

The articulatory and acoustic characteristics of Polish sibilants and their consequences for diachronic change

Véronique Bukmaier & Jonathan Harrington

Institute of Phonetics and Speech Processing, University of Munich, Germany

bukmaier@phonetik.uni-muenchen.de

jmh@phonetik.uni-muenchen.de

The study is concerned with the relative synchronic stability of three contrastive sibilant fricatives /s ʂ ʧ/ in Polish. Tongue movement data were collected from nine first-language Polish speakers producing symmetrical real and non-word CVCV sequences in three vowel contexts. A Gaussian model was used to classify the sibilants from spectral information in the noise and from formant frequencies at vowel onset. The physiological analysis showed an almost complete separation between /s ʂ ʧ/ on tongue-tip parameters. The acoustic analysis showed that the greater energy at higher frequencies distinguished /s/ in the fricative noise from the other two sibilant categories. The most salient information at vowel onset was for /ʧ/, which also had a strong palatalizing effect on the following vowel. Whereas either the noise or vowel onset was largely sufficient for the identification of /s ʧ/ respectively, both sets of cues were necessary to separate /ʂ/ from /s ʧ/. The greater synchronic instability of /ʂ/ may derive from its high articulatory complexity coupled with its comparatively low acoustic salience. The data also suggest that the relatively late stage of /ʂ/ acquisition by children may come about because of the weak acoustic information in the vowel for its distinction from /s/.

1 Introduction

While there have been many studies in the last 30 years on the acoustic (Evers, Reetz & Lahiri 1998, Jongman, Wayland & Wong 2000, Nowak 2006, Shadle 2006, Cheon & Anderson 2008, Maniwa, Jongman & Wade 2009), perceptual (McGuire 2007, Cheon & Anderson 2008, Li et al. 2011) and articulatory characteristics of sibilants (Narayanan, Alwan & Haker 1995), the large majority of these have been focused on the two-way distinction between alveolar /s/ and post-alveolar /ʂ/. Here we are concerned with the comparatively rarer three-way place contrast in sibilants in Polish. Apart from Swedish and Mandarin Chinese, Standard Polish is one of the very few languages that distinguishes lexically between dental /s/ (e.g. *sali* /sali/ ‘room, GEN’), retroflex /ʂ/ (e.g. *szali* /ʂali/ ‘scale, GEN’), and alveopalatal /ʧ/ (e.g. *siali* /ʧali/ ‘sown’) sibilants (Gussmann 2007, Żygis, Pape & Jesus 2012a).

In recent years, these three sibilants have been analysed physiologically for Polish in Toda, Maeda & Honda (2010), for Mandarin in Proctor et al. (2012), and in both these languages

by Hu (2008). These studies have shown that the three sibilants differ articulatorily not only in tongue position, but also in tongue posture. The fricatives are also distinguished from each other by two other tongue shape properties. Firstly, whereas in /ʃ s/ the vertical orientation of the tongue tip is typically upward-facing, it is downward-facing for /ç/. Secondly, while the tongue tip tends to be curled back to a greater extent for /ʃ/ than for /s/, the degree to which it is retracted has been shown to be somewhat less in Polish and Mandarin than in Indian languages (Hamann 2002a, b; Hu 2008; Toda et al. 2010): for these reasons, there is a greater resemblance in tongue shape between /ʃ ç/ in Polish than in Indian languages.

There is some evidence from the physiological analysis of four Polish L1 speakers in Bukmaier et al. (2014) for greater variability in /ʃ/ than in the other sibilants. At a slow speech rate, /ʃ ç/ were clearly differentiated in tongue-tip orientation such that /ʃ/ was a sub-laminal production in which the underside of the tongue tip/blade made contact with the place of articulation. However, at a fast speech rate, these orientation differences were much less in evidence such that /ʃ/ resembled /s/ in being supra- rather than sub-laminal. Hu's (2008) physiological analysis of Mandarin Chinese also pointed to a greater articulatory variability in /ʃ/ than in the other two fricatives.

As far as the acoustics are concerned, many studies in the last 50 years have shown that the place of articulation distinction between English /s ʃ/ can be based to a large extent on the spectral characteristics of the fricative noise (Whalen 1991, Shadle & Mair 1996, Evers et al. 1998, Stevens 1998, Jongman et al. 2000, Shadle 2012): more specifically, the shorter front cavity in /s/ causes the energy in the spectrum to be shifted towards higher frequencies, so that both acoustically and perceptually (Fujisaki & Kunisaki 1977, Mann & Repp 1980), a higher spectral centre of gravity (Forrest et al. 1988) differentiates it from /ʃ/. For the three-way place contrast in Polish, these spectral characteristics in the noise can separate /s/ from the other sibilants (Żygis et al. 2014a, b), but as various studies (Jassem 1995, Żygis & Hamann 2003, Nowak 2006) have shown, the centre of gravity in the noise by itself is generally insufficient for the /ʃ ç/ separation.

The issue of whether formant transitions contribute to the acoustic and perceptual distinction of place of articulation within fricatives is still unresolved. Some of the first studies to address this issue (Harris 1958, Heinz & Stevens 1961) showed that formant transitions were not necessary for the distinction between sibilants but that they were for the non-sibilant /f θ/ separation in English. On the other hand, although acoustic studies showed evidence of formant transitions extending well into the fricative noise (Soli 1981), subsequent research suggested that vowel transitions were perceptually less important for the perceptual distinction between place of articulation in sibilant and non-sibilant fricatives (LaRivière, Wintz & Herriman 1975, Jongman 1989). However, most of these studies were based on languages with only two sibilant fricatives. By contrast, a more recent cross-linguistic investigation by Wagner, Ernestus & Cutler (2006) showed that the effectiveness of formant transition cues was language-dependent: more specifically, listeners were shown to rely on formant transitions to a greater extent in languages like Polish which has fricatives such as /ʃ ç/ that are largely undifferentiated in the fricative noise. These results were consistent with those by Nowak (2006) in which L1 Polish listeners identified Polish sibilants from isolated sections of friction noise and in VCV sequences with the transitions into the following vowel removed. Nowak's (2006) results showed that, while fricatives could be reliably identified from the noise section, formant transitions were essential for the separation of /ʃ ç/ in VCV sequences. Compatibly, Toda et al. (2010) showed how the quite different tongue shapes for /ʃ ç/ contributed to the differences between these sibilants in vowel formant transitions.

