The realization of voicing opposition in alveolar fricatives in Hungarian. Preliminary study on articulation and acoustics

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

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

Abstract

The simultaneous articulation of the turbulent noise of fricatives and vocal fold vibration poses difficulties due to their conflicting pressure requirements. Previous studies found advanced tongue root and narrower obstacle in voiced fricatives than in voiceless ones. The first helps to maintain vocal fold vibration, while the latter helps to achieve the appropriate amount of turbulence.

In our study 12 subjects produced /izi/ and /isi/ sequences in pre-focal position. Headset microphone-, EGG- and tongue ultrasound (US)-signals were recorded. Cessation and restart points of voicing, and the voiceless part ratio (VR) were measured in the EGG-signal. CoG, SD, skewness and kurtosis were measured in the acoustic signal at 11 equally distanced time points in the fricatives. The midsagittal tongue contours were analyzed in the US signal in the closest image to the 0%, 50% and 100% points of the fricatives' total duration. Voicing characteristics of /z/ and /s/ were compared by LMM, the further spectral features were analyzed by GAMM, and the tongue contours were analyzed by polar GAMM.

The VR, the cessation and restart point of voicing were distinctive, although some of them had large VR in /z/ realizations. That may be resulted not only by the laryngeal settings but also by the supraglottal settings. The present study found tongue contour differences between the two fricatives at 50%, of the fricatives, and also at 0% and 100% point, but in less subjects' speech: suggesting advanced tongue root and narrower constriction in /z/ realizations and speaker dependent timing of gestures. The spectral measures did not reflect the US results in one-on-one way. That is explicable by the quantal relations of the two domains (Stevens, 1968), and we suggest that they are also a result of further articulatory maneuvers that are applied in the voiced and voiceless fricative pairs (see Liker & Gibbon, 2011, 2013, 2018).

Email addresses: `graczi.tekla.etelka@nytud.hu` (Tekla Etelka Gráczi), `csapot@tmit.bme.hu` (Tamás Gábor Csapó), `bartokmarton@gmail.com` (Márton Bartók), `deme.andrea@btk.elte.hu` (Andrea Deme), `marko.alexandra@btk.elte.hu` (Alexandra Markó)

Beszédtudomány – Speech Science 2020. 22–56. DOI: 10.15775/Besztud.2020.22-56

1. Introduction

Fricatives are produced with turbulent airflow through a narrow constriction in the oral cavity. The participating articulators and the size of the vocal tract in front of the constriction determine the resulting acoustic patterns (Fant, 1960; Shadle, 1991).

In order to produce a high intensity turbulent noise, the cross-sectional area of the oral obstacle must be smaller than that of the glottis, and the intraoral pressure needs to be larger than the atmospheric. The high intraoral pressure also means high supraglottal pressure, therefore the transglottal pressure differential decreases. The continuous increase of the pressure above the glottis also leads to an increase in the area of the glottis. The loss of the transglottal pressure differential and the increasing pressure in the glottis hinders vocal fold vibration. The vocal folds are first forced to vibrate slower than to stop vibration and stay apart (Bickley & Stevens, 1986; Stevens, 1997). As a result, voiced obstruents, hence also voiced fricatives may become partially (or totally) devoiced (Smith, 1997).

Although the intraoral pressure rises during the production of obstruents before full articulatory closure or constriction is reached (Müller & Brown Jr., 1980), and decreases only when the obstacle starts opening (or other articulatory strategies cause its decrease via the initiation of volume expansion of the oral cavity), vocal fold vibration does not cease in all voiced obstruents. There is a narrow range of pressure in which both voicing and friction could be maintained (Ohala & Solé, 2010). This can be reached by adjusting the cross-sectional area of the glottis and the constriction to approximately equal values (Stevens et al., 1992).

Voicing may also be maintained by other articulatory maneuvers / articulatory strategies through active or passive enlargement of the vocal tract during voicing. For instance, slower pressure build up can be achieved by expanding the area behind the obstacle (Docherty, 1992; Fuchs & Perrier, 2003). Further, lowering the larynx, enlarging the oral cavity, lowering the tongue, or forwarding the radix may also be used to expand the oral cavity, and thus to decrease intraoral pressure. The relaxation of the larynx with the resulting supraglottal area increase cause the slackening of the muscles close to the tongue surface that results further passive expansion of the cavities (Svirsky et al., 1997). While most of the earlier studies investigating articulatory maneuvers that may aid the maintenance of voicing concentrated on stops, more recent studies also aim to describe articulatory maneuvers in fricatives. Narayanan and colleagues (1995) studied /f θ s ʃ v ð z ʒ/ in American English in four speakers by MRI. They found advanced tongue root and larger pharyngeal area in voiced fricatives than in voiceless ones. Fuchs and her colleagues (2007) found that voicing during the frication interval was a less reliable discriminator of the voicing contrast, especially in Southern speakers of German and in word final position. Their results also showed that the relative voicing duration and the amount of tongue palate contact correlated who did not devoice, and that voiced fricatives showed more anterior articulation than voiceless ones (especially postalverolars).

The articulatory timing of voiced and voiceless fricatives were also found to be different in English and Croatian. In EPG-studies /s/ realizations were found to need longer time to achieve the largest contact surface at the constriction in both languages than the realizations of /z/ (Liker & Gibbon, 2013, 2018).

Studies on the area of contact using EPG showed contradictory results. Fletcher (1989) and Tabain (2001) did not find differences between the alveolar and the postalveolar voicing counterparts, while McLeod and her colleagues (2006) and Fuchs and her colleagues (2007) found a greater percentage of anterior tongue palate contact in the production of voiced fricatives than that of voiceless ones. Since vocal fold vibration results in lower intraoral pressure that would lead to lower turbulence intensity, narrower constriction may appear in order to avoid turbulence to be low in intensity. Additionally, the greater amount of contact may be attributed to a narrower medial groove. This hypothesis is also supported by the absence of this tendency in subjects, who devoiced voiced fricatives in their production (Fuchs et al., 2007).

Liker and Gibbon (2011) also studied the groove of the constriction in the anterior and the posterior part of the constriction in /ʃ/ and /ʒ/. Although the anterior groove was smaller in most speakers than the posterior, no consistent difference was found between the two counterparts. The anterior groove did not show major differences, while the posterior groove was larger in voiced fricatives in 3 out of 5 subjects' pronunciation. This shows that the parallel maintenance of the targets of voicing and friction exhibits differences across speakers.

The maintenance of vocal fold vibration in voiced fricatives and thus articulatory patterns were found to vary not only across speakers but also across languages (Shih et al., 1999).

The motivation of the research that the present study belongs to was to analyze the articulatory differences across the consonant duration. Our goal was to detect if the distinction that is present at the mind point of the fricative is also present already at the start and still at the end of the fricatives. We also aimed to analyze how the acoustic distinction is apparent during the consonants.

In the present study, we analyzed articulatory and acoustic patterns of voiced and voiceless alveolar fricatives in Hungarian. Our goals were (i) to describe the articulatory and acoustic distinction of voiced and voiceless alveolar fricatives in Hungarian, and (ii) to describe the timing relations of the articulatory and acoustic features of the voicing contrast in these fricatives.

2. Methods

2.1. Subjects

Twelve native female speakers of Hungarian were recorded. None of them had any speech or hearing impairments. Their age was between 20 and 27 years (mean: 22.25 years, sd: 1.5 years). All were given information on the procedure before the recordings both orally and written. All of them signed an informed consent before the recording.

2.2. Speech material

The analysed material is a part of a larger corpora we recorded previously (Markó et al., 2019). During the recordings, mini dialogues were introduced one by one to the subjects. Their task was to read the first utterance only for themselves (this part served as a context to the target sentence), and then to read aloud the next utterance (the target sentence) as an answer to the first one. The target sentences started with VCV# /nɛ/. Here we analysed only the /isi/ and /izi/ targets (5 per speaker). All target words occurred in focus position.

2.3. Recordings

The speech signal was recorded with a Beyerdynamic TG H56c tan omnidirectional condenser microphone at 44.1 kHz sampling rate and the tongue movement was recorded in midsagittal orientation using the "Micro" ultrasound system (Articulate Instruments Ltd.) with a 2–4 MHz / 64 element 20 mm radius convex ultrasound transducer at 83 fps. The vocal fold activity was captured by an electroglottograph (D200, Laryngograph LtD.) at 44 kHz. The speech signal was also recorded by the electroglottograph through a clipped microphone that was placed on the helmet used to stabilize the ultrasound probe at a fix distance (10-15 cm) away from the mouth. This speech signal was used to time-align the EGG- and the speech signal of the ultrasound recordings.

The segmentation of vowels was carried out by forced alignment (Mihajlik et al., 2010) and corrected manually in Praat (Boersma & Weenink, 2019), on the basis of the F2 trajectory.

