Tongue root position in VC sequences with regard to the phonetic realization of obstruent voicing: A preliminary study on Hungarian

Tekla Etelka Gráczi[1,5], Tamás Gábor Csapó[2,5], Andrea Deme[3,5], Kornélia Juhász[4,5], Alexandra Markó[3,5]

[1]*Hungarian Research Institute for Linguistics*
[2]*Budapest University of Technology and Economics, Department of Telecommunications and Media Informatics*
[3]*Eötvös Loránd University, Department of Applied Linguistics and Phonetics*
[4]*Eötvös Loránd University, Doctoral School of Linguistics*
[5]*MTA-ELTE "Lendület" Lingual Articulation Research Group*

graczi.tekla.etelka@nytud.hu, csapot@tmit.bme.hu, deme.andrea@btk.elte.hu, juhasz.kornelia8@gmail.com, marko.alexandra@btk.elte.hu

Abstract

In this paper we studied the tongue root position in VC-sequences with regard to the phonological voicing of the consonant and its phonetic realization. /iz/ and /is/ sequences were recorded embedded into carrier sentences in 12 speakers' pronunciation. Simultaneous audio, ultrasound and EGG-recordings were carried out. The speakers were categorized based on the voiceless part ratio (EGG-signal analysis) and CoG (speech signal analysis) of the /z/-realizations. We compared the tongue root position between the two sequences as a function of the speakers' categories. The results showed that there is a large interspeaker difference in /z/-realizations and that this is in a strong interrelation with the tongue root position patterns. Possible explanations are discussed for the speakers whose pronunciations are exceptions from the expected difference between the voiced and voiceless contexts.

Keywords: tongue root advancement, voicing counterparts, voiced fricatives, individual differences, ultrasound tongue imaging

1. Introduction

Vocal fold vibration and obstruent production are conflicting articulatory goals (e.g. Stevens 1997). In the case of the fricatives, this means that the simultaneous articulation of the turbulent noise and vocal fold vibration have contradictory pressure requirements. In order to produce intense turbulent noise high intraoral pressure needs to be achieved. The glottal vibration, however, need low intraoral pressure in order to maintain the transglottal pressure drop. Various articulatory gestures may be used to maintain the vocal fold vibration: e. g. lowering the larynx, enlarging the oral cavity, lowering the tongue, tongue root advancement. Tongue root advancement was found in the voiced compared to the voiceless counterpart segments in various languages (e.g. stops: Westbury 1983; Ahn 2018; Coretta 2020, fricatives: Narayanan et al. 1995).

The ratio of vocal fold vibration in voiced fricatives varies not only across speakers (Fuchs et al. 2007, for Hungarian see e.g. Gráczi 2012) but also across languages (Shih 1999). The tongue-palate contact measures showed diverse pictures with regard to voicing, which may be caused by variability in voicing: the tongue-palate contact was strongly related to the amount of voicing present in voiced fricatives in devoicing speakers (Fuchs et al. 2007). Combining the results on the tongue-palate contact varying with the amount of voicing present in the voiced fricative's realization (Fuchs et al. 2007), we may ask if the tongue root position also shows variability with the amount of voicing in voiced fricatives.

The present experiment introduces a pilot study on the timing pattern of tongue root movement regarding phonological and phonetic voicing. Our hypothesis was that in the pronunciation of speakers whose /z/ realizations tend to maintain voicing throughout (most of) its duration, larger differences can be found in the tongue root position both in the fricative and in the preceding vowel between /z/ and /s/ than in the pronunciation of /z/-devoicing speakers.

2. Methods

Nonsense words /izi/ and /isi/ were initial words of sentences, where the first /i/ bore word stress and pitch-accent. The 10 sentences (5 with both target words) appeared in randomized order among distractors. 12 native female speakers of Hungarian (20-27 ys) were recorded using the AAA ultrasound system of Articulate Instruments Ltd. at 81.67 fps. The ultrasound transducer was fixed below the speakers' chin by the ultrasound stabilization headset designed for speech recordings (Articulate Instruments Ltd). The speech signal was recorded with a Beyerdynamic TG H56c tan omnidirectional condenser microphone. The ultrasound data and the audio signals were synchronized using the tools provided by Articulate Instruments Ltd. Electroglottograms were captured by D200 device (Laryngograph Ltd.). The speech signal was recorded by the EGG as well, ultrasound and EGG recordings were synchronized through the two speech signals.

Segmentation was carried out on the speech signal by forced alignment (Mihajlik et al. 2010), then manual correction was carried out on the first /i/- and on the fricative-realizations in Praat (Boersma & Weenink 2019).

