# Dynamic articulatory and acoustic features of Hungarian sibilants as a function of phonological voicing

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

[1]Hungarian Research Centre for Linguistics, [2]MTA–ELTE "Lendület" Lingual Articulation Research Group, [3]Eötvös Loránd University, [4]Dept. of Telecomm. and Media Informatics, Budapest Univ. of Technology and Economics,

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

## ABSTRACT

It is a phonetic cliché that obstruents' phonological voicing shows various phonetic patterns due to aerodynamic constraints. We analysed articulatory and acoustic features of Hungarian intervocalic sibilants /z, ʒ, s, ʃ/.) Midsagittal Ultrasound Tongue Imaging was used to track tongue root, tongue tip and tongue blade along the total duration of the consonants (83 images per seconds). From the acoustic signal, centre of gravity (CoG), spectral skewness, voicing profile (all at 11 equidistant temporal points), and voiceless part ratio were extracted. Tongue root was advanced in /z, ʒ/ compared to /s, ʃ/; however, in some cases, its backward movement was also found in both voiced and voiceless sibilants, which might appear as compensation for the lowering of the tongue front to maintain the intraoral pressure for intense turbulent frication: CoG did not lower, neither skewness increased considerably at the end of the consonants despite the restart of phonation in the voiced sibilants.

**Keywords**: tongue root movement, tongue front movement, spectral features, phonation, sibilants.

## 1. INTRODUCTION

Phonation and obstruent production are contradictory targets: phonation requires low supraglottal air pressure, while the oral seal leads to the increase of intraoral pressure [1, 2]. This contradiction is more enhanced in sibilants, as their aerodynamic target is a high intensity turbulent frication which can be reached by increasing the air pressure behind the oral obstacle.

One of the possible articulatory manoeuvres to maintain phonation is to produce voiced obstruents with an advanced tongue root (TR) compared to their unvoiced counterparts [3, 4, 5]. While the tongue root advancement of voiced consonants is a well described phenomenon, its dynamic features and its interrelation with the spectral features are less documented so far. In a study of Polish and Italian stops, TR moved forward during the preceding vowel for both voiced and voiceless consonants, and it was more advanced for voiced stops in both languages already at the start of the closure, than for their voiceless counterparts [6]. In a comparison of Hungarian /z/ and /s/, [7] found that the TR was more advanced in half of the speakers (6 out of 12) at the start of the consonant duration, and in 75% of them (8 speakers) during the entire duration of the consonant.

The contradictory aims of phonation and obstruent production might also lead to partial devoicing of the voiced obstruents. The presence of TR advancement was not found to vary among speakers as a function of phonetic voicing, i.e., this phenomenon was recognised both in the group of speakers who tended to maintain voicing, and in the group of speakers who devoiced their /z/-realisations [7]. The degree of TR's advancement, however, was found to vary with the ratio of phonation to the consonant duration [8].

The present study raises the question how the TR position changes during the production of intervocalic voiced and voiceless sibilants, and how this change interacts with the contradictory targets of turbulent frication and phonation. We hypothesized that (i) the TR is advanced in voiced targets along the sibilants' total duration to maintain phonation, (ii) the TR position is different in alveolar and postalveolar places of articulation due to the volume difference in the obstacle-vocal fold distance, (iii) tongue front and TR position changes along the sibilants' time interval in a manner to maintain intense turbulent frication.

## 2. METHODS

Four Hungarian sibilants /s, z, ʃ, ʒ/ were analysed in nonsense words of the structure /l/$V_1$_$V_1$/l/, where the analysed sibilants appeared at the onset of the second syllable, and the two neighbouring vowels were identical, either /ɒ, aː, ɛ/ or /i/. Two of these sequences *(lesel, leszel)* are meaningful words in Hungarian. Each item was read out five times by three native female speakers of Hungarian (aged between 35 and 47 ys) in a randomized order. Altogether 4 (C) * 4 (V) * 5 (repetitions) = 80 items per speaker were analysed. The items were recorded together with further nonsense words that are out of the focus of the current study. The recording was carried out by AAA speech ultrasound system (Articulate Instruments LtD.), which recorded midsagittal tongue contours at a frame rate of 83 image/s. The speech signal was recorded by Beyerdynamic TG H56c omnidirectional condenser microphone.