Studies of the acquisition of Polish sibilants have shown that children acquire /ʃ/ relatively late and typically after the other sibilants have been acquired (Łukaszewicz 2006, 2007). The articulatory instability in /ʃ/ and the findings from language acquisition might also be related to the diachronic change of the three-way /ʃ ç/ to a two-way distinction as a result of an /s ʃ/ merger in both the Min variety of Mandarin (Duanmu 2006, Chuang & Fon 2010) and in several Polish dialects (Żygis, Pape & Czaplicki 2012b). One of the main motivations for



Figure 1 (Colour online) The placement of the two sensors on the surface of the tongue.

the present study was to investigate the synchronic basis for the diachronic collapse of the /s ʂ/ contrast towards /s/. The more specific aims were to analyse both the physiological and acoustic characteristics of these three fricatives in order to assess whether the identification of /ʂ/ is disadvantaged in comparison with the other two fricatives. In order to do so, we carried out an electromagnetic articulographic (henceforth EMA) study of nine Polish L1 speakers producing these sibilants and assessed the acoustic distinctiveness of the three fricatives from each other in both the noise and transitions.

2 Method

2.1 Data collection and speakers

Acoustic and speech movement data were acquired using electromagnetic articulography in a soundproof booth at the IPS in Munich (AG501, Carstens Medizintechnik) in order to obtain measurements of the horizontal, vertical, and lateral position of the articulators. For the EMA recordings, two sensors were placed on the tongue (Figure 1): one on the midline 1 cm behind the tip of the tongue (TT) and the other on a level with the molar teeth at the tongue back (TB). Additionally, two sensors were placed on the upper and lower lip, i.e. on the skin just above and below the lips. Four additional sensors were fixed to the maxilla (to the tissue just above the teeth), the nose bridge, as well as to the left and right mastoid bones in order to correct for head movement. For the present study, only the data from the sensor attached to the tongue tip were analysed. The acoustic speech signal was recorded synchronously with the physiological data using a Sennheiser ME66 supercardioid microphone with bass rolloff filter turned on (−6 dB at 200 Hz) positioned at a distance of approximately one metre in front of the subject. Audio data was recorded with a National Instruments Compact DAQ multichannel data acquisition front-end, with USB connection to a notebook computer. Synchronization of the audio and speech movement signal was carried out in the post-processing of the data after the recording session (see Hoole & Zierdt 2010 for further details of the post-processing of acoustic and articulatory data).

The subjects in this experiment were nine L1 Standard Polish speaking adults spanning an age range between 19 and 28 years and included four male and five female speakers. Six speakers were born and went to school in dialectal regions with a three-way sibilant contrast (two each from Silesia, Lesser Poland and Greater Poland). The remaining three speakers

Table 1 Distribution of CV sequences in real words and non-words used as target words.

	Word 1	Word 2	Word 3	Non-word
/sa/	sara (/sara/ forename, FEM)	sama (/sama/ alone)	sawa (/sava/ forename, M)	sasa
/sɛ/	serce (/sertsɛ/ heart)	–	–	sese
/sɔ/	sowa (/sɔva/ owl)	–	–	soso
/ʂa/	szary (/ʂari/ grey)	szafa (/ʂafa/ cupboard)	–	szasza
/ʂɛ/	szereg (/ʂerek/ line)	–	–	szesze
/ʂɔ/	szorstka (/ʂorstka/ rough)	–	–	szoszo
/ɕa/	siatka (/ɕatka/ net)	siano (/ɕano/ hay)	–	siasia
/ɕɛ/	siekacz (/ɕekacɕ/ incisor)	–	–	siesie
/ɕɔ/	siodło (/ɕɔdɔ/ saddle)	–	–	siosio

were born and lived most of their lives (i.e. went to school) in dialect regions in which the alveolar/retroflex contrast is neutralized (two from Mazovia and one from Kashubia). These three subjects were nevertheless included in our analysis because they were judged by an L1-Polish speaker with linguistics training to be speakers of Standard Polish with no perceptible regional colouring. None of the participants had lived outside of Poland for more than two years at the time of recording.

2.2 Speech material and experimental set-up

The participants produced symmetrical CVCV (e.g. /sɛsɛ/) non-words (in which C=/s, ʂ ɕ/ and V=/a ɛ ɔ/) as well as Polish disyllabic real words (Table 1) with initial CV sequences (in which C=/s ʂ ɕ/ and V=/a ɛ ɔ/). All target words were embedded in the carrier phrase ‘Ania woła [TARGET WORD] aktualnie’ (literally ‘Ania shouts [TARGET WORD] currently’), where the target word was produced with a nuclear pitch accent. The participants read the sentences aloud as they were automatically presented to them on a computer screen one at a time in randomized order. In cases of mispronunciations and productions of incorrect prosody, the participants were asked to repeat the sentence.

The recording session consisted of ten blocks alternating between slow and fast speech rates. In order to define individual speech rates as well as to adjust the corresponding recording time, participants were asked to read examples of the speech material at a self-selected fast and slow speech rate in a pretest prior to the actual recording. The display incorporated a progress bar linked to the desired speech rate that was defined for each speaker and condition based on the mean durations of the pre-recording and that indicated the time frame for each token. For each speech rate, each of the 22 target words containing nine non-words and 13 real words (Table 1) was repeated ten times in randomized order. Some word initial CV sequences occur more often in the onsets of Polish disyllabic real words e.g. /sa-/, /ɕa/- and /ʂa/-word onsets, as a result of which there were more (near) minimal pairs for these CV sequences (see Table 1, row 1: /sara/, /sama/, /sava/; row 4: /ʂari/, /ʂafa/; row 7: /ɕatka/, /ɕano/). Because of this skewed distribution of CV sequences, the materials for this study included between two (/ɕa/ and /ʂa/ onsets) and three (/sa/ onsets) target words, with other, rarer sequences only being represented with one target word (e.g. /sɛ sɔ ʂɛ ʂɔ ɕɛ ɕɔ/). Table 1 contains the complete distribution of CV sequences.

The experiment contained 3960 (22 target words × 10 repetitions per speech rate × 9 speakers) sentences, of which 3895 sentences were analysed in this study. The data loss of 65 tokens was due to technical problems during the recording session and post-processing. The total number of analysed sibilant–vowel combinations for both real and non-words is given in Table 2.

Table 2 Total number of analysed CV sequences in real and non-words separately for each sibilant–vowel combination.

