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

Stefano Coretta

1. Introduction

It is known that tongue root plays a role in maintaining voicing in voiced stops in English. Recent ultrasound tongue imaging work has confirmed that the tongue root is advanced in voiced consonants. Tongue root advancement has been shown to be present also when vocal fold vibration is not present [1]. An interesting question arising from this is weather tongue root advancement might be correlated to other factors like, the focus of this paper, vowel duration. Several studies showed that vowels followed by voiced stops are longer than vowels followed by voiceless stops. Different languages show different magnitudes of such durational differential and some languages do not show it at all. Given the connection between voicing and tongue root advancement, it is natural to ask whether tongue root advancement is also linked to vowel duration. If this is indeed the case, then one would expect tongue root advancement to play a role in language that have the voicing effect, but not in languages that do not show it.

To test this hypothesis, I conducted an acoustic and articulatory study that looked at vowel duration and tongue contours to assess the possible link between consonantal and vocalic tongue gestures.

Although several attempts have been put forward to explain the effect of voicing on vowel durations, to date no consensus has been reached. A recurrent theme focusses on the difference in gestural implementation that characterises voiced stops in comparison with voiceless stops.[1] One of the first accounts that attributed the voicing effect to a difference in production is that of Halle and Stevens [5], subsequently reiterated by Chomsky and Halle [4]. According to this account, voicing in vowels is produced with a state of the glottis that diverges from the configuration necessary to produce voiced consonants, due to the aerodynamics of the vocal tract. On the contrary, it is claimed that voiceless stops do not require any specific glottal configuration and thus the voicing perpetuated during the vowel naturally ceases at closure. The authors thus hypothesise

[1] However, see ... for a perceptually inclined account.

that, to allow the glottal state to change in voiced stops from sonorant voicing to obstruent voicing, the vowel is lengthen so that enough time is available for the change to happen without compromising the quality of voicing during the vowel. Although such account seemed promising at that time, later studies could not demonstrate that obstruent voicing is any different from sonorant voicing.

Tongue root advancement differences seems like a promising area of enquiry since its link to voicing has been already confirmed. On the same line of the laryngeal hypothesis, I put forward a similar account in which it is tongue root advancement rather than fold configuration that requires extra time during the vowel to be implemented. If tongue root advancement plays a role in determining the duration of preceding vowels through the extension mechanism described before, than one expects languages with the voicing effect to show a systematic advancement of the tongue root in voiced stops. On the contrary, tongue root advancement in languages without the voicing effect should be absent or less systematic.

2. Methodology

2.1. Participants

Native speakers of Italian and Polish have been recruited in Manchester and in Italy. Four speakers per language participated in the experiment. The participants received a compensation of £10 (or equivalent in Euro).

2.2. Equipment set-up

An Articulate Instruments set-up was used for this study. This is constituted by a TELEMED C3.5/20/128Z-3 ultrasonic transducer plugged into a TELEMED Echo Blaster 128 unit. A synchronisation unit (P-Stretch) was plugged into the Echo Blaster unit and used for automatic audio/ultrasound synchronisation. A Movo LV4-O2 Lavalier microphone plugged into a ... was used for audio recording. Articulate Assistant Advanced v2.17.1 running on a Windows laptop was used for the acquisition of audio and ultrasonic signal. A stabilisation headset produced by Articulate Instruments was used for probe stabilisation [].

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/, and C_2 = /t, d, k, g/ (e.g. /pata/, /pada/, /poto/, etc.). A bilabial stop /p/ was chosen as the first consonant to reduce influence on the following vowel (although cf. Vazquez-Alvarez and Hewlett [8]). Only coronal and velar stops were used as target consonants since labial consonants cannot be imaged with ultrasonography. All possible combinations were employed, yielding to a total of 12 target words. The words were embedded in medial position within a frame sentence. Prosodically similar sentences were used to ensure comparability between languages. The frame sentence was *Dico X lentamente* 'I say X slowly' for Italian, and *Mówię X teraz* 'I say X now' for Polish.

2.4. Procedure

The stimuli were randomised for each participant and repeated six times (the order was kept the same in each of the six repetitions due to software constraints). Before recording the stimuli, the occlusal plane of the participant was imaged using a bite plate [7], and the palate asking the participant to swallow water.