The segment boundaries were automatically labelled [9] and corrected manually in Praat [10]. The voiceless part ratio (VPR) of the sibilants was retrieved via Praat's voice report function (fraction of locally unvoiced frames; $f_0$ range was set to 75–300 Hz, for further settings, the standard values were kept). Centre of Gravity (CoG) and spectral skewness were measured automatically unfiltered along the time course of the sibilants at 11 equidistant time points with 0.01 s window. No filtering was used in order to include the effect of the presence or absence of voicing in the results of the spectral measures. The presence or absence of phonation at the same time points was retrieved by the Pitch (ac) function of Praat with the same settings as VPR for all three speakers. Voicing profile is defined by the ratio of phonation present at a time point in the specific consonant [see 11]. While VPR shows how much percent of each consonant realisation's duration is devoiced in percentages, the voicing profile shows how many of all the consonant realisations had maintained/restarted the phonation at each analysed time point. For example, a VPR of 0% means one token with phonation all along its total duration, while a voicing profile of 0% at measurement point 5 means that none of that phoneme's realisations exhibited phonation at this specific point of their duration.

Each tongue contour along the sibilants' duration was drawn manually in the AAA software by the first two authors. Approximately 12.7 tongue contours per sibilant realisations, altogether 3031 tongue contours were manually tracked and analysed.

The statistical analyses were carried out in R [12]. The possible difference of VPR among the consonants and vowel context was analysed using Linear Mixed Models (LMM) [13, 14, 15]. VPR was set as a dependent variable. The most complex model included (PHONOLOGICAL) VOICING, PLACE OF ARTICULATION (henceforth: PoA) and VOWEL (CONTEXT) as fixed factors allowing for their interactions along with random slope on VOWEL, POA and VOICING by SPEAKERS. The model selection (avoiding overfitting, and model simplification by anova()) resulted in the model where the fixed factors were VOICING and VOWEL allowing for interaction, and the random effects included only random intercept by SPEAKER. Pairwise comparisons were carried out via Tukey post hoc tests [16].

Generalised additive mixed models (GAMMs) were run to analyse CoG and skewness differences among the consonants in general, and along their duration [17, 18]. CoG, and skewness served as dependent variables. CONSONANT was set as ordered factor with contrast treatment, and set as parametric term. The smooth terms included TIME, TIME by CONSONANT (ordered factor). Random smooth was included on TIME by ITEM. No further autoregression treatment was needed.

Tongue contours were analysed by polarGAMs, separately for each speaker [19]. Vertical position of the tongue points (*y* dimension of coordinates) was set as dependent variable. The parametric terms contained the ordered factors of CONSONANT with contrast treatments. The smooth terms contained the TONGUE POINT (*x* dimension of the coordinates) by the CONSONANT, the NORMALISED TIME POINT within the consonant independently and by the CONSONANT (ordered factor), and the interaction of the TONGUE POINT and TIME POINT by the CONSONANT (using the te() function). In order to obtain *p*-values for each important consonant pair, two models were run with differently ordered CONSONANT levels. Due to computation capacity no random terms, but AR1 type autoregression correction was used.

The tongue front and root movements were analysed at three points of the estimated tongue contours taken from the TIME POINT * TONGUE POINT interactions of the polarGAMMs. The first column represents the tongue tip. It was not chosen from the edge of the estimates, as the GAMs methods may result in fuzzy data at the edges of the estimated ranges. This "tongue tip" point (TT) was taken from the closest point to the most anterior point of the tongue contour with largest movement excursion during the alveolars (for each speaker). The "tongue blade" point (TB) was selected based on the largest excursion during the postalveolars behind the TT point. The "tongue root" (TR) point was selected based on the heat map results, choosing the most typical point for the movement of the TR's area.