	/a/	/ɛ/	/ɔ/
/s/	708	354	358
/ʃ/	530	348	354
/ç/	534	355	354

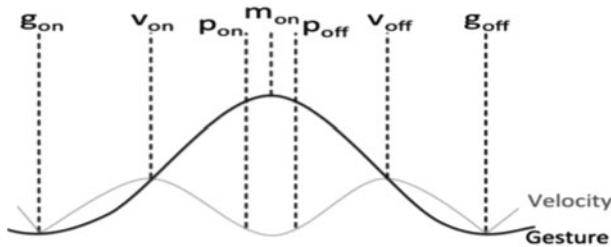


Figure 2 Schematic representation of landmark positions: gestural onset (g_{on}), maximum velocity at gestural onset (v_{on}), onset of constriction plateau (p_{on}), maximum in constriction (m_{on}), offset of constriction plateau (p_{off}), maximum velocity in gestural offset (v_{off}), and gestural offset (g_{off}).

2.3 Data analysis

2.3.1 Physiological analysis

The post-processing of the physiological raw data was done semi-automatically in Matlab (version MathWorks MATLAB R2012a) including rotation of the data so that they were parallel to the occlusal plane (Hoole & Zierdt 2010).

The articulatory annotation of the three sibilants was based on the vertical movement of the tongue tip (TT) and the TT tangential velocity. Physiological labels included seven different landmarks (Figure 2). Typically, a complete CVC movement cycle was divided into a CV or opening phase, a nucleus, or quasi target phase, and a VC or closing phase. Onsets and offsets of opening and closing gestures were determined by using a 20% threshold criterion of the tangential velocity signal (Hoole & Mooshammer 2002, Hoole et al. 2009). The vowel nucleus was then defined as the interval between CV offset and VC onset.

Using the landmarks in Figure 2, we extracted data from the vertical and horizontal positions of the tongue tip (TT) and the tongue dorsum (TD). We also analysed the orientation of the TT since this potentially provided information about the retroflex, in which the tongue tip is often known to be curled upwards (Ladefoged 2001; Hamann 2002a, b; Toda et al. 2010; Bukmaier et al. 2014).

2.3.2 Acoustic analysis

The synchronized acoustic data was digitized at 25,600 Hz and automatically segmented and labelled using forced alignment (Munich Automatic Segmentation tool; Schiel 2004). Calculations were made of spectra (256 point discrete Fourier transform with a 40 Hz frequency resolution, 5 ms Blackman window, and a frame shift of 5 ms) and of formant frequencies (F1–F4; pre-emphasis of -0.8 , 20 ms Blackman window with a frame shift of 5 ms).

For the acoustic analysis of the fricative noise, spectra were extracted at the temporal midpoint between the acoustic onset and offset of each sibilant. These spectral data were reduced to a set of coefficients using the discrete cosine transformation (DCT) after converting

the Hz to the mel scale. For an N -point mel-scaled spectrum, $x(n)$, extending in frequency from $n = 0$ to $N-1$ points over the frequency range of 500–3500 mel (414–10 313 Hz), the m th DCT-coefficient C_m ($m = 0, 1, 2$) was calculated with the following equation:

$$C_m = \frac{2k_m}{N} \sum_{n=0}^{N-1} x(n) \cos\left(\frac{(2n+1)m\pi}{2N}\right)$$

These three coefficients C_m ($m = 0, 1, 2$) encode the mean, the slope, and curvature, respectively, of the signal to which the DCT transformation was applied (Harrington 2010). Since preliminary studies of these had shown that the sibilants were optimally distinguished in the fricative noise from C_1 and C_2 (i.e. from the slope and curvature of the spectrum respectively), all further quantifications of the sibilants were based on these coefficients.

The articulatory and formant data were speaker-normalized using standard normalization (Lobanov 1971). More specifically, where $x_{P,i,T}$ is a raw value of an articulatory or formant parameter P from speaker i at time point T , the corresponding normalized value $X_{P,i,T}$ was given by the following formula:

$$X_{P,i,T} = (x_{P,i,T} - x_{P,i,m}) / x_{P,i,s}$$

where $x_{P,i,m}$ and $x_{P,i,s}$ are the mean and standard deviation respectively,

calculated across all frames of the same parameter for the same speaker

When normalization was applied to the data in the fricative noise, ($x_{P,i,m}$, $x_{P,i,s}$) were calculated from all frames of data between the fricatives' acoustic onset and offset; when normalization was applied to the formant parameters, ($x_{P,i,m}$, $x_{P,i,s}$) were calculated from all frames extending between the acoustic vowel onset and offset.

We also carried out a Gaussian classification of the acoustic (spectral, formant) data in order to determine the degree of separation between the three fricative places of articulation. Classification was based on quadratic discriminant analysis (Srivastava, Jermyn & Joshi 2007) in which there was a training and a testing stage. During the training stage, each fricative class consisting of a number of observations in a two-dimensional acoustic space was modelled as a bivariate Gaussian distribution; in the testing phase, observations were classified as one of the fricative classes based on the greatest posterior probability. The relationship between training and testing was accomplished using the leave-one-out procedure in which, iteratively for each of the nine speakers in turn, a given speaker's data were classified following training on the data of the other eight speakers. For the fricative noise, the two parameters were C_1 and C_2 as defined above extracted at the acoustic temporal midpoint of the fricative; for the vowel, the two parameters were F2 and F3 at the acoustic vowel onset. In vowel classifications, training and testing were additionally carried out using this leave-one-out procedure separately in each of the three /a ε ə/ vowel contexts. The classifications as described above were separately accomplished in the slow and fast rate contexts (thus four classifications: two (slow/fast) based on C_1 and C_2 and two (slow/fast) on F2 and F3 at the acoustic vowel onset).

3 Results

The results are presented below separately for the fricative noise (Section 3.1) and for vowel transitions (Section 3.2). In both cases, the aim was to determine the extent to which there was separation between the three fricative places of articulation and to assess how far these two sets of cues provide complementary information for this purpose.¹

¹ We also tested whether there was any effect of lexical frequency on the classification scores. Lexical frequencies for the real words were obtained using SUBTLEX-PL, a freely available database of Polish