2.4. Analyses

The EGG signal was analyzed in Praat (Boersma & Weenink, 2019). Voiced fragments within the target vowels were labelled automatically and corrected manually. Cessation point of voicing and restart point of voicing were also labelled. We also calculated the voiceless part ratio as the ratio of the duration of the unvoiced part to the total duration of the fricative. Additionally, the cessation point and the restart point of the vocal fold vibration were also calculated as shown below, where V_{end} is the cessation point of voicing, $V_{restart}$ is the restart time of voicing, Fr_{start} is the start point of the fricative, and Fr_{end} is the endpoint of the fricative:

- voiceless part ratio: $((V_{restart} - V_{end}) / (Fr_{end} - Fr_{start})) * 100$
- cessation point of voicing: $((V_{end} - Fr_{start}) / (Fr_{end} - Fr_{start})) * 100$
- restart point of voicing: $((V_{restart} - Fr_{start}) / (Fr_{end} - Fr_{start})) * 100$

Figure 1 shows a partially devoiced /z/ realization. It can be seen that the intensity of the vocal fold vibration decreases as frication is superimposed on it towards the middle of the consonant, and that it ceases (first arrow) as the target is reached in the frication component. The vocal fold vibration restarts after the frication target is reached, which is supposedly caused by the release of the constriction which allows the air pressure to decrease above and in the glottis, and to reach the transglottal pressure differential that allows vocal fold vibration.

The midsagittal tongue contours were manually traced in the AAA software (Articulate Instruments Ltd.) and then extracted in the Cartesian coordinate system. Ultrasound recordings insist of tongue contour images at every 12^{th} s, therefore any tongue contour in between these time points are averaged images from the closest before and the closest after images. Therefore we did not take the tongue contours at the exact 0%, 50% and 100% time points of the consonants, but we selected the closest "real" image to these points. At the starting and endpoints, the first/last image frame within the fricative was chosen.

The present analyses addressed the "anterior", "mid", and "posterior" parts of the tongue which terms refer to the parts of the tongue that can be seen in the ultrasound tongue contour. The real parts of the tongue that these terms approximately refer to are the following. The "anterior part" corresponds to the tongue tip and/or the tongue blade. The "posterior part" corresponds to the back and the root of the tongue. The "mid part" corresponds to the tongue body. These are, however, only approximations of the denoted regions, since (except from very few cases, where the tongue contour is very short) it cannot be reliably decided if the tongue contour seen in the ultrasound images is in fact the entire contour, or not. In the present study, several items had to be excluded from the analysis for technical reasons (no tongue contour could be detected or only a very short line appeared that was evidently only a small part of the subject's tongue surface): one /s/ realization by sp09 (all three measurement points), one 100% tongue contour in one /s/ realization by sp11, and seven 100% point tongue contours in /z/ realizations (one in the production of each of the following speakers: sp02, sp05, sp06, sp07, sp08, sp09, and sp11).

The acoustic analyses were also carried out in Praat (Boersma & Weenink, 2019). The spectral measurements were done both on the total duration of the consonant and at the three measurement points we listed above.

In order to measure the spectral characteristics for the total duration of the fricative, the fricative was extracted from the speech recording with rectangular window and transformed to spectrum slice with fast Fourier transformation. In order to measure the spectral features, the total speech recording was transformed to spectrogram using the Burg algorithm 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 Fourier transformation.

The center of gravity (CoG), the standard deviation of the spectral shape (SD), the skewness and the kurtosis was calculated (2^{nd} power) at each 10% of the consonant duration between 0% and 100%.

First, the actual manner of articulation of the /z/ realizations were grouped based on the CoG and EGG-data, then the fully voiced tokens with low CoG values were checked visually to separate the voiced fricatives and the approximant-like realizations. Partially devoiced /z/ realizations were not grouped further into minor groups.

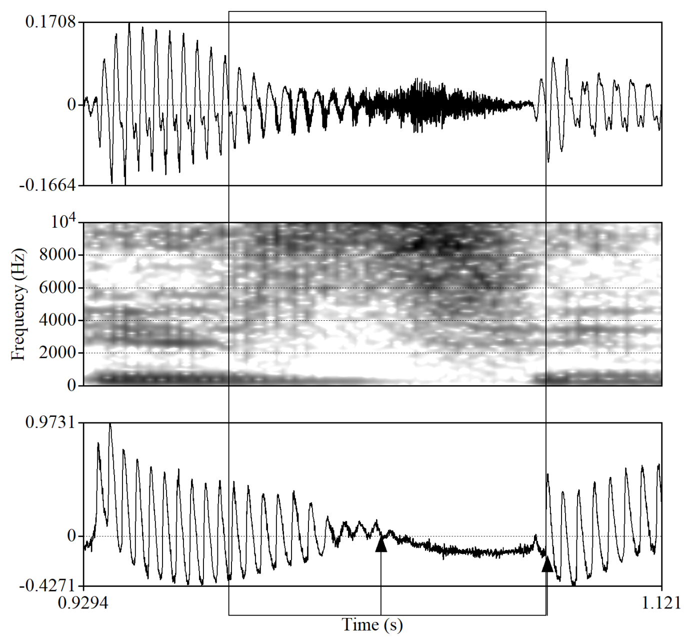

Figure 1: Partially devoiced /z/ in an /izi/ sample and its labeling. The top most box of the figure includes the oscillogram of the speech signal, the middle box includes the spectrogram of the speech signal, while the lowest includes the oscillogram of the EGG-signal. The rectangle shows the increase and decrease of the intensity of the frication noise in the speech signal. The decrease of the intensity of the vocal fold vibration can be observed. The first arrow shows the cessation of the vocal fold vibration, while the second marks the restart of voicing.

2.5. Statistical analyses

The statistical analyses were carried out in R (R Core Team, 2019).

Linear mixed models (Bates et al., 2015) were built in order to test the difference of the voiceless part ratio between the two fricatives, the cessation time, and the restart time of vocal fold vibration. In these models, voiceless part ratio, cessation time and restart time of voicing were used as the dependent variables. First a random intercept model was fitted (using the speakers) in a base model. The second model also included the consonant as fixed effect. The third one was further expanded with random slope for the consonants. The best fitting model was selected as the final model as determined on the basis of the Akaike Criterion (AIC-number) (Akaike, 1974) by using the anova() function. P-values were calculated by anova() in the lmerTest package (Kuznetsova et al., 2017).

The four spectral measures were analyzed by generalized additive mixed models (GAMM; Wood, 2017), the tongue contour was analyzed by polar GAMM (Coretta, 2019b) which is a modified version of GAMM especially for ultrasound tongue imaging. GAMM is a model that was elaborated for non-linear data, that are better described by fitting any function on the fixed effect (Wieling, 2018). This statistical approach determines the non-linear pattern automatically.

Tongue curves were analyzed by polar GAMMs in rticulate package (ultrasound tongue imaging in R; Coretta, 2019b). The models were built and compared separately for all speakers based on the suggestion of Coretta (2019a). The models were built for the temporal midpoints of the fricatives. Three models were built with maximum likelihood estimation. The horizontal placement of the measurement point (x coordinate value) was analyzed as a function of the vertical measurement point (y coordinate value). The models included a reference smooth by x-axis value and the consonant as fixed effect with the interaction between the x-axis value and the consonant.

The results of both GAMM and polar GAMM models include the comparison of the factor groups in general across the entire time interval/tongue contour, while the estimated difference can also be traced back along the time interval/tongue contour and the phases/parts can be detected where there are differences between the factor groups. Therefore if the difference consequently appears but only in a smaller time interval/region of the tongue contour and does not lead to an overall significant difference, it still can be detected.

The statistical analysis of the acoustic data (CoG, SD, SK, and KU) was also carried out by means of GAMMs with maximum likelihood estimation (mgcv package: Wood, 2017) in R. The models included the reference smooth of time, and the consonant as fixed effect. Random effect smooth of time was included in the model. We also fitted a separate model that included the random effect smooth of the time by consonant as the fixed effect. Autocorrelation was found in the data for CoG, SD and skewness, therefore it was incorporated in the model to remove its effects.

In the case of the GAMMs and polar GAMMs, the best fitting models were also selected on the basis of AIC determined using compareML of the itsadug package (van Rij et al., 2017). The smooth curves of the tongue contours were extracted from the fitted GAMMs used for the statistics. The smooth curves of the acoustic measures were extracted from a model separated for the subjects, in that the reference smooth and the consonant were included also allowing for interaction effects.

3. Results and discussion

3.1. Voicing features based on the EGG-signal

The devoicing pattern was different across the speakers (Figure 2). Six speakers did not devoice their /z/ realizations, or only once out of the five repetitions (sp01, sp02, sp03, sp04, sp11, sp12). Five subjects often realized /z/s with devoicing (sp05, sp07, sp08, sp09, sp10), and one subject (sp06) pronounced all 5 /z/s with a voiceless part ratio above 85%. Her /s/ realizations had the lowest amount of voicing, i.e. the highest voiceless part ratio among the 12 subjects. Figure 2 illustrates the voiceless part ratio of /z/ and /s/ realizations. The distinction of these consonants is evident in all speakers, even in the ones that tend to devoice their /z/ tokens. The linear mixed model with random slope for the consonant was found to describe the results best. The voiceless part ratio was significantly different between the two fricatives ($F(1, 11.095) = 62.820$, $p < 0.001$).