## 3. RESULTS

### 3.1. Acoustic data

The VPR of the voiced sibilants showed large variation within speakers, as well as between speakers. Eliminating the outliers, the VPR of spk01's voiced sibilants ranged between 0–75% (/z/), 0–60% (/ʒ/), spk02's between 0–100% (both), spk03's between 97–100% (/z/) and 67–100% (/ʒ/). Spk01's voiceless sibilants had 85–100% IQR of VPR (both), spk02's ones had 100–100% (/s/) and 80–100% (/ʃ/), and spk03's /s, ʃ/ had 100–100%, 67-100%, respectively. The best-fitting model was the one that included PHONOLOGICAL VOICING, VOWEL

CONTEXT and their interaction as fixed effects, and random intercept by SPEAKER ($R_m^2$=0.309, $R_c^2$=0.513). This means that the POA did not considerably affect the VPR. The main effect of PHONOLOGICAL VOICING ($F$(1,229)=137.597, $p$<0.01) and its interaction with VOWEL CONTEXT ($F$(3,229)=2.817, $p$=0.04) had a significant effect on this measure. The post hoc test of the interaction did not reveal any significant effect of the VOWEL CONTEXT on the difference between sibilants of the same phonological voicing.

The mean CoG and the mean skewness were significantly different among the four sibilants according to the parametric terms of the GAMMs ($R^2$=0.958, and 0.820, respectively). The smooth terms showed significant differences for both measures' change along the time course among all consonants. Plotting the smooth's estimated differences showed that the voiced counterparts' CoG-values were significantly different from appr. the 30% to 100% (both alveolars and postalveolars), the skewness from appr. the 25% to 85% (alveolars) and 70% (postalveolars) of the consonant duration. Figure 1 indicates that the CoG started increasing, and the skewness started decreasing for each speaker's sibilants at the time point where the voicing profile revealed a rapid drop in the phonation possibility, i.e., from which time point the phonation tended to cease in the speaker's pronunciation. While skewness of the voicing counterparts converged after the restart of voicing, CoG values remained significantly different between them.

### 3.2. Tongue positions and movements

Tongue position, and tongue movement were analysed by polarGAMs ($R^2$ = 0.855, 0.885, 0.886 for the three speakers respectively). The smooth of the TONGUE POINT averaged across the duration was significantly different between both voicing and PoA counterparts (Table 1, $p \leq 0.02$). This means that the tongue positions in general were different across the consonants. The smooth of the NORMALISED TIME POINT and the tensor of the interaction of TONGUE POINT and NORMALISED TIME POINT can be better understood by investigating the estimated smooths.

Figure 2 introduces three typical points for TT, TB, TR of the tongue contours. The TR's position was more advanced in the alveolars than in the postalveolars; however, it was speaker-specific, whether it was only a difference between the sibilants of the same phonological voicing category or in general. The TR was more advanced in the voiced sibilants during the entire duration in all speakers' pronunciation. Its position, however, changed during the consonant and moved backward for all four sibilants of spk03, and for the postalveolars of spk02. The TR in spk03's /s/ moved into a more anterior position, and in /z/, it did not change its position. In spk01's speech the TR moved backward, and in the last third of the production, the TR moved forward again in all 4 sibilants. The backward movement during the consonants may appear to tighten the pharyngeal region in order to control the intraoral pressure, which is responsible for the intense turbulent noise of the sibilants.

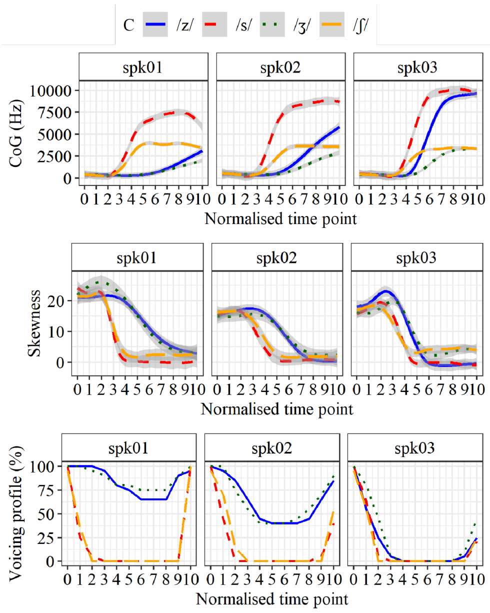


**Figure 1**: CoG (Hz, top), and spectral skewness (mid) estimated by GAM-smoothing by ggplot2 [20], and voicing profile (bottom) along the duration of the sibilants. (Color online.)

| C-pair | $F_{spk01}$ | $F_{spk02}$ | $F_{spk03}$ |
|---|---|---|---|
| /z/ vs. /s/ | 41.355 | 29.165 | 14.197 |
| /ʒ/ vs. /ʃ/ | 13.998 | 69.641 | 3.681 |
| /s/ vs. /ʃ/ | 44.290 | 29.165 | 38.582 |
| /z/ vs. /ʒ/ | 14.699 | 7.998 | 4.861 |