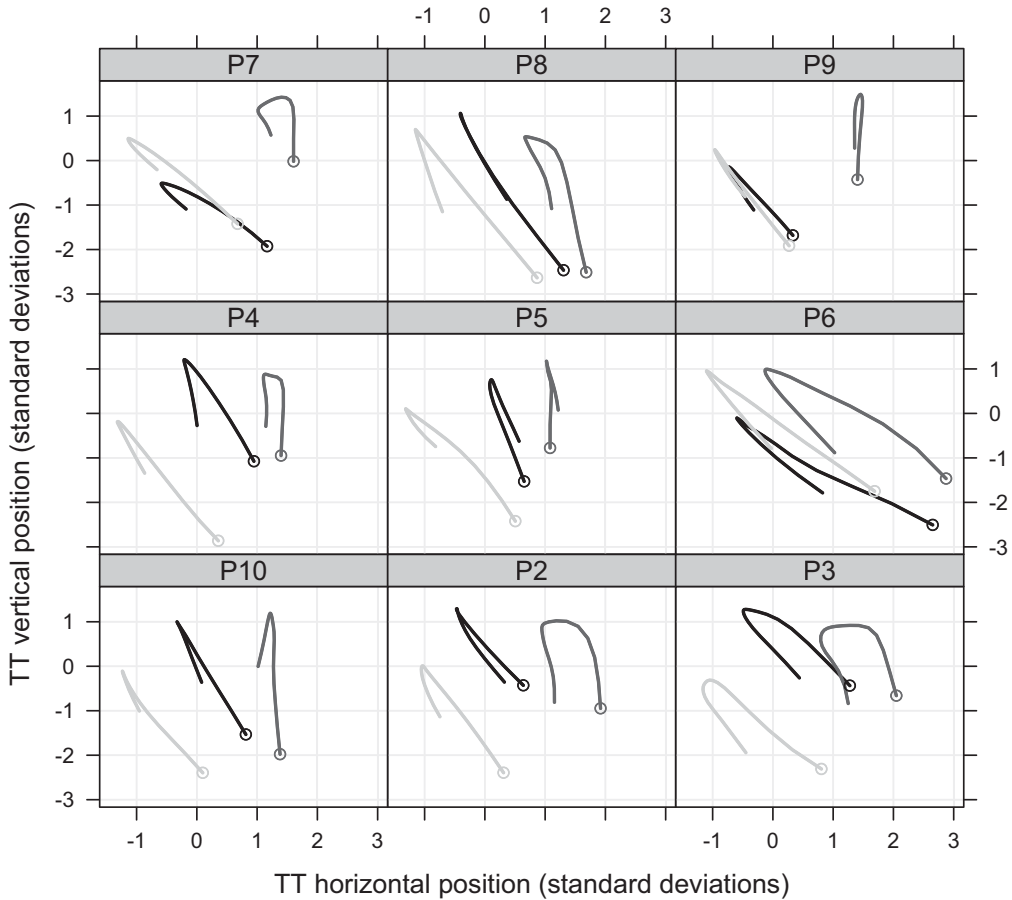


Figure 3 Lobanov-normalized vertical (y -axis) and horizontal (x -axis) TT trajectories averaged across vowels between the two velocity maxima separately for the dental (grey), retroflex (dark grey) and alveopalatal (black) fricative (with the circle marking the starting point of the trajectories).

3.1 Frication

3.1.1 Physiological analysis

The aggregated tongue-tip data in [Figure 3](#) shows a clear separation between the fricatives for each of the nine speakers. For most subjects, /s ʃ/ had the most fronted and retracted positions respectively, with /ç/ located along the front–back dimension between the other two sibilants. Additionally, the tongue tip was generally lower for /s/ than for the other two fricatives; and

word frequencies with 101 million words from movie subtitles (Mandera et al. 2014). The database includes word frequencies transformed to the Zipf scale, a logarithmic scale of frequency per billion words (for further details see also van Heuven et al. 2014). We ran a mixed model with the binary response correctly or incorrectly classified consonant from the spectral data (Table 3) as the dependent variable, with place of articulation (three levels: /s ʃ ç/) and lexical frequency as fixed factors and with the speaker (9 levels) and word (22 levels) as random factors. We also ran another mixed model with classification scores from the vowel transitions (Table 4) as the dependent variable. There was no influence of lexical frequency on classification scores in either case.

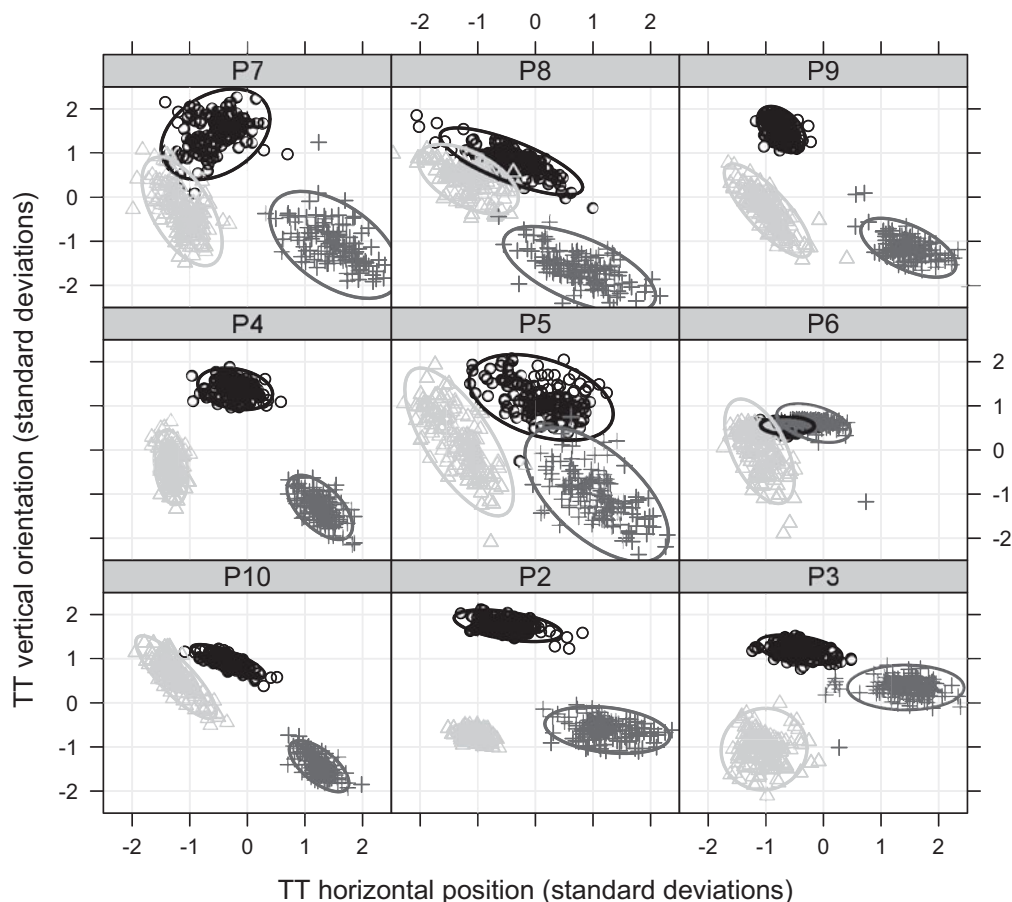


Figure 4 Lobanov-normalized TT orientation (y -axis) and horizontal TT position (x -axis) shown separately for each speaker at the moment of the maximum constriction (m_{on} in Figure 1) and separately for /s/ (grey triangle), /ʃ/ (dark grey cross) and /ç/ (black circle). The confidence ellipses are 2.47 standard deviations and enclose 95% of the data points.