Figure 3 shows the normalized time point of the cessation and the restart of the voicing during the fricative. The voicing in the /z/ realizations of sp06, who devoiced these consonants in almost their entire duration, ceased before the 10% of the total duration was reached. The other speakers, who devoiced their /z/ realizations in the present study (sp05, sp07, sp08, sp09, and sp10) maintained voicing at least until the 25% time point of the fricative was reached. Sp10, however, did not show large variability, while the other subjects' voicing cessation time point varied. Voicing tended to restart at an earlier time point in devoiced /z/ realizations than in the /s/ realizations in those speakers' pronunciation who favored devoicing. The cessation time of the voicing was significantly different between the two fricatives ($F(1, 11.105) = 56.961$, $p < 0.001$). The restart time of voicing was analyzed in a subset of the data which includes only the subjects' who had partially devoiced /z/ realizations. Here, the model not including random slope fitted the data best, and the restart of voicing was significantly different between the two fricatives ($F(1, 52.180) = 12.879$, $p < 0.001$).

3.2. Tongue contours in the fricatives

3.2.1. Tongue contours at the mid point of the fricatives' duration

The analysis of the midsagittal tongue contours during the fricative production may reveal what maneuvers the specific speaker used to maintain vocal fold vibration. The models of polar GAMM including non-linear random slope for the consonant were found to have the lowest AIC in the case of 11 speakers, the second model (without non-linear random slope) was proven to describe data at the midpoint of the consonants the best. Each model had higher r2 than 0.95 which means that they explained at least 95% of the actual data. Table 1 includes the results for the comparison of /s/ and /z/ of these models. As seen in Table 1, the two fricatives were significantly distinct in the pronunciation of the

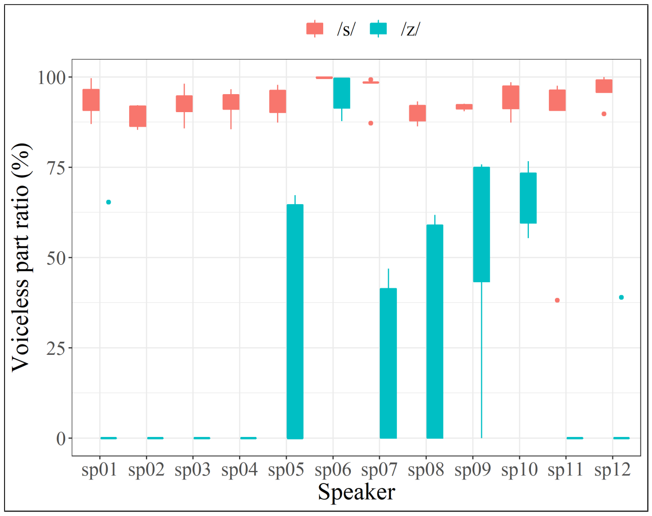

Figure 2: Voiceless part ratio (%) of /z/ and /s/

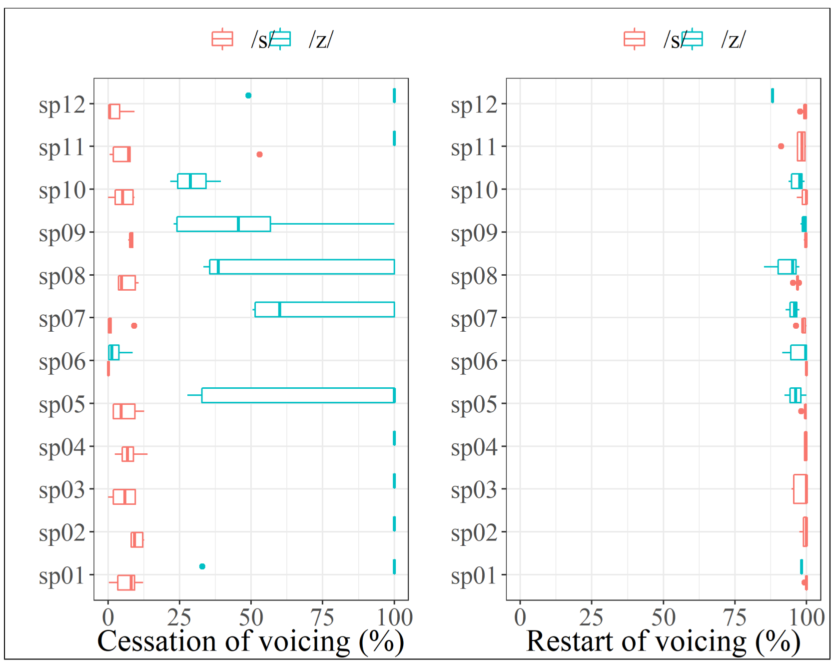

Figure 3: The normalized time of the cessation and the restart of the voicing. (The realizations that were voiced throughout their entire durations appear with a "cessation" of 100% in the left panel, but do not appear in the right panel.)

subjects sp01, sp02, sp05, sp07 and sp11 in their global tongue contours. The smooth of the models fitted on the speakers' data are shown in Figure 4. The estimated difference of the tongue contours is shown in Figure 5. The intervals in that the mean and the 95% confidence intervals of the estimated difference is not equal to zero are shown by red dashed lines. The posterior part of the tongue is on the right side of both figures. Figure 4 shows that even if only 5 out of the 12 subjects had global or large tongue contour differences between the two fricatives, there were consequent differences in some regions of the tongue contours in most subjects' pronunciation.

Table 1: The t- and p-values of the polar GAMM models for the tongue contour differences at the midpoint of the fricatives between /s/ and /z/ realizations.

	sp01	sp02	sp03	sp04	sp05	sp06	sp07	sp08	sp09	sp10	sp11	sp12
t	-2.770	-5.742	0.091	0.953	-2.213	-1.431	-7.394	1.347	-1.071	3.626	3.183	-1.317
p	*0.039*	*<0.001*	0.928	0.341	*0.028*	0.154	*<0.001*	0.179	0.286	*<0.001*	*0.002*	0.190

The posterior part of the tongue was lower in /z/ realizations, and the mid region and/or the anterior region of the tongue contours were also different between the two fricatives in 10 speakers' pronunciation. Sp10, however, produced /z/ realizations with higher vertical tongue position in the posterior tongue contour region and without further differences at the other regions of the tongue. The estimated difference of the tongue contours was very low in the case of sp11; however, it was consistent throughout the entire tongue contour.

3.2.2. Tongue contours at the start point of the fricatives' duration

The smooth of the models of polar GAMMs of the tongue contours at the start point (0% duration) of the fricatives are shown in Figure 6, and their estimated differences are shown in Figure 7. The directionality of the tongue contours is identical to those in Figure 4 and 5, i.e., the posterior part is on the right of the panels.

The t- and p-values of the polar GAMM for the difference between the tongue contours at the start point of the consonants is shown in Table 2. In sp03 and

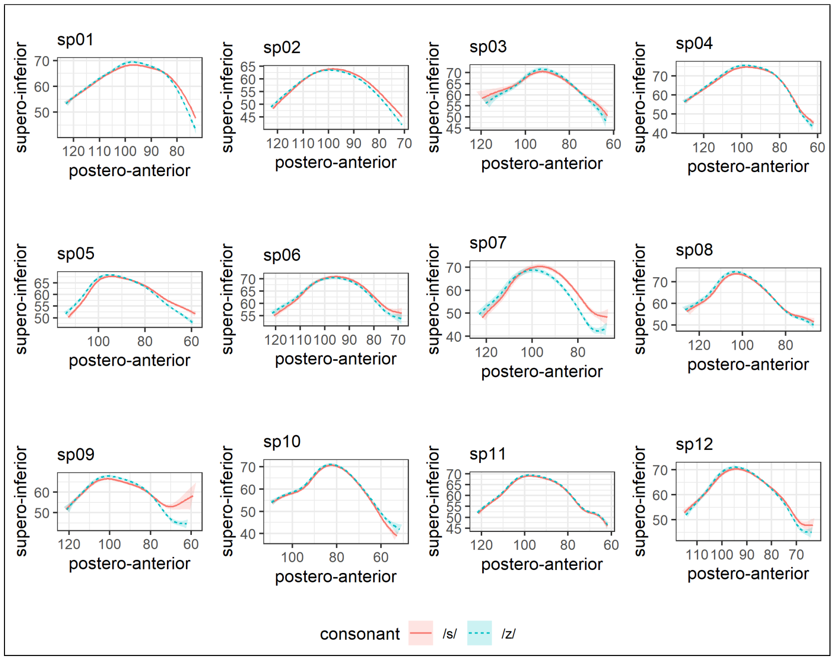

Figure 4: The smooth (mean, 95% CI) of tongue contours at the midpoint of the fricatives. The posterior part of the tongue contour is on the right side of the figure, the anterior part is on the left side.

sp12 the first, basic model yielded the lowest AIC values that did not include the consonant as factor. In the case of the other speakers, either the model including the consonant as factor with random slope for the consonants, or the one without random slopes (sp04) yielded the lowest AIC value. The best fitting models explain at least the 88.8% of the deviance.

