its relation to stop consonant voicing," J. Acoust. Soc. Am. **73**(4), 1322–1336.

Wieling, M. (**2018**). "Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English," J. Phon. **70**, 86–116.

Wood, S. (**2011**). "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models," J. R. Stat. Soc. Series B Stat. Methodol. **73**(1), 3–36.

Wood, S. (**2017**). *Generalized Additive Models: An Introduction with R*, 2nd ed. (Chapman and Hall/CRC, ■). AQ22

Wood, S. N. (**2003**). "Thin plate regression splines," J. R. Stat. Soc. Series B Stat. Methodol. **65**(1), 95–114.

Yanagihara, N., and Hyde, C. (**1966**). "An aerodynamic study of the articulatory mechanism in the production of bilabial stop consonants," Stud. Phonologica **4**, 70–80.