The tongue contours were traced manually at each frame along the time course of the /iz/ and /is/ sequences in AAA (Articulate Instruments Ltd.). The voiceless part ratio (ratio of duration without vocal fold vibration to the entire consonant duration) was automatically labeled and manually corrected in the EGG-signal in Praat (Boersma & Weenink 2019).

The CoG of the fricatives were measured at five equidistance measurement points. The total speech recording was transformed to spectrogram with a window length of 0.005 s, time step of 0.002 s, frequency step of 20 Hz, in the range of 0 to 21000 Hz using Gaussian window, and the spectral slice was taken at the time points to be measured with fast transformation. Generalized additive mixed models (GAMM) (R: R Core Team 2019, mgcv: Wood 2017, itsadug: van Rij et al. 2017, rticulate: Coretta 2019) were run. The tongue contours were analyzed separately for each speaker (Coretta 2019). Two models were built: one for the vowel tongue contours in the V and one for the tongue contours in the VC sequence, where the V was set to the 0-0.5 time interval, the C to the 0.5-1 interval. A basic model and a model with C contrast (with contrast treatment) were compared with $\chi 2$ test. Autocorrelation correction (first order autoregressive model) was carried out. The final, best fitting models were built up as the followings. The Y-position of the tongue contour points was the dependent variable. The models included the reference smooth of the X-position of tongue contours and of the normalized time course in the V or VC sequence, and also their interaction (tensor). All the smooths and tensors were included with and without nesting for the difference smooths of the C.

The first analyses of the results made evident that the speakers need to be grouped into four speaker types based on the results of the voiceless part ratio and CoG-analyses. The detailed results on the voiceless part ratio and the acoustic measures are discussed in Gráczi et al. (2020).

The realization of a fricative was considered partially devoiced if the voiceless part ratio was above 25%. In the case of sp01-sp04, sp11 and sp12, no or one partially devoiced /z/-realizations appeared. In the case of sp05, sp07, sp08 2 or 3 realizations of /z/ phonemes were partially devoiced. The remaining three speakers, sp06, sp09, sp10 had 4 or 5 partially devoiced /z/-realizations.

The CoG values were compared between /z/- and /s/-realizations in order to select the speakers who tend to pronounce approximant-like /z/-realizations. This was found to be important based on the first observations of the tongue contour results of the speakers who tended to maintain vocal fold vibration across their /z/-realizations. In the case of sp01 and sp03 the CoG of /z/-realizations raised close to the typical values for /s/s, while sp02, sp04, sp11 and sp12 had very low CoG values along the time course of the analyzed voiced fricatives. While the former speakers can be grouped together as speakers pronouncing frication and voicing simultaneously, the latter four speakers tended to pronounce approximant-like realizations.

The results for the voiceless part ratio and the CoG combined gives four speaker groups: speakers who tend to pronounce /z/ as voiced approximants (Group1: sp02, sp04, sp11, sp12), voiced fricatives (Group2: sp01, sp03), sometimes voiced, sometimes partially devoiced speech sound (Group3: sp05, sp07, sp08), devoiced speech sounds (Group4: sp06, sp09, sp10).

3. Results

Results regarding the tongue contour differences by the consonant contrast of the GAMMs are shown in **Table 1** and **2**. The results for X-position are significant if the tongue shape averaged across the duration is significantly different between the contexts. This means that whenever the difference appeared, it was altogether large enough to be significant already in the vowel duration in 6 speakers. The tensor is significant if the difference between the tongue shapes of the two contexts change across the time. **Table 1** shows the results for the model on the vowel duration. In 6 out of the 12 speakers the tongue shape was different even averaged across the time between the two contexts within /i/-realizations. In 8 speakers the tongue shape formed significantly different along the time course of /i/-realizations between the two contexts. Three speakers that did not have a significant difference averaged across the time domain had large enough change to be considered significant.

When considering the VC time course, the time-averaged tongue shape differed significantly in 9 speakers between the two contexts, and the change in the tongue shape was significant in 10 of them. The results also show speakers who do not tend to have significant difference between the two contexts.