The results for the regions of the midsagittal tongue contours are the followings. As shown in Table 2, the tongue contours at the start point of the fricative duration did not show any differences between the two consonants in five subjects' pronunciation (sp03, sp04, sp08, sp11, sp12). And according to Figure 6 and 7, in line with midpoint data, the posterior part of the tongue contour shows a difference between the consonant pairs at the start of the pronunciation in the seven further subjects' samples. This difference was relatively small in the case of sp06. In this speaker's case this small difference was the

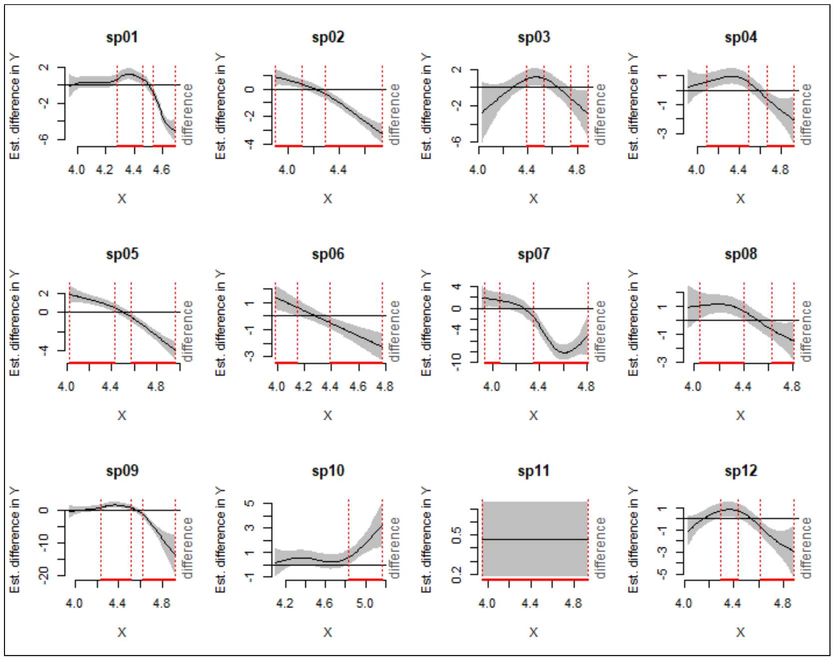

Figure 5: The estimated difference of the smoothed tongue contours at the midpoint of the fricatives. The red dashed lines indicate the intervals in which the mean and CI of the estimated differences of the contours are not equal to zero. The anterior part of the tongue contour is on the left side of the figure, the posterior part is on the right side.

only difference, while in the other seven subjects the difference was larger and appeared in the anterior and/or mid region of the tongue contour, as well. In one subject this difference did not appear at the backmost part of the tongue (sp01), but slightly anterior to that. The mid part of the tongue contour was different in five subjects' production between the two consonants at the start point of the total duration (sp01, sp02, sp07, sp09, sp10). In the case of sp02 and sp07 the entire mid-posterior part of the tongue appeared distinct at this time point. The anterior part of the tongue showed difference in four speakers' pronunciation at the start time point of the fricatives (sp01, sp02, sp05, sp07). Although in two subjects' production (sp01, sp07) only a small anterior region showed difference between the two fricatives, but a large posterior-mid region was distinct.

Table 2: The t- and p-values of the polar GAMM models for the tongue contours at the start point of the fricatives. ('-' denotes cases where that the first, basic model had the lowest AIC score.)

	sp01	sp02	sp03	sp04	sp05	sp06	sp07	sp08	sp09	sp10	sp11	sp12
t	-0.137	-6.901	-	0.560	-3.963	-0.201	-5.74	0.511	0.685	-1.303	1.493	-
p	*0.910*	*<0.001*	-	0.120	*<0.001*	0.841	*<0.001*	0.610	0.494	0.194	0.137	-

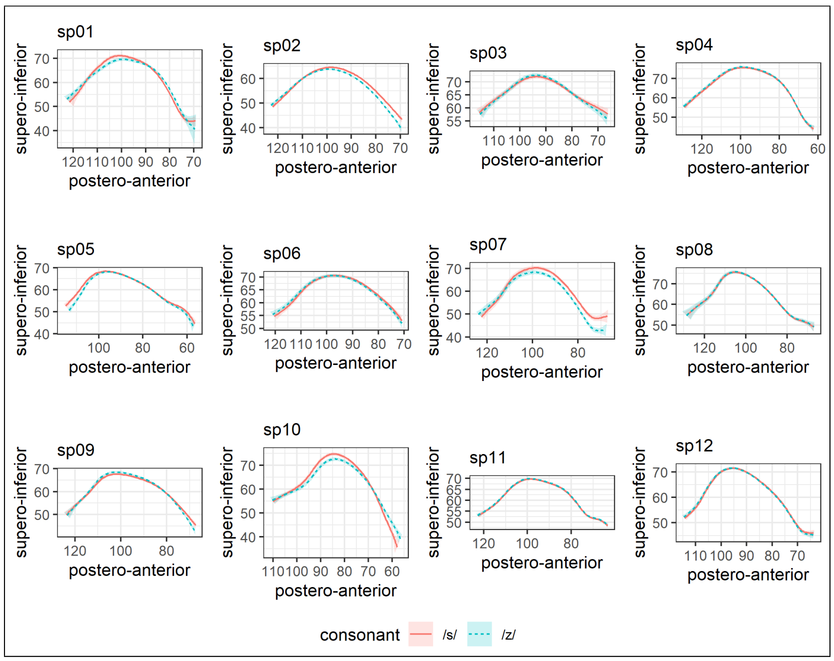

Figure 6: The smooth (mean, 95% CI) of tongue contours at the start point of the fricatives. The posterior part of the tongue contour is on the right side of the figure, the anterior part is on the left side.

3.2.3. Tongue contours at the endpoint of the fricatives' duration

The smooth of the tongue contours at the endpoint of the consonant duration are shown in Figure 8, their estimated differences are shown in Figure 9. The directionality of the tongue contours is again identical to those in Figures 4, 5, 6, 7.

The global difference between the two fricatives showed the following results. In the case of most speakers (sp01, sp02, sp05, sp06, sp07, sp09, sp10, sp11, sp12)

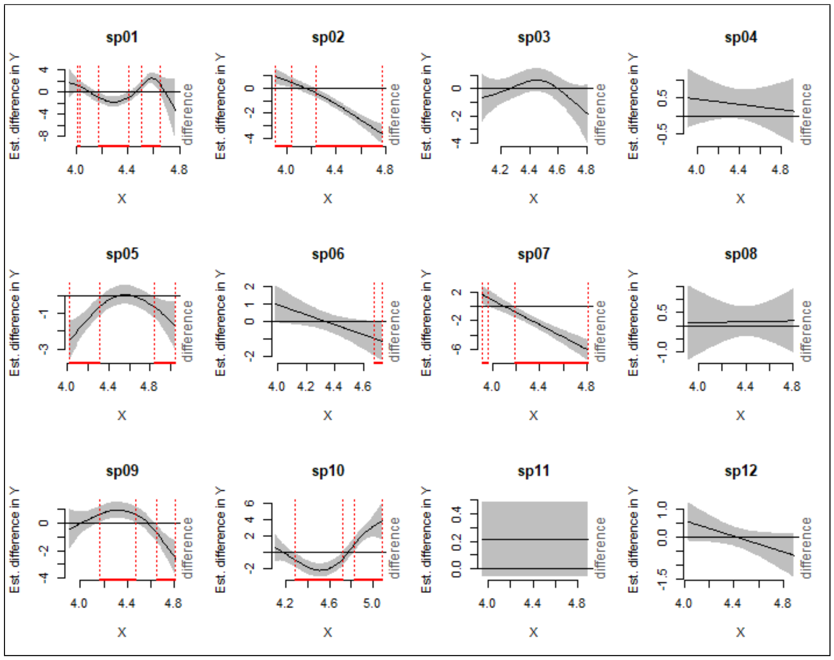

Figure 7: The estimated difference of the smoothed tongue contours at the start point of the fricatives. The red dashed lines indicate the intervals in that the mean and CI of the estimated differences of the contours are not equal to zero. The anterior part of the tongue contour is on the left side of the figure, the posterior part is on the right side.

the third polar GAMM model yielded the lowest AIC score, i.e., the model that included the consonant as factor and random slope on the consonants. In case of sp03 and sp04 the second model had the lowest AIC value in which the consonant was included as factor but no random slopes were added. In the case of sp08 the first, basic model yielded the lowest AIC score, that did not include the consonant as factor. The t- and p-values of the polar GAMMs are shown in Table 3. The results showed that the difference of the tongue contours between the two fricatives at the endpoint of the consonant was significant in six out of the twelve subjects (sp02, sp03, sp05, sp06, sp07, sp09). The best fitting models explain at least the 92.1% of the deviance.