**Table 1**: The polarGAM-results.

TT and TB followed similar movement patterns for all sibilants within the speakers. It raised in the first part of the duration and then lowered from the third or half of the consonant duration in spk01's and spk02's pronunciation (except for spk02' /s/). Both points lowered along the consonant duration of spk03. In the present results, spk03 had no significant difference in TT or TB at all, spk02 had significant difference between the postalveolars between 0–40% of the duration, but not at all in alveolars. There was significant difference in the front of the tongue in spk01's alveolars between the 30–100% of the consonant duration, and in 0–80% of that in the postalveolars.

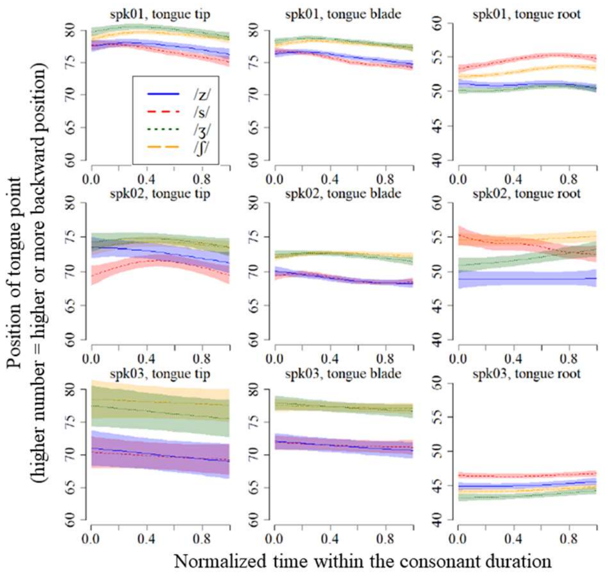


**Figure 2**: The tongue tip, tongue blade, and tongue root movement along the consonant duration extracted from the polarGAM. (Color online. Color and line type coding identical to Fig. 1.)

## 4. DISCUSSION AND CONCLUSIONS

The present study investigated dynamic articulatory and acoustic features, and voicing characteristics of four Hungarian sibilants regarding phonological voicing. Tongue root, tongue tip and tongue blade positions and movements were analysed by ultrasound tongue imaging. CoG, and spectral skewness, as well as the voicing profile and voiceless part ratio were also compared across the sibilants.

TR was advanced in voiced sibilants compared to the voiceless ones along the production to maintain phonation. This result contradicts the findings of a previous study on Hungarian /z/ and /s/, where this difference was only found for a part (most) of the speakers. The difference might appear due to the larger variability of phonetic context in the present study (four vowels were used here, while only /i/ context in [7]).

We hypothesised that the TR position was different between the alveolar and the postalveolar places of articulation due to the volume difference between the obstacle-vocal fold distances. In general, TR was found to have a more posterior position for postalveolars than for alveolars, probably as a result of POA differences. However, in some speakers' cases, phonological voicing seemed to override this trend, as voiceless alveolars exhibited a more posterior TR position than voiced alveolars. As TT and TB positions also showed considerable intraspeaker differences, these results may have an interrelation. Analysing English /s/ and /ʃ/, the groove width was found to be longer in the postalveolar sibilant [21]. Groove width difference was also found to vary, however, in a speaker-specific manner between Croatian /ʃ/ and /ʒ/ [22]. Together these findings may indicate that the target of the intense turbulent frication might be reached with somewhat speaker-specific tongue positions, which requires parallel EPG- and UTI-analysis in order to better understand the effect of the place of articulation and voicing together in its speaker-specific behaviour.

The third hypothesis was that tongue front and TR position changes along the sibilants' duration in a manner to maintain the intense turbulent frication. Intense turbulent frication might be reached by higher intraoral pressure behind the obstacle. This might be reached by i) higher tongue front, i.e., a narrower seal, or ii) (within a specific range of seal volume) with the backward movement of TR, i.e., constricting the pharyngeal region to increase the supraglottal pressure. The need of the backward movement might appear as a reason of the tongue fronts lowering found in each speakers' sibilants (transition to the following vowel), but might be also induced (as an additional reason) by the transition to the following vowel. In spk01's pronunciation, a fronting movement appeared in the last 10-20% of the sibilants. In her case, the phonation restarted sooner than in the other two speakers' pronunciation, which might be achieved by this fronting.