Table 1: *Results for the smooth term for X-position and the tensor of the normalized time (t) and X-position by consonant contrast from the GAMM within the V duration. Grey cells indicate significant differences.*

Speaker	N	s(X, by =C.ord)		te(X, t, by =C.ord)	
		F	p	F	p
sp01	912	1.863	0.173	6.643	< 0.001
sp02	1239	0.450	0.502	3.840	0.004
sp03	1003	2.544	0.009	1.256	0.239
sp04	1487	2.676	0.102	1.578	0.143
sp05	1003	5.126	0.024	7.199	< 0.001
sp06	831	3.441	0.064	5.234	< 0.001
sp07	1086	4.147	0.042	2.155	0.037
sp08	1080	23.867	< 0.001	9.419	0.002
sp09	1184	8.025	0.005	3.946	0.002
sp10	2372	3.455	0.063	2.166	0.017
sp11	922	6.315	0.002	0.341	0.559
sp12	1473	2.885	0.090	0.116	0.734

Table 2: *Results for the smooth term for X-position and the tensor of the normalized time (t) and X-position by consonant contrast from the GAMM within the VC duration. Grey cells indicate significant differences.*

Speaker	N	s(X, by =C,ord)		te(X, t, by =C.ord)	
		F	p	F	p
sp01	2213	8.368	< 0.001	5.773	< 0.001
sp02	2652	54.880	< 0.001	2.889	0.003
sp03	2581	11.474	< 0.001	1.101	0.326
sp04	3194	0.008	0.930	2.795	< 0.001
sp05	2957	0.088	0.768	15.603	< 0.001
sp06	2131	5.634	< 0.001	0.032	< 0.001
sp07	2730	10.096	< 0.001	15.972	< 0.001
sp08	2667	1.849	0.174	4.896	< 0.001
sp09	2519	78.42	< 0.001	14.85	< 0.001
sp10	4349	7.365	0.006	3.384	< 0.001
sp11	2483	5.875	< 0.001	3.072	0.019
sp12	1473	3.328	0.002	0.806	0.458

The estimated differences and its 95% confidence intervals of the entire tongue contours are shown in **Figure 1.** The tongue root is positioned to the right of the differential lines. The five illustrated time points are the 50% and 75% of the vowel duration, the VC-boundary and the 25% and 50% time point of the consonant. The expected difference was found if the tongue root was advanced in the voiced context, i.e. the differential line decreases or stays in the negative domain, and is indicated by a red dashed line which stands for the significant differences.

In Group1, where the speakers tended to pronounce approximant-like /z/-realizations, only one speaker had the expected difference at the tongue root, while two speakers did not: sp04 had no difference and sp11 had a significant but opposite difference (her tongue root was more advanced in the /s/-realizations). The further two speakers' tongue root was found to be significantly advanced already during the preceding vowel's duration in the voiced phoneme context.

In Group2, where most of the /z/-realizations were voiced fricatives, both speakers had the expected difference, i.e. advanced tongue root in the voiced fricatives, however, in one speaker's (sp01) pronunciation this difference became significant only during the consonant's duration. The other speaker (sp03) had the expected tongue root pattern already during the preceding /i/-realization.

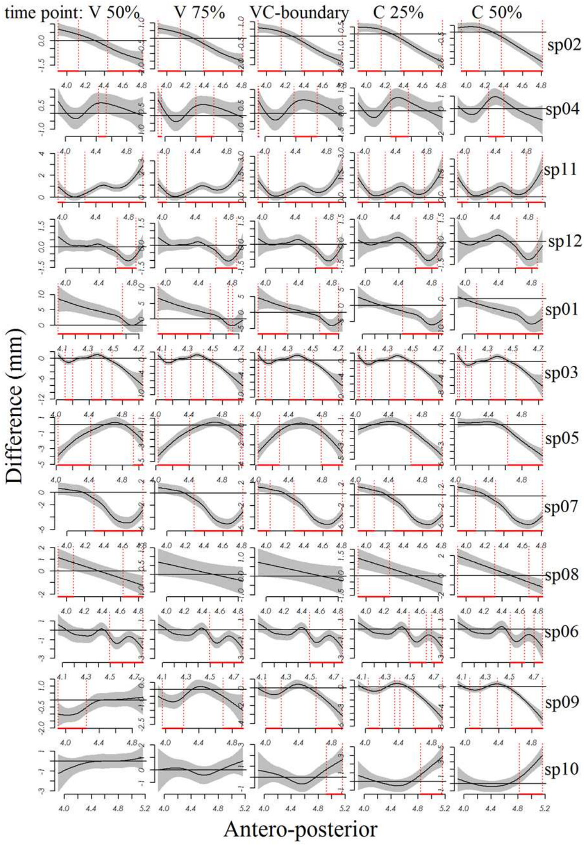

Figure 1: *The estimated tongue contour difference of /is/ - /iz/ and its 95% confidence interval.*