The various regions of the tongue contours showed, however, further differences. Only three speakers (sp04, sp08, sp11) did not have any difference between the two consonants. Difference was present in the posterior part of the tongue contours in eight subjects (sp01, sp02, sp05, sp06, sp07, sp09, sp10, sp12). However, in the case of sp01 and sp10 the difference was the opposite of that found in the other subjects', since in sp01's and sp10's production, the posterior tongue region was higher in the voiced fricative than in the voiceless one. While in the case of sp12 this was the only tongue contour region that showed any difference between the two fricatives, in sp03 it was only the anterior part that showed any difference. In the pronunciation of sp01, sp02, sp07, and sp09 the anterior part and in some cases the mid part of the tongue contour also showed a difference, and in the case of sp05 and sp10, the mid part (or a nearby region) of the tongue contour showed further differences.

Table 3: The t- and p-values of the polar GAMM models for the tongue contours at the start point of the fricatives. ('-' denotes cases where means that the first, basic model yielded had the lowest AIC score.)

	sp01	sp02	sp03	sp04	sp05	sp06	sp07	sp08	sp09	sp10	sp11	sp12
t	-1.192	-2.882	-2.220	0.542	-4.861	-2.417	-5.457	-	-4.796	-0.543	1.719	-1.385
p	0.234	*0.004*	*0.029*	0.124	*<0.001*	*0.017*	*<0.001*	-	*<0.001*	0.587	0.087	0.168

3.2.4. Comparison of the estimated differences of the tongue contours at the start, mid and endpoint of the fricatives

Comparing the tongue contours observed within the group of either /z/ or /s/ realizations at the start, mid and endpoint of the fricative duration, the following four distinct tendencies were found (compare Figure 5, 7 and 9).

a) There was no difference in the tongue contours at the start and the end of the consonants, but there was a difference at the midpoint (sp04, sp08, sp11).

b) There was no difference in the tongue contour at the start point of the consonants, but the midpoint showed a greater difference, and the endpoint still showed some, relatively smaller difference (sp12).

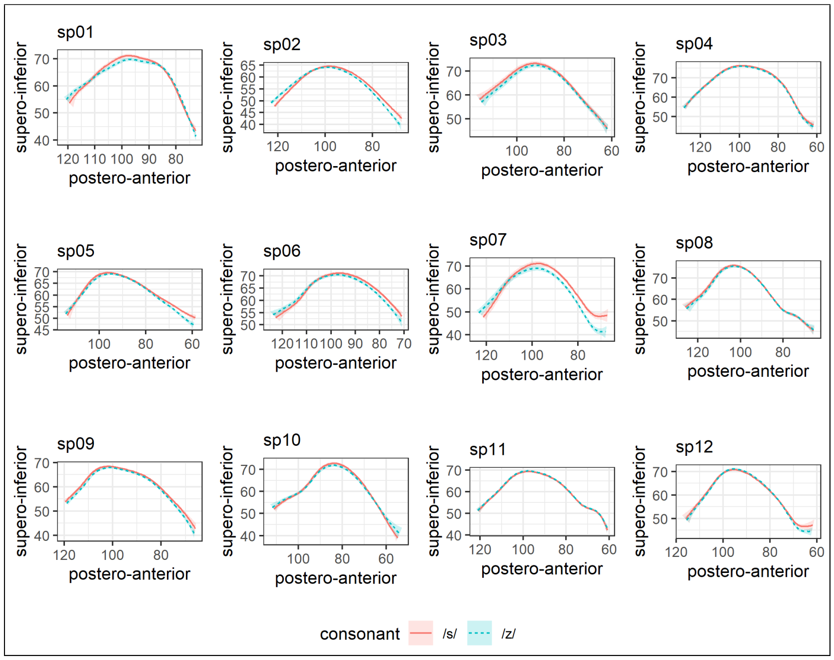

Figure 8: The smooth (mean, 95% CI) of tongue contours at the endpoint of the fricatives. The posterior part of the tongue contour is on the right side of the figure, the anterior part is on the left side.

c) There was some difference in the tongue contours at the start of the fricatives, and the difference was large at both the mid and endpoint of the consonant (sp03, sp06).

d)) Six speakers produced a great distinction in tongue contours throughout the entire fricative duration (sp01, sp02, sp05, sp07, sp09, sp10).

There were some evident tendencies in the differences also with regard to which region of the tongue was concerned.

a) In the cases, in which there was any difference between the two fricatives, the posterior part of the tongue showed a difference (with the only exception of sp03, who showed a difference elsewhere on the tongue contour at the end of her fricatives' duration. The posterior part of the tongue contour was higher in /z/ than in /s/ realizations, with the exception sp10

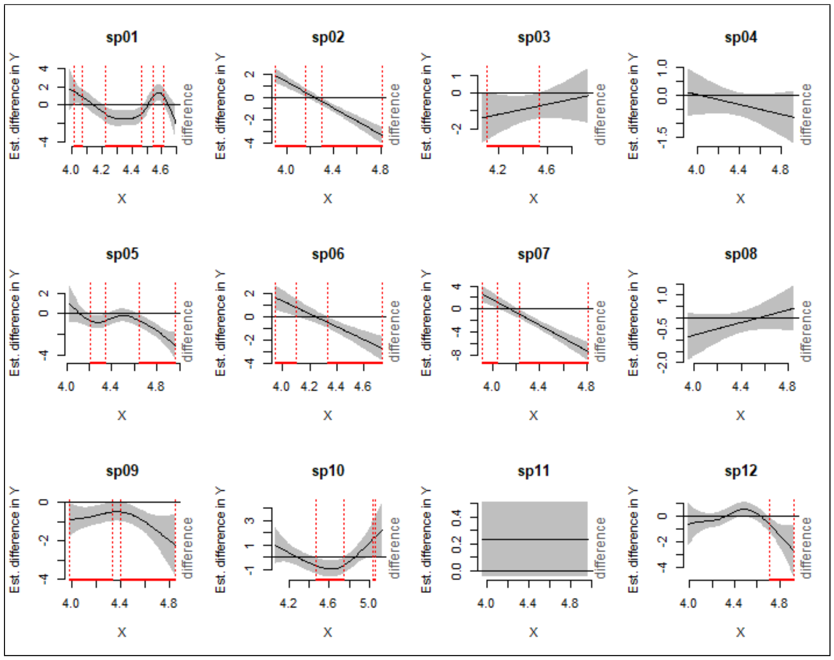

Figure 9: The estimated difference of the smoothed tongue contours at the endpoint of the fricatives. The red dashed lines indicate the intervals in that the mean and CI of the estimated differences of the contours are not equal to zero. The anterior part of the tongue contour is on the left side of the figure, the posterior part is on the right side.

who showed elevation in /z/ realizations in the entire tongue contour, and sp01 who showed elevation in /z/ realizations only at the endpoint.

b) In most cases, there was a difference also at the anterior and/or mid part of the tongue as well. If there was a difference in the position of the anterior part of the tongue, it was always raised higher in /z/ than in /s/ realizions. The difference in the mid region varied across subjects at the start- and endpoints, while it was higher in /z/s at the mid time point of the fricatives.

The subjects who showed devoicing or who did not could not be separated based on the differences in the tongue contours between the two fricatives.

3.3. Spectral measures

Realization of voiced fricatives can be diverse. For instance, in order to maintain voicing, /z/ may be realized approximant-like, but in this case there is no frication. If, however, the speaker favors to maintain the frication noise, voicing may cease during the consonant duration. Also, there are various possibilities between these two ends of the scale, in that the amount of the turbulence and the voicing can vary. Figure 10 shows two /izi/ realizations. The one on the left was produced by sp01, whose voiced fricatives appeared with voicing throughout their total duration in general. This particular realization on the left of the figure had lower intensity friction than the realization on the right side which was pronounced by sp05. This latter /z/ token was realised with partial devoicing. Their CoG at the midpoint of the token on the left of the figure was 658 Hz, while it was 6840 Hz for the token in the right.