Based on the results of the present study, we conclude that the advanced tongue root of the voiced sibilants was found as expected compared to their voiceless counterparts. The tongue root was found to move backward during the duration of both voiced and voiceless sibilants that might appear to maintain intense turbulent frication of sibilants. Probably the relatively frequent occurrence and high ratio of devoicing found in these speakers appeared through the tongue front lowering due to the backward movement.

Due to the large interspeaker variability the analysis is planned to be extended to include a larger number of speakers. As vowel context also seems to have an effect on the data, this factor should also be further analysed in future studies. Naturally, due to the non-linear interrelation of articulation and acoustics [23], and the limited amount of information that midsagittal tongue contours give on the articulation, the spectral consequences of articulatory settings and changes can only be studied with questions necessarily being left open.

## 5. ACKNOWLEDGEMENTS

The work was supported by the Bolyai János Research Scholarship of the Hungarian Academy of Sciences (A.D., T.G.Cs.), ÚNKP-22-3 (K.J.), ÚNKP-22-5 (A.D., T.G.Cs.).

## 6. REFERENCES

[1] Stevens, KN. 1998. *Acoustic Phonetics.* MIT Press

[2] Stevens, K. N. 1997. Articulatory–acoustic–auditory relationships. In Hardcastle, W. J., Laver, J. (eds.): *The Handbook of Phonetic Sciences.* Blackwell. 462−506.

[3] Westbury, J. R. 1983 . Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *JASA* 73, 1322–1336.

[4] 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.

[5] Narayanan, S. S., Alwan, A. A., Haker, K. 1995 . An articulatory study of fricative consonants using magnetic resonance imaging. *JASA* 98, 1325–1347.

[6] 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, 245-259. https://doi.org/10.1121/10.0000556

[7] Gráczi, TE., Csapó, TG., Deme, A., Juhász, K., Markó, A. 2021. Tongue root position in VC sequences with regard to the phonetic realization of obstruent voicing: A preliminary study on Hungarian. In Tiede, M. Whalen, DH., Gracco, V. (eds.), *Proc. of the 12th International Seminar on Speech Production.* New Haven CT , Amerikai Egyesült Államok : Haskins Press. 198−201

[8] 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, 90–100.

[9] 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, 1588–1600.

[10] Boersma, P., Weenink, D. 2022. *Praat: doing phonetics by computer.* http://www.praat.org/

[11] Shih, C., Möbius, B., Narasimhan, B. 1999. Contextual effects on consonant voicing profiles: A cross-linguistic study. Proc. 14th ICPhS. 989–992.

[12] R Core Team 2022. R: *A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.* https://www.R-project.org/.

[13] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *J SS* 67, 1–48.

[14] Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *J SS* 82, 1–26.

[15] Bartoń, K. 2022. *MuMIn: Multi-Model Inference.* R package version 1.47.1, https://CRAN.R-project.org/package=MuMIn.

[16] Lenth, R. 2022. *emmeans: Estimated Marginal Means, aka Least-Squares Means.* R package version 1.8.3, https://CRAN.R-project.org/package=emmeans.

[17] Wood, SN. 2017 *Generalized Additive Models: An Introduction with R 2nd edition* . Chapman and Hall/CRC.

[18] van Rij, J., Wieling, M., Baayen, R., van Rijn, H. 2022. *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs.* R package version 2.4.1.

[19] Coretta S 2022 . *rticulate: Ultrasound Tongue Imaging.* R package version 1.7.3, https://CRAN.R-project.org/package=rticulate

[20] Wickham. H. 2016. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag.

[21] Fletcher, S.G., Newman, D.G. 1991. [s] and [ʃ] as a function of linguapalatal contact place and sibilant groove width. *JASA* 89, 850−858.

[22] Liker, M., Gibbon, FE. 2011. Groove width in croatian voiced and voiceless postalveolar fricatives. In Proc. of the XVII International Congress of Phonetics. Hong-Kong, 12−21 August, 2011.

[23] Stevens, KN. 1989. On the quantal nature of speech. *J Phonetics* 17, 3−45.