Table 3: *Appearance of the expected tongue root position in the speaker groups. (✓ = the tongue root is significantly advanced in the /z/-context compared to the /s/- context, ✗ = the tongue root is not advanced in the /z/- context compared to the /s/- context)*

Group	Speaker	Expected difference appeared during	
		the preceding /i/	the consonant
(1)	sp02	✓	✓
	sp04	✗	✗
	sp11	✗	✗
	sp12	✓	✓
(2)	sp01	✗	✓
	sp03	✓	✓
(3)	sp05	✗	✓
	sp07	✓	✓
	sp08	✓	✗
(4)	sp06	✓	✓
	sp09	✓	✓
	sp10	✗	✗

In Group3, where the speakers devoiced 2-3 /z/-realizations out of the 5, the tongue root position difference was very variable in one speaker's (sp08) pronunciation: the expected difference was detectable during vowel production but disappeared towards the consonant. In one speaker's (sp05) pronunciation the expected pattern did appear already in the vowel, but it reached the significance level only during the consonants' duration. The third speaker (sp07) showed advanced tongue root in the voiced phoneme's context already during the preceding vowel's durations.

In Group4, where most of the /z/-realizations were pronounced (partially) devoiced, only one speaker (sp10) of the three had different pattern from the expected one: her tongue root was neither advanced during in the /z/-context nor in the /z/-realizations, while the two further speakers had advanced tongue root already during the preceding vowel in the voiced phoneme context.

4. Discussion

The results showed that the tongue root was significantly advanced in the /z/-realizations in 8 out of the 12 speakers compared to the /s/-realizations, and this difference was significant already in during the preceding vowel in 6 out of them. Although each group consist only of a few, 2-4 speakers, their typical /z/-realizations are important in order to discuss what reasons or consequences these results have.

In the case of the speakers who tend to produce approximant-like /z/-realizations (Group1), the exceptions might be explained in two ways. In their case the tongue tip might not be raised as high, close to the alveolar region that the frication is not produced/reached. Either the non-advanced tongue root makes the approximant-like realizations necessary in order to avoid devoicing, or the approximant-like realization makes unnecessary to advance their tongue root during the voiced phoneme's realizations. The speakers who produce voiced /z/-realizations but with frication (Group2) the tongue root was found to be advanced. This strengthens the need to separate these two groups, as their articulatory strategies can be hypothesized to be different.

Most speakers with 2-5 partially devoiced /z/-realizations (Group3, Group4) advanced their tongue roots in the voiced context and most of them reached a significant difference already during the preceding vowel between the two contexts. The realizations though (often) became (partially) devoiced. There might be two reasons again that might contribute to these results. One is that the results appear generalized for their speech sounds averaged that shades the possible within-speaker differences. The other is that advancing the tongue root is only one articulatory gesture in order to maintain voicing. Further articulatory maneuvers have not been studied in this specific analysis.

The tongue root position was found to appear as expected but disappear towards the consonant in one speaker (sp08, Group3). The possible explanation might be that her realizations were diverse. She produced roughly the half of her /z/-phonemes partially devoiced. The voiceless part ratio ranged between 0% and appr. 60% in her voiced phoneme realizations, which means a large within-speaker variability. Possibly the tongue root started advancing during the preceding vowel, but not always sufficiently to maintain voicing.

Another reason has to be taken into consideration behind the results discussed in the above two paragraphs. Westbury (1983) and Coretta (2020) found tongue root advancement also in some cases for voiceless stops. Coretta (2020) took over the reasoning by Westbury (1983) that the raise of the anterior part of tongue (towards the closure in stops) may mechanically forward the tongue root. As the present study did not include statistical analysis that would provide results on this phenomenon for voiceless fricatives, we cannot include this as being proven for this material, however, the question has to be addressed in our future work.

We can see in **Figure 1,** that the significant difference at the midpoint of the consonants ranges roughly between 1-5 mm.

Coretta (2020) raised the question analyzing stops whether this magnitude is considerable. He argues that the anterior wall of the pharynx is able to move 5 mm (Rothenberg 1967), and 4 mm difference was found in Twi between vowels with and without advanced tongue root (Kirkham & Nance 2017). The tongue root displacement in vowels compared preceding (truly) voiced and voiceless (unaspirated) stops ranged between 0-1.5 mms, and were proven to be significant in Coretta's results (2020). Altogether we can hypothesize that the present 1-5 mm range for the significant tongue root difference might be considered as relevant.