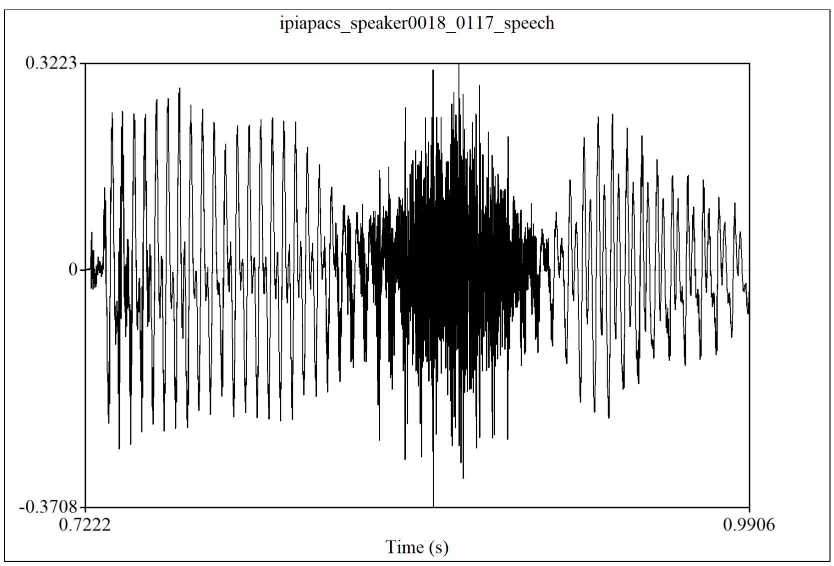

Figure 10: Oscillograms of the acoustic signal in two realizations of /izi/ (left: sp01, right: sp05).

The measurement points throughout the duration of the consonant will be short addressed as “time”. The model of best fit was the one which included consonant as a factor, and a random slope for time in the case of all four acoustic variables. Autocorrelation was detected in CoG, SD and skewness, but not in kurtosis and was involved therefore in the model. We found that in general, only CoG was significantly different between /s/ and /z/ ($t = -10.054$, $p < 0.001$), while the other three measures did not show significant differences in the global comparison of data as handled as two distinct time series. The best fitting models explain the 86.5% of the deviance in CoG, 72.3% of the deviance in SD, 82.5% of the deviance in skewness and 66.9% of the deviance in kurtosis.

3.3.1. CoG

The CoG curve of /s/ realizations was fairly similar across the speakers (Fig. 11), an abrupt increase appeared in the first 10-20% of the duration and an abrupt decrease in the last 10-20% of the consonant, while the middle part showed a slow change or plateau. This was expected as the CoG in voiceless fricatives is largely affected by the reach of the target constriction and then by the release which leads to the target of the following vowel. The realizations of /z/, however, showed various patterns (Figure 11). In eight speakers , the CoG slowly increased until the 50% of the consonant duration, then it showed the same pattern as /s/ realizations (slow increase in CoG: sp03, sp05, sp07, sp08, sp09, faster increase in CoG: sp01, sp06, sp10). The distinction between the two fricatives disappeared or was very low from the 50-60% of the consonants durations (sp01, sp03, sp06, sp07, sp09), and from approx. 30% in sp10. As observed in data on the temporal organization of voicing, some of these speakers frequently devoiced their /z/ realizations (e.g., sp06), or varied in their devoicing pattern (e.g. sp07), while others had voicing throughout the entire duration of /z/ realizations (sp01, sp02, sp03, sp04, sp11, sp12). Four speakers (sp02, sp04, sp11, sp12) /z/ tokens were often realized with a very low CoG in its entire duration, that either could be a result of approximant-like, or that of a very low intensity turbulent noise with voicing throughout the fricative's duration (as observable in the left panel of Figure 10).

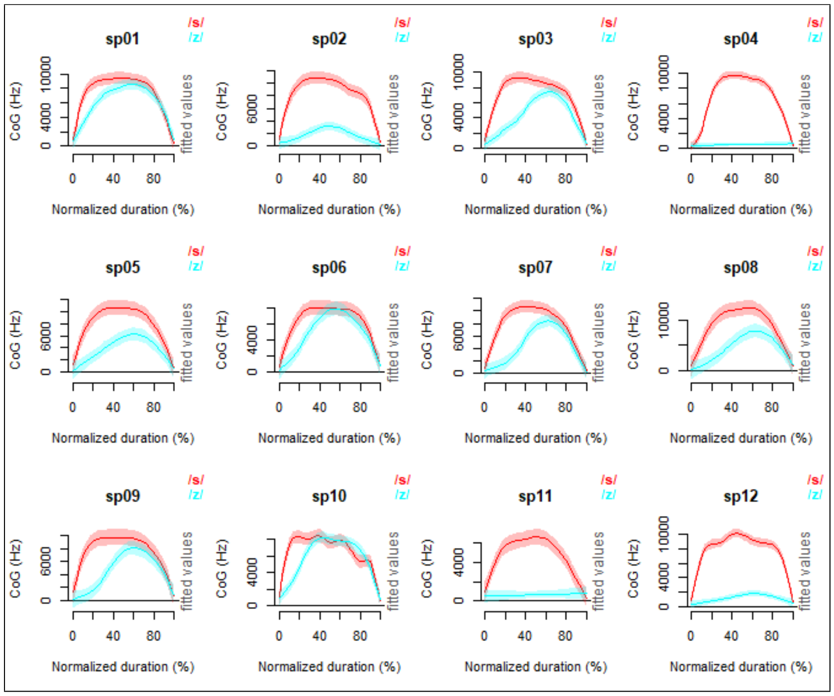

Figure 11: Fitted smooth curves of CoG by speaker (mean, 95% confidence interval)

3.3.2. Spectral variability (SD)

Spectral SD is shown in Figure 12. In the case of /s/ the curve was rather similar across speakers, which is expected and explicable as follows. While voicing ceases abruptly and the turbulent noise gets more intense, the CoG increases, and the higher frequency regions become more dominant in the signal than the lower regions. This leads to a fast increase in SD at the start of /s/. The constriction release, and thus the loss of dominance of the higher frequency regions leads to a fast decrease not only in CoG, but also in SD at the end of the fricative, and the interval between the 10-to-90% of the duration showed a valley in most speakers in /s/. After the cessation of voicing and the reach of the target/maximum CoG, the low-frequency region of the spectrum does not add to the variability of the spectrum. As a result of the above, in general, the shape of the spectral SD in /s/ showed the following trend: an abrupt increase followed by a somewhat variable valley, and then by an abrupt increase again.

In the case of sp10 and sp11 the first, increasing part of the above outlined tendency was slow, and the first SD maximum was reached at the point where the second decrease was expected based on the other speakers' results. The valley did not appear in their case, but the abrupt decrease following the slow increase did.

The realizations of /z/ showed a dome shape in the production of six speakers (sp02, sp04, sp05, sp09, sp11, sp12), while in all the further speakers', a valley appeared in these tokens as well, similarly to /s/ realizations. The shape of the spectral SD in /z/ can be explained by voicing and CoG results taken together: the cessation of voicing and the appearance of lower and higher frequency components.

As we found the spectral SD time series curves to be variable, we can conclude on no systematic tendencies with regard to the distinction of the voicing contrast in spectral SD.

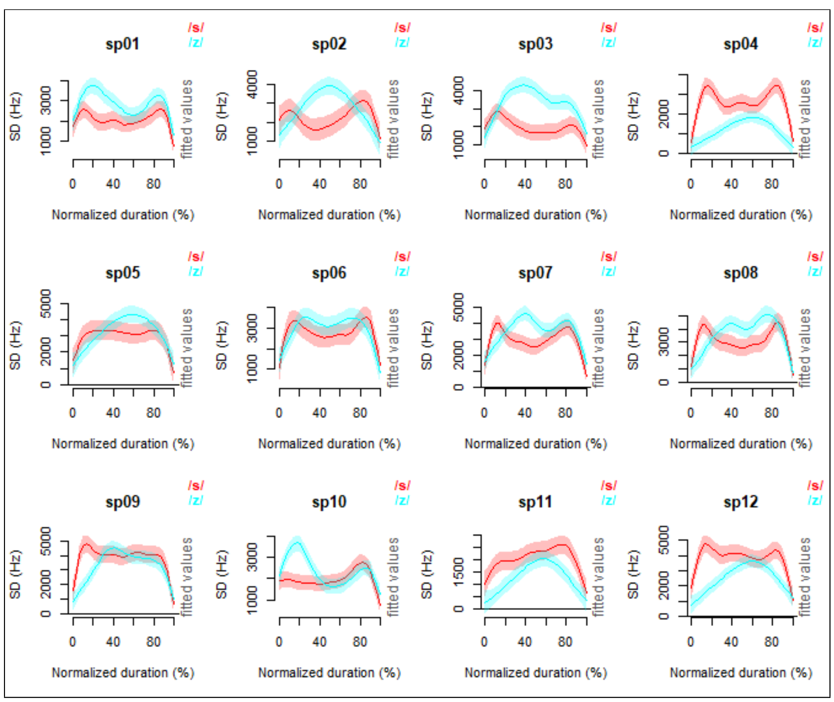

Figure 12: Fitted smooth curves of SD by speaker (mean, 95% confidence interval)

3.3.3. Skewness of the spectral components

The skewness decreased abruptly at the start of the consonants and increased abruptly at the end (Fig. 13). This value was less affected by voicing even in the voiced fricatives: /z/ realizations showed less abrupt decrease at their starting phase; however, the overlap between the members of the consonant pair was reached in the first 30% to 40% of the fricatives regardless of the speakers' devoicing tendencies. The two consonants were distinct throughout their entire duration only in four speakers (sp02, sp04, sp11, sp12). Each of these four speakers had voicing throughout the entire duration of /z/ realizations, and showed low CoG, dome shaped spectral SD, and low ratio of overlap between the two fricatives in spectral SD.