/ʃ/ tended to reach the highest position, perhaps as the tongue tip unfolded from an initially curled position.

Subsequent analyses showed that various combinations of two physiological parameters provided a very clear separation between the three fricative places of articulation. One of the most effective of these was for the combination of the horizontal position of the tongue tip and its vertical orientation (Figure 4). Recall that the latter provides information about the sensor's rotation along the front–back axis. Since the tongue tip can be expected to be curled back in /ʃ/, but not in /ç/, then the sensor which is affixed just behind the tongue tip should be rotated for /ʃ/ about the axis that is perpendicular to the sagittal plane – or at least to a greater extent than it is in /ç/. This, as Figure 4 shows, was the case for eight out of nine speakers, in which the rotation was greater for /ʃ/ than for /ç/: note in particular that this is the distinguishing feature for two speakers (P5, P8) for whom /ʃ ç/ were otherwise undifferentiated as far the horizontal position of the TT was concerned. Figure 4 also shows that, with the exception of P6, there was almost complete separation between the three fricatives on these two dimensions for the remaining speakers. Thus the general conclusion is that /s ʃ ç/ were separated from each other as far as tongue-tip posture is concerned.

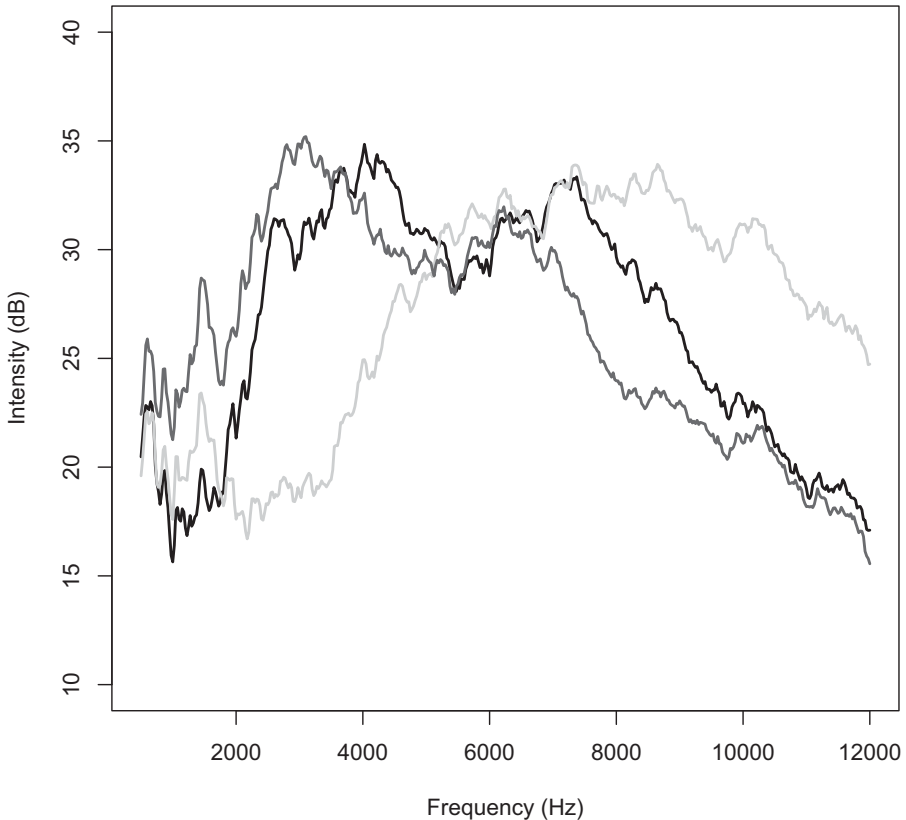


Figure 5 Ensemble-averaged fricative spectra at the fricatives' temporal midpoints separately for the dental (grey), retroflex (dark grey) and alveopalatal (black) sibilants. The averaging was done across all fricatives produced by all speakers.

3.1.2 Acoustic analysis

We now consider the extent to which the clear physiological separation between the three fricatives was matched acoustically. The ensemble-averaged spectra in [Figure 5](#) show that /s/ was separated from the other two fricatives by greater energy at higher frequencies, but that the ensemble-averaged spectral shapes for /ʃ ç/ were quite similar (see Appendix). We tested various combinations of spectral parameters at the fricatives' temporal midpoint including spectral moments (Forrest et al. 1988). The two which were most effective in separating the places of articulation were those that are proportional to the linear slope (C_1) and curvature (C_2) derived from the discrete cosine transformation, calculated after transforming the frequency axis to the mel scale as described in 2.3. For C_1 , if a regression line were drawn through the three spectra, then, as [Figure 5](#) suggests, /s/ would be differentiated from the other two by its rising as opposed to falling slope. For C_2 , the greater the resemblance of the ensemble-averaged spectrum to a parabolic shape, then the greater the values on C_2 . There is a clear parabolic shape in evidence for the ensemble-averaged /s/ spectrum in [Figure 5](#), and the generally higher amplitude levels over a mid-frequency range for /ç/ than for /ʃ/ may provide some basis for their differentiation on this parameter.

For most speakers, [Figure 6](#) shows an overlap of /ʃ ç/ in the $C_1 \times C_2$ space, whereas /s/ was clearly separated from the other two sibilants (except for speaker P6). The data in [Figure 6](#) were consistent with the classifications (see [Section 2.2](#) above for details) which showed for the slow rate of speech ([Table 3](#)) 96% correct classification for /s/ as opposed to