Finally, we found altogether four speakers (2 from Group1, 1-1 from Group3 and Group4) who did not have tongue root advancement in the realizations of /z/ compared to /s/ phonemes. We have already discussed a few reasons behind these results for the speakers of Group2 above in this section. However, we have to add that if we consider that the results are based on comparison of the two speech sound sets, which may shed away tongue root advancement in both sets that results in a non-significant p-value, and some speakers had also diverse results in Coretta's study (2020) analyzing vowels preceding stops.

5. Conclusions

The present study introduced a pilot study on tongue root movement timing comparing VC-sequences with voiced and voiceless fricatives. One vowel (/i/) and one fricative pair (/z/ and /s/) were studied.

The results showed that the tongue root movement and the realizations of the voiced fricatives have to be studied combined, and also further articulatory maneuvers have to be taken into consideration. The future work has to explore an expanded vowel and consonant set in order to see the coarticulatory effects of both the vowels' and the consonants' articulatory gestures. The further work has to be expanded also by numerifying the tongue root displacement in itself, and by the categorical analysis of the speaker groups being transformed into scaled speech sound features.

6. Acknowledgements

The authors acknowledge the kind help of Valéria Krepsz in sound labeling. The research was funded by the Bolyai János Research Scholarship of the Hungarian Academy of Sciences, the UNKP-20-5 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund, and the Thematic Excellence Program of the Ministry for Innovation and Technology.

7. References

Ahn, S. (2018). The role of tongue position in laryngeal contrasts: An ultrasound study of English and Brazilian Portuguese. J Phonetics, 71, 451-467. doi:10.1016/j.602wocn.2018.10.003.

Boersma, P. & Weenink, D. (2019). Praat: doing phonetics by computer. http://www.fon.hum.uva.nl/praat/download_win.html. (Downloaded: 2019. october 1.)

Coretta, S. (2019). rticulate: Ultrasound Tongue Imaging in R. R package version 1.5.0.9000. https://github.com/stefanocoretta/rticulate

Coretta, S. (2020). Longer vowel duration correlates with greater tongue root advancement at vowel offset: Acoustic and articulatory data from Italian and Polish. JASA 147(1), 245-259. https://doi.org/10.1121/10.0000556

Fuchs, S. Brunner, J. & Busler, A. (2007). Temporal and spatial aspects concerning the realizations of the voicing contrast in German alveolar and postalveolar fricatives. Advances in Speech-Language Pathology, 9(1), 90-100.

Gráczi, T. E. (2012). Zörejhangok akusztikai fonetikai vizsgálata a zöngésségi oppozíció függvényében. [Acoustic characteristics of obstruents with regard to voicing opposition]. PhDthesis. Budapest: Eötvös Loránd University. http://doktori.btk.elte.hu/lingv/graczitekla/diss.pdf

Gráczi, T. E., Csapó, T. G., Bartók, M., Deme, A., & Markó, A. (2020). The realization of voicing opposition in alveolar fricatives in Hungarian. Articulation and acoustics. Beszédtudomány–Speech Science 1. 21-55.

Kirkham, S., & 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.

Markó, A., Bartók, M., Csapó, T. G., Deme, A. & Gráczi, T. E. (2019). The effect of focal accent on vowels in Hungarian: articulatory and acoustic data. In: Calhoun, S., Escudero, P., Proc. 19th ICPhS. (pp. 2715-2719). http://intro2psycholing.net/ICPhS/papers/ICPhS_2764.pdf

Mihajlik, P., Tüske, Z., Tarján, B., Németh, B. & Fegyó, T. (2010). Improved recognition of spontaneous Hungarian speech: Morphological and acoustic modeling techniques for a less resourced task. IEEE Transactions on Audio, Speech and Language Processing, 18(6), 1588-1600.

Narayanan, S. S., Alwan, A. A. & Haker, K. (1995). An articulatory study of fricative consonants using magnetic resonance imaging. JASA, 98, 1325-1347.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

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

Shih, C., Möbius, B. & Narasimhan, B. (1999). Contextual effects on consonant voicing profiles: A cross-linguistic study. Proc. 14th ICPhS. 989-992.

Stevens, K. N. 1997. Articulatory–acoustic–auditory relationships. In Hardcastle, W. J., & Laver, J. (szerk.): 1997. 462 - 506. The Handbook of Phonetic Sciences. Blackwell: Oxford.

van Rij, J., Wieling, M., Baaye, R. & van Rijn, H. (2017). itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs. R package version 2.3.

Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. JASA, 73(4), 1322-1336.

Wood, S. N. (2017). Generalized Additive Models: An Introduction with R (2nd edition). Chapman and Hall/CRC.

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