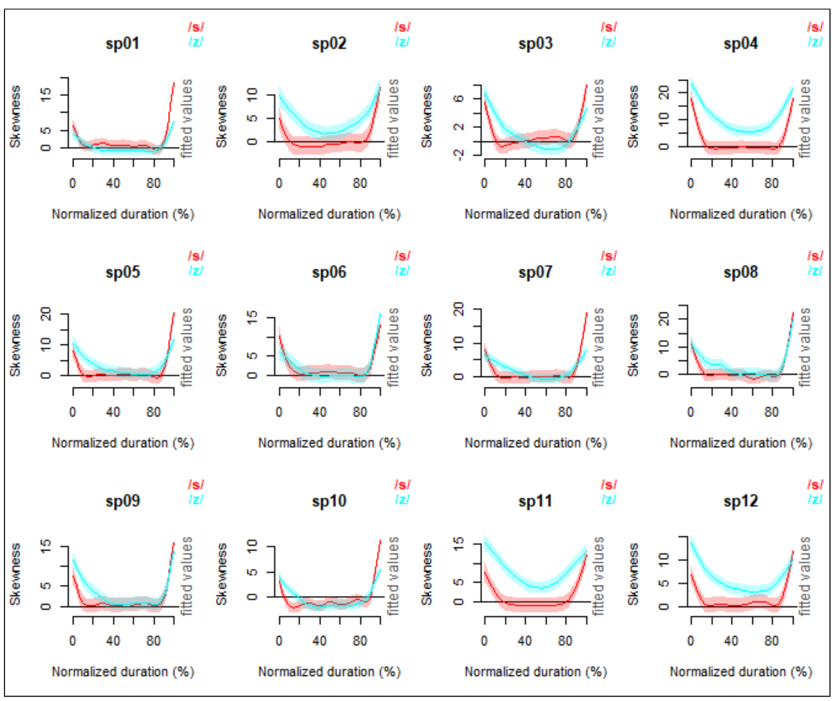

Figure 13: Fitted smooth curves of skewness (mean, 95% confidence interval)

3.3.4. Kurtosis of the spectral components

Compared to the other parameters, kurtosis stayed fairly constant throughout the entire consonant duration (Fig. 14). With respect to kurtosis values in first 10-20% of the fricative duration three distinct tendencies were found: i) it showed either minor or no decrease in the two fricatives (sp01, sp05, sp06, sp07, sp08, sp09), ii) it showed great decrease in both fricatives (sp03, sp04), or iii) it showed smaller decrease in /s/ realizations and larger decrease in /z/ realizations (sp02, sp11, sp12). In five speakers (sp02, sp03, sp04, sp11, sp12), kurtosis of the members of the fricative pairs was different but only in a smaller portion of the total consonant duration, typically at the start of the fricatives , while in other speakers /s/ and /z/ realizations did not show any difference in the kurtosis.

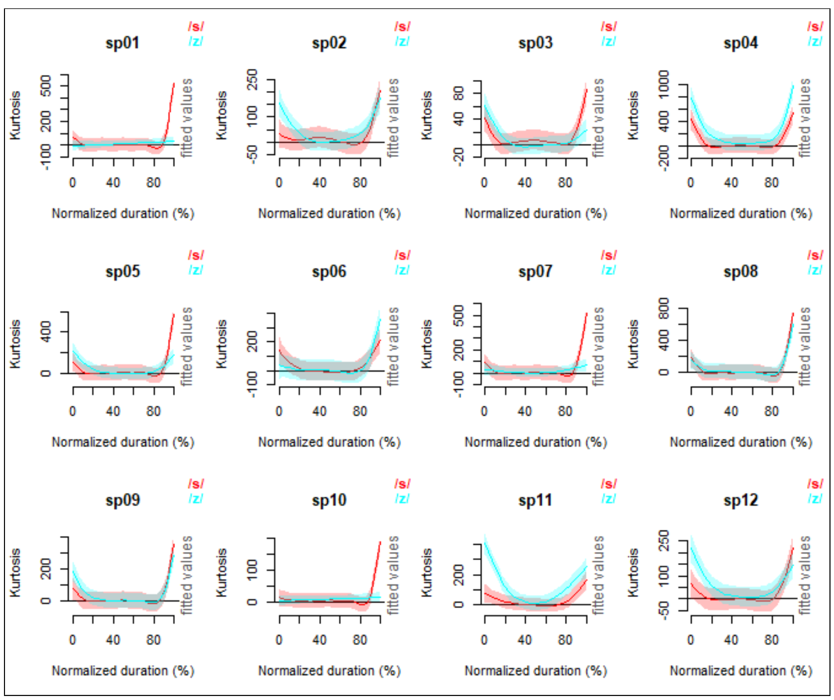

Figure 14: Fitted smooth curves of kurtosis (mean, 95% confidence interval)

4. Conclusions

In the present paper we aimed to (i) describe the articulatory and acoustic distinction of voiced and voiceless alveolar fricatives in Hungarian, and (ii) to describe the timing relations of the articulatory and acoustic features of the voicing contrast in these fricatives.

Since the fricatives in question were analyzed in intervocalic position, the voicing contrast was hypothesized to be apparent in vocal fold vibration throughout the entire consonant duration in most cases in /z/ realizations, i.e., the voiceless part ratio was expected to stay low, close to 0% in /z/. Earlier studies showed that the voicing characteristics of voiced obstruents are diverse across languages and speakers (e.g., Shih et al., 1999), and that fricative realizations in Hungarian were also diverse across speakers both in read and spontaneous speech (e.g., Bárkányi & Kiss, 2009; Gráczi, 2012). In line with these results, interspeaker variation was also demonstrated in the present study revealing speakers not devoicing their phonologically voiced fricatives, or partially devoicing some of their phonologically voiced fricatives, and one speaker who devoiced all of his/her phonologically voiced fricatives. Nevertheless, the devoiced part ratio was lower in /z/ realizations than in /s/ realizations in all 12 subjects, and voicing ceased later and restarted earlier in voiced fricatives than in voiceless ones. This latter finding is the result of the fact that the target in /z/ is voicing, and the reason of the cessation of voicing in /z/ is the intraoral pressure build up (Bickley & Stevens, 1986; Stevens, 1997), while in /s/ the target is voiceless. Therefore, the vocal fold settings may be different in the two consonants even in those fragments where no vocal fold vibration is found, and the voiced phoneme realizes as devoiced. While the restart of voicing in /z/ depends on the decrease of intraoral pressure, as the vocal folds are already set for voicing, in the case of /s/, the restart of voicing is primarily controlled by the voicing gesture of the following vowel.

Articulatory results of previous MRI-, EPG- and ultrasound-studies (e.g., Rothenberg, 1967; Kent & Moll, 1969; Perkell, 1969; Westbury, 1983; Ahn, 2018;

Coretta, in press) provided evidence of the articulatory maneuver of advancing tongue root in voiced fricatives to avoid cessation of voicing. This articulatory maneuver broadens the posterior region of the oral/pharyngeal cavity which results in slower intraoral pressure build up and thus in slower/less frequent devoicing in phonologically voiced fricatives. In our study, this maneuver was also demonstrated, as we found lower position of the posterior region of the midsagittal tongue contour in most speakers in /z/, and this region approximately corresponds the root of the tongue. However, while Coretta (in press) found that the advanced tongue root was already present in the preceding vowel in the case of voiced stops, in the present study we did not find this difference at the starting point of fricatives in seven out of twelve (i.e., in 58.3% of the) speakers. The difference in these results might be a consequence of a difference in the phonetic realisation of the voicing contrast across languages. However, this assumption is to be analyzed in future studies which extend their scope beyond the one vowel context of /i_i/. In the present data we also found that the difference in the vertical position of the posterior region of the tongue contours (i.e., the alleged advanced tongue root) was present in most subjects also at the end of the fricatives. This may be due to the fact that the volume expansion of the vocal tract at the end of the fricative may also help the earlier restart of voicing in the phonologically voiced fricatives, which were realised as devoiced.