2.5. Data processing

Durational measurements were obtained from acoustics. A first rough annotation was achieved through force alignment using the SPASS force aligner. The criteria for the annotation of the acoustic data were the following:

- the target vowel onset was marked by visual inspection at the time of appearance of higher formants in the spectrogram following the burst of /p/
- the target vowel offset was marked by visual inspection at the time of disappearance of the higher formants in the spectrogram; this point also corresponded with the consonantal onset
- closure offset was marked by visual inspection at the time of formants onsets of the following vowel
- burst onset was determined automatically by a Praat implementation of the algorithm described in Ananthapadmanabha et al. [2]

Finally, vowel duration, consonant duration (including burst and VOT), and closure duration (from vowel offset to burst onset) were also automatically extracted with a Praat script.

2.6. Analysis

The contours were normalised by applying offsetting and rotation relative to the participant's occlusal plane [7]. Generalised additive mixed effects models [9] were used for the statistical analysis of tongue contour data in R [6]. Duration measurements were subject to linear mixed effects models using `lme4` in R [3].

3. Results

3.1. Vowel duration and voicing

A linear mixed effect regression model was fit on the Italian vowel durations with duration as the outcome variable, vowel quality (/a, o, u/), voicing and place of articulation of the following consonant, sentence duration as fixed effects, random intercepts by speaker and word, and random slopes for voicing by speaker. An interaction between voicing and vowel quality was also included, which turned out to be significant. P-values were obtained through likelihood ration tests comparing a model including voicing with a null model without voicing. According to the model, Italian vowels are 19.5 msec ($\pm$5.5) longer if followed by a voiced stop ($\chi^2(3) = 18.5$, p = 0.000337).

For Polish, the same model structure was used, excluding the voicing-vowel interaction (which was not significant). Surprisingly, the model reported a partially significant 8 msec ($\pm$3) effect of consonantal voicing on the preceding vowel ($\chi^2(1) = 5.4$, p = 0.02). The exploration of the random slopes by speaker showed that one speaker showed a particularly higher slope for voicing, meaning that the effect of voicing was stronger in his data. This observation will come handy while discussing about the results of the tongue contour data.

3.2. Tongue contours

The Italian tongue contour analysis showed that voiced stops are produced with advancement of the root of the tongue. Generalised additive mixed effect models were fitted for each speaker: the y-coordinates of the contours were the outcome variable; the x-coordinates the parametric term; a reference smooth term for X, with difference smooths for X by voicing, vowel quality, and place of articulation of following consonant; random smooths by word. A first-order autoregressive model was included, given the high auto-correlation residuals as obtained by visual inspection of the residuals. For two participants out of four, the root was significantly more front in voiced stops in both vocalic contexts (/a, o/). On the other hand, one participant had significant tongue root advancement only following /a/, while the fourth participant didn't show advancement at all. For Polish, three out of four speakers did not have tongue root advancement, while the fourth speaker showed significant advancement in both vocalic contexts.

4. Discussion

The presence of tongue root advancement in Italian but not in Polish initially supports the idea that vowel are longer if followed by voiced stops to allow time for the root to reach a position suitable for consonantal voicing. The reported absence of the voicing effect in Polish could then be ascribed to the absence of tongue root advancement in the voiced consonants of Polish. However, such overstatement is complicated by two facts that emerged from the data analysed in this study. First, tongue root advancement was found in one of the Polish speakers and it was absent from one of the Italian speakers. Second, even though the effect was not as big as for Italian, the Polish duration data nonetheless showed a difference of 8 msec. However, as mentioned above, one Polish speaker had a particularly high slope value for voicing, and incidentally this is also the same and only Polish speaker who showed root advancement. Moreover, a difference of 8 milliseconds could be considered to be quite small in any case, although statistically detectable.[2]

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

[2] T. V. Ananthapadmanabha, A. P. Prathosh, and A. G. Ramakrishnan. Detection of the closure-burst transitions of stops and affricates in continuous speech using the plosion index. *The Journal of the Acoustical Society of America*, 135(1):460–471, 2014.

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

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

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

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

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

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

[9] Simon Wood. *Generalized additive models: an introduction with R.* CRC press, 2006. ISBN 1584884746.

[2] A model fitted on the data excluding the outlier speaker leads in fact to a smaller estimate of about 6 msec.