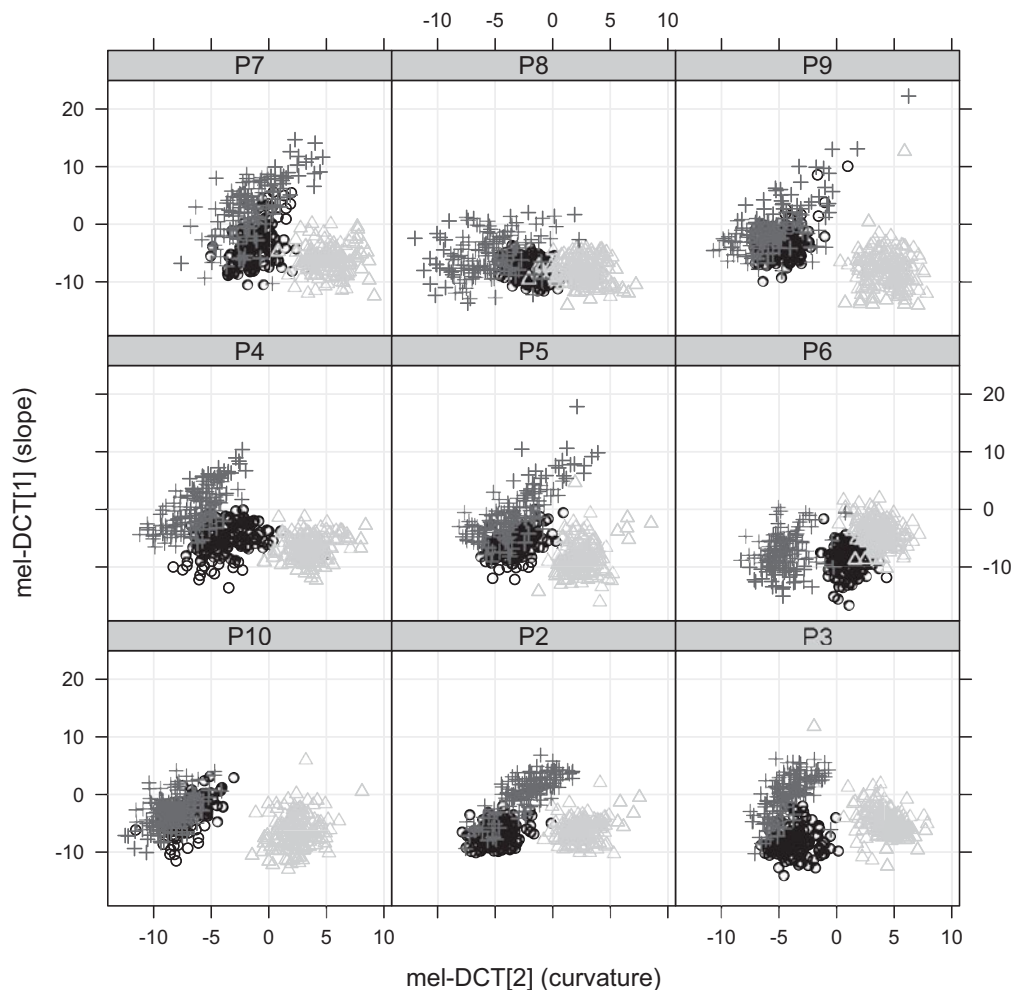


Figure 6 First (slope) and second (curvature) mel-scaled DCT coefficients separately for the dental (grey triangle), retroflex (dark grey cross) and the alveopalatal (black circle) sibilants averaged across vowel contexts and speech rates.

77% and 63%, respectively, for /ʂ ʐ/. **Table 3** shows a high degree of /ʂ ʐ/ confusion for the slow rate of speech with 25% of /ʐ/ being misclassified as /ʂ/ and 23% of /ʂ/ misclassified as /ʐ/. **Table 3** also shows that the classification scores at the fast rate of speech showed a broadly similar pattern.

We tested the influence of place of articulation and rate on classification scores. We also tested the influence of whether or not the sibilant had occurred in a real or non-word. For this purpose, we ran a mixed model with the binary response correctly or incorrectly classified consonant as the dependent variable, with fixed factors that included place of articulation (three levels: /s ʂ ʐ/), word-type (two levels: real word/non-word), and rate (two levels: slow/fast); and with the speaker (nine levels) and word (22 levels: the separate words and non-words in **Table 1**) as random factors. We also included all the interaction terms between the fixed factors in the model. We assessed the influence of word-type and rate by comparing two models: one with all the factors included, as outlined above; and one that differed from this by dropping word-type and rate. A comparison of these two models (one full with another

Table 3 Results for correct classification of the three sibilants with Gaussian training/testing in the $C_1 \times C_2$ space based on the leave-one-out method i.e. total of nine speakers ($k = 1, 2 \dots 9$), test on speaker k , train on all other eight speakers. The results are shown for classifications from the slow speech rate and from the fast rate in parentheses.

	/ç/	/s/	/ʒ/
/ç/	.63 (.61)	.12 (.11)	.25 (.28)
/s/	.03 (.04)	.96 (.95)	.01 (.01)
/ʒ/	.23 (.23)	.00 (.00)	.77 (.77)

without word-type and rate) showed no significant differences: thus neither word-type nor rate had any significant influence on classification scores. Predictably, classification scores were significantly influenced by place of articulation ($\chi^2 = 67.2, p < .001$).

3.2 Coarticulatory effects on adjacent vowels

In the preceding section, we showed that the very clear separation between the three fricatives based on the tongue configuration was not matched by the acoustic analysis of the fricative noise, which showed a substantial /ʒ ç/ confusion. Here we apply a similar type of analysis to the onset of the transitions into the vowel.

3.2.1 Physiological analysis

With the exception of speaker P4 (for whom the TB trajectories of dental and alveolopalatal were quite similar), Figure 7 shows that the vertical TB position was higher in vowels following /ç/ (indicated by higher vertical TB values), while in vowels following /s ʒ/ the vertical TB position was lower (indicated by lower vertical TB values). These findings suggest that /ç/ exerted a strong coarticulatory influence on the following vowel. Figure 7 also shows that, with the exception of speaker P6, /ʒ/ had a more retracted tongue body position compared to /s/. Thus, there is considerable information in the tongue dorsum at the vowel onset and often throughout the vowel for the distinction between the three fricatives.

3.2.2 Acoustic analysis

For all speakers, the F2 transition data in Figure 8 shows higher F2 values following /ç/, consistent with the observations of the physiological analysis in Figure 7. Although /ʒ ç/ overlapped in F2, they were separated to a certain extent by the lower F3 for /ʒ/.

Figure 9 illustrates further the strong coarticulatory influence of /ç/ on the vowels causing marked F2 raising for all vowels and F1 lowering in an /ε/ context. Thus, these data provide further evidence that vowels in a /ç/ context are strongly palatalized.

The results of the leave-one-out classification based on a two-parameter model of F2 onset and F3 onset show a high classification score of 91% at the slow rate for /ç/ with equal confusion between the other two fricatives on these parameters (Table 4). Although the identification rates of /s ʒ/ at the slow rate (82%, 71% respectively) were well above chance level (33%), there was also marked confusion between them (26% /ʒ/ misclassified as /s/ and 13% of /s/ misclassified as /ʒ/). Table 4 also shows a similar pattern of classification scores for the fast speech condition. A mixed model with the binary response correct/incorrect classification score based on these classifications from the combined F2 and F3 onset and with the same fixed and random factors as deployed earlier (Section 3.1.1) showed no significant effects for either rate or for word-type. Thus once again, the classification scores were unaffected by rate or word-type (whether or not the word was a real word or a non-word). Predictably, the classification scores were significantly influenced by consonant place of articulation ($\chi^2 = 23.6, p < .001$).

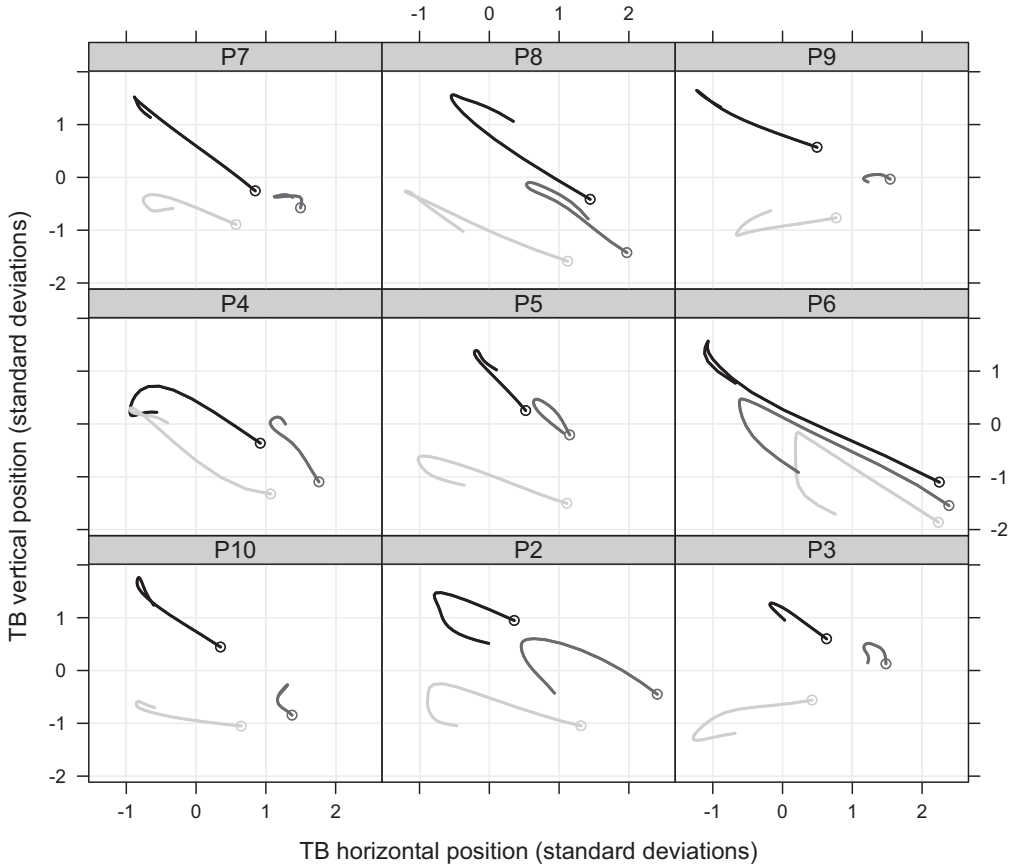


Figure 7 Trajectories of the vertical and horizontal TB (tongue body) position between the acoustic vowel onset and offset following the dental (grey), retroflex (dark grey) and the alveopalatal (black) sibilants separately for each speaker.

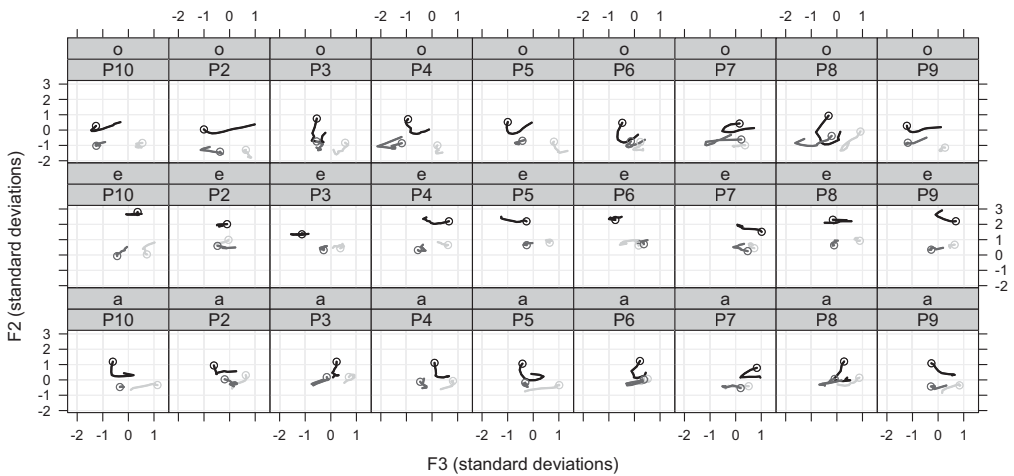


Figure 8 Aggregated, time-normalized trajectories in an F2 (y-axis) × F3 (x-axis) space between the acoustic vowel onset and its temporal midpoint shown separately for each vowel following the three sibilants: dental (grey), retroflex (dark grey) and alveopalatal (black). The averaging was carried out following Lobanov-normalization of the speaker data.

Table 4 Results for correct classification of the three sibilants with Gaussian training/testing based on the F2 and F3 values of the three values using the leave-one-out method i.e. total of nine speakers ($k = 1, 2 \dots 9$), test on speaker k , train on all other eight speakers. The results are shown for classifications from the slow speech rate and from the fast rate in parentheses

	/ç/	/s/	/ʂ/
/ç/	.91 (.88)	.04 (.06)	.05 (.06)
/s/	.05 (.03)	.82 (.83)	.13 (.14)
/ʂ/	.04 (.05)	.26 (.24)	.71 (.71)

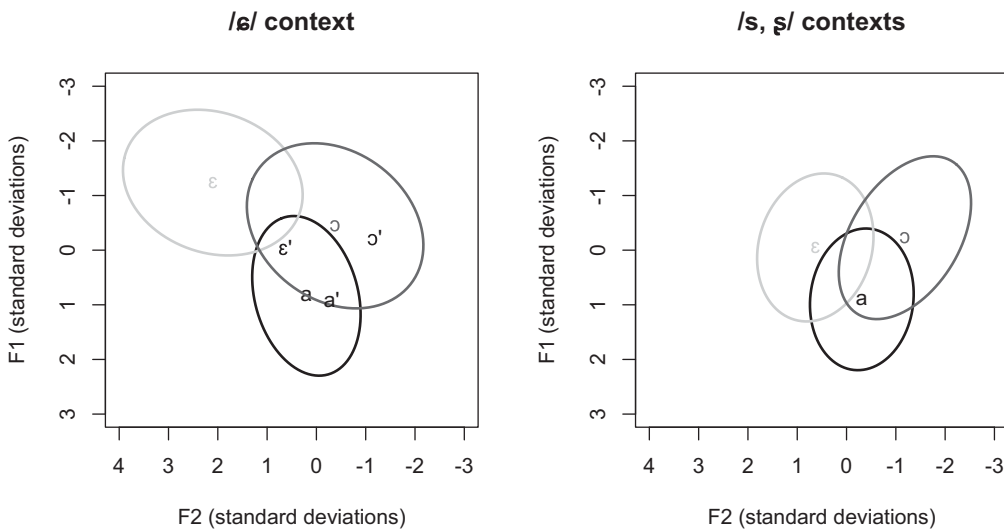


Figure 9 F1 (y -axis) and F2 (x -axis) extracted at the vowels' temporal midpoints in /ç/ (left panel) and in /s ʂ/ (right panel) contexts for /ε/ (grey), /a/ (black) and /ɔ/ (dark grey). The black vowel symbols (a^* \epsilon^* $\text{\textcircled{O}}^*$) in the left panel are the centroids extracted from the ellipses in the right panel, i.e. the centroids of the formants in vowels following /s ʂ/. The ellipses are 2.47 standard deviations and include at least 95% of tokens.

4 General discussion

The main aim of the present study has been to shed light on the acoustic and articulatory characteristics of the three Polish sibilants /s ʂ ç/ and to test whether the greater phonetic instability in /ʂ/ may be the source of the reduction of the three-way contrast to a two-way distinction that has been observed in certain Polish varieties and in Mandarin Chinese. We begin by considering the degree to which the fricatives were separated in the noise and transitions in turn.

Earlier studies have generally reported a very high separation between /s ʂ ç/ when listeners are presented with noise sections alone (Nowak 2006). Our physiological data for nine speakers shows quite unequivocally that /s ʂ ç/ were all distinguished on the basis of the position and configuration of the tongue. In particular, /s/ was (predictably) shown to have a very forward tongue-tip constriction, and it was most retracted for /ʂ/: this result is consistent with a physiological analysis of fricatives in Mandarin Chinese by Hu (2008) and Proctor et al. (2012), who showed a more retracted position for /ʂ/ than for /s ç/. The tongue-tip retraction in our data came about because the tip was (as is typical for retroflex consonants) curled back towards the hard palate. This posture was also the main characteristic that differentiated

it from /ç/; that is, /ç ʂ/ differed according to the rotation of the tongue tip about the axis that is perpendicular to the sagittal plane. Just these two parameterizations of the tongue tip (horizontal position, rotation) were sufficient for an almost complete separation between /s ʂ ç/. Thus the very high perceptual distinction between these fricatives based on noise found by Nowak (2006) is likely to be related to their marked physiological differences in the tongue position and orientation found in our study.

Our acoustic analysis of the fricative noise was consistent with that of Nowak (2006) and others (e.g. Jassem 1995) in showing a very clear separation between /s/ and the other two categories based on the greater concentration of energy at higher frequencies. According to Halle & Stevens (1997), theoretical considerations of vocal tract modeling suggest that energy typically found in the region associated with the second formant frequency should be lower for /ʂ/ than for /ç/; thus, /ç/ has a much narrower palatal constriction that suppresses back cavity resonances leading to an energy increase in the spectral region close to F2. In general, such a difference should result in a slightly greater weighting of spectral energy towards the lower frequency values for /ʂ/ than for /ç/. This is exactly what is evident from the ensemble-averaged spectra in Figure 5 above, which show a spectral peak in the vicinity of 2 kHz (i.e. in the region of F2) for /ʂ/ which is absent for /ç/. These observed differences are consistent with the findings by Li, Edwards & Beckman (2007), who found that energy in this F2 region of the noise spectrum effectively distinguished between /ʂ ç/ in Mandarin Chinese. Beyond these differences, and consistently with Nowak (2006), our study shows very similar spectral shapes for /ʂ ç/: that is, /ʂ ç/ differed principally in that a similar spectral shape occurred at slightly lower frequencies for the retroflex. Compatibly, Žygis & Hamann (2003) showed that a lower spectral centre of gravity of the noise separated /ʂ/ from /ç / for a female speaker, although not in their male speaker. In the semi-open classification test in which we trained and tested the three fricative categorizations based on a DCT parameterization, although their classification rates were well above chance, around 25% of /ʂ ç/ were nevertheless confused with each other. This result suggests that, in spite of the very clear physiological distinction, the acoustics of the fricative noise alone are unlikely to provide sufficient information in more casual, spontaneous speech for their separation.

Numerous studies in the last 50 years have shown that formant transitions provide contributory information to fricatives' place of articulation distinctions. This was shown to be especially so for the non-sibilants /f θ/ in English (Harris 1958). However, other studies have shown that formant transitions into the following vowel can also be important for the /s ʃ/ separation (Delattre, Liberman & Cooper 1962, Soli 1981, Whalen 1991, Lisker 2001, Gordon, Barthmaier & Sands 2002, Wagner et al. 2006, Li et al. 2007; see also Wagner et al. 2006 for a comprehensive review). Drawing on an analysis of Shona fricatives, Bladon, Clark & Mickey (1987) were among the first to suggest that formant transitions may be critical in languages with a three-way place of articulation contrast in sibilants. The results from our study show that the second formant transition provided especially salient information for identifying /ç/. In agreement with Nowak (2006) and Sawicka (1995), our results also show that /ç/ exerted a strong coarticulatory influence on adjacent vowels causing them to be palatalized: in particular, our physiological data showed a raised tongue-dorsum position at vowel onset extending well into the vowel for /ç/ for all speakers and contexts (Figure 4 above) and a concomitant raised F2 throughout the first half of the vowel once again in all speakers and contexts. This finding is also consistent with a perceptual study by Lisker (2001) who showed that English listeners were able to separate /ʂ ç/ quite reliably only on the basis of acoustic information in the vowel. A further new finding from our study is that it is not just F2 but also F3 that may contribute to this distinction. The acoustic theory of speech production predicts that retroflex consonants should be associated with F3 lowering (Fant 1960) and the results from our study show that F3 of /ʂ/ is lower than for the other two fricatives. F3 lowering for /ʂ/ was also found in the acoustic analysis of the Toda language by Gordon et al. (2002). The main result from our semi-open categorizations based on a two-dimensional space of F2 and F3 at vowel onset was that classification scores were well above chance and that almost

90% of /ç/ could be identified from this information in the vowel. While the classification scores for /ʃ/ are