At the midpoint of the fricatives, the anterior or mid region of the tongue contour was higher in /z/ than in /s/ in most speakers who showed lower posterior regions in /z/ than is /s/. This tendency was also observable at the boundaries of the fricative, although the position difference of the mid region was the opposite (and showed higher position in /s/) in some speakers'. Although we cannot exactly tell if these 'anterior' points always reflect the position of the tongue tip or the tongue blade (as the ultrasound is not always able to show the entirety of the midsagittal view of the tongue surface), we may claim that our results are in line with those of Narayanan and colleagues' (1995). The MRI-study of Narayanan and colleagues' (1995) revealed that the alveolar fricatives in English may be articulated with either the blade or the tip. Stevens and colleagues (1992) found smaller cross-sectional area at the oral constriction than at the glottis for voiced fricatives. These results are also in line with our results, as the higher position of the anterior regions we found here does also suggest a similarly narrow constriction, which may provide support to the fricative to reach its target frication intensity through the lower pressure build up (Fuchs et al., 2007). The higher position in the mid region of the tongue contour in /z/? partly contradict previous results, as we expected a larger oral volume, that is, a lower tongue position behind the obstacle, which would help to maintain voicing (Docherty, 1992; Fuchs & Perrier, 2003).

Three further questions regarding the articulatory maneuvers speakers may utilize to avoid devoicing in phonologically voiced fricatives cannot be addressed in the present study, in which we analyzed 2D midsagittal ultrasound data with start, mid and end point measurements. We cannot determine i) the groove size, ii) the area of the tongue-palate contact, nor iii) the timing of the oral articulation of the fricatives at hand. EPG-studies of English and Croatian showed that /s/ realizations needed longer time to achieve the largest contact surface at the constriction in both languages than /z/ realizations (Liker & Gibbon, 2013, 2018). The tongue-palate contact was found to correlate with the occurrence of devoicing (Fuchs et al., 2007).

The GAMM analysis of acoustic parameters in /z/ and /s/ as two sets of time series data taken at eleven measurement points showed, that only CoG was significantly different between the /z/ and /s/ realizations in general. The detailed analysis of the time series data, however, showed that there are differences in the time course of each analyzed spectral measure with a considerable variability across speakers in /z/ realizations, and much less variability in /s/ realizations. /s/ realizations had an abrupt increase in CoG, and in spectral variability and an abrupt decrease in skewness in the first 10-20% of their duration, while the change was opposite in its direction, but similarly abrupt at the final 10-20% of the fricatives. Kurtosis varied with regard the abruptness in the first 10-20%. In the middle portion of the fricative we found a dome like shape in CoG and most often a valley in the other spectral measures.

The CoG shapes of /z/ realizations varied across the subjects. Taken together, spectral results suggest that voiced and partially voiced fricative realizations both occurred, as well as approximant-like realizations. However, devoiced part ratio data and the midsagittal tongue contour shape together are not enough to describe the articulatory-acoustic relationships. Most importantly, articulation and acoustics show aquantal relationship (see Stevens, 1968), but two further reasons are also important to mention here. One of these is that the measurement of the distance of tongue and the palate is difficult and unreliable in ultrasound images as the palate cannot be traced in detail (despite the use of wet swallowing or other tricks that may partially reveal the palate contour), but neither can we be sure if an ultrasound image includes the entirety of the midsagittal tongue surface. The second main reason is that the earlier EPG-studies not only revealed differences of the grooves, tongue palate contact and its timing between the voicing counterparts (e.g. Liker & Gibbon, 2011, 2013, 2018; Fuchs et al., 2007), but also among the subjects Liker & Gibbon (2011). This variability cannot be analyzed in midsagittal tongue contours.

In this study, we conducted a pioneering work on investigating the articulation and acoustics of Hungarian alveolar fricatives. We demonstrated that the phonetic realisation of the voicing contrast in Hungarian /s/ and /z/ requires an appropriate laryngeal-oral coordination, as shown by previous studies, and is not merely the result of differences at the laryngeal level (e.g. Narayanan et al., 1995; Fuchs et al., 2007; Coretta, in press). We showed half of our twelve speakers showed some devoicing of phonologically voiced fricatives, while half of them showed no devoicing at all. We also replicated previous findings revealing an important role of tongue root displacement in the maintenance of voicing in phonologically voiced fricatives, but we also described further distinct articulatory strategies that may aid this articulatory/acoustic goal. Our results may contribute greatly to our knowledge on speaker-dependent phonetic variability of fricative voicing.

References

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

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Bárkányi, Zs., & Kiss, Z. (2009). Word-final fricative contrasts in Hungarian. A phonetic approach. URL: http://budling.nytud.hu/~cash/papers/buphoc09-slide.pdf Előadás a BuPhoC 2009. nov. 5-i ülésén.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Bickley, C. A., & Stevens, K. N. (1986). Effects of a vocal tract constriction on the glottal source: experimental and modeling studies. *Journal of Phonetics*, *14*, 373–382.

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

Coretta, S. (2019a). *Assessing mid-sagittal tongue contours in polar coordinates using generalised additive (mixed) models*. doi:10.31219/osf.io/q6vzb. pre-print page: https://www.researchgate.net/publication/335475997_Assessing_mid-sagittal_tongue_contours_in_polar_coordinates_using_generalised_additive_mixed_models.

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

Coretta, S. (in press). Longer vowel duration correlates with greater tongue root advancement at vowel offset: Acoustic and articulatory data from Italian and

Polish. *Journal of Acoustic Society of America*, . Preprint downloaded from `https://osf.io/zrqyx`.

Docherty, G. J. (1992). *The Timing of Voicing in British English Obstruents*. Berlin – New York: Foris Publications.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Fletcher, S. G. (1989). Palatometric specification of stop, affricate, and sibilant sounds. *Journal of Speech and Hearing Research*, *32*, 736–748.

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.

Fuchs, S., & Perrier, P. (2003). An EMMA/EPG study of voicing contrast correlates in German. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (p. 1057–1060). Barcelona.

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

Kent, R. D., & Moll, K. L. (1969). Vocal–tract characteristics of the stop cognates. *Journal of the Acoustical Society of America*, *46*, 1549–1555. doi:`10.1121/1.1911902`.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). LmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26.

Liker, M., & Gibbon, F. (2011). Groove width in Croatian voiced and voiceless postalveolar fricatives. In *Proceedings of the 18th International Congress of Phonetic Sciences. Hong Kong* (p. 1238–1241).

Liker, M., & Gibbon, F. (2013). Differences in EPG contact dynamics between voiced and voiceless lingual fricatives. *Journal of the International Phonetic Association*, *43*, 49–64. doi:10.1017/S0025100312000436.

Liker, M., & Gibbon, F. (2018). Tongue-Palate Contact Timing during /s/ and /z/ in English. *Phonetica*, *75*, 110–131.

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 S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (p. 2715–2719). Melbourne, Canberra: Australasian Speech Science and Technology Association Inc. URL: http://intro2psycholing.net/ICPhS/papers/ICPhS_2764.pdf.

McLeod, S., Roberts, A., & Sita, J. (2006). Tongue/palate contact for the production of /s/ and /z/. *Clinical Linguistics and Phonetics*, *20*, 51–66.

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.

Müller, E. M., & Brown Jr., W. S. (1980). Variations in the supraglottal air pressure waveform and their articulatory interpretation. *Speech and Language: Advances in Basic Research and Practice*, *4*, 317–389.

Narayanan, S. S., Alwan, A. A., & Haker, K. (1995). An articulatory study of fricative consonants using magnetic resonance imaging. *Journal of the Acoustical Society of America*, *98*, 1325–1347.

Ohala, J. J., & Solé, M. J. (2010). Turbulence and phonology. In S. Fuchs, M. Toda, & M. Zygis (Eds.), *Turbulent sounds: An interdisciplinary guide* (p. 37–102). Berlin & New York: De Gruyter Mouton.

Perkell, J. S. (1969). Physiology of speech production: Results and implication of quantitative cineradiographic study. *M.I.T. research monograph*, *53*.

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

Rothenberg, M. (1967). The breath-stream dynamics of simple-released-plosive production. 7186 (Basel: Biblioteca phonetica).

Shadle, C. H. (1991). The effect of geometry on source mechanisms in fricative consonants. *Journal of Phonetics*, *19*, 409–424.

Shih, C., Möbius, B., & Narasimhan, B. (1999). Contextual effects on consonant voicing profiles: A cross-linguistic study. In *Proceedings of the 14th International Congress of Phonetic Sciences* (p. 989–992). San Francisco, CA.

Smith, C. L. (1997). The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics*, *25*, 471–500.

Stevens, K. N. (1968). *The Quantal Nature of Speech: Evidence from Articulatory-acoustic Data*. Northwestern University.

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

Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., & Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of Acoustic Society of America*, *91*, 2979–3000.

Svirsky, M., Stevens, K. N., Matthies, M., Manzella, J., Perkell, J., & Wilhelms-Tricarico, R. (1997). Tongue surface displacement during bilabial stops. *Journal of the Acoustical Society of America*, *102*, 562–571.

Tabain, M. (2001). Variability in fricative production and spectra: Implications for the hyper- and hypo- and quantal theories of speech production. *Language and Speech*, *44*, 57–94.

van Rij, J., Wieling, M., Baaye, R., & 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. *The Journal of the Acoustical Society of America*, *73*, 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. *Journal of Phonetics*, *70*, 86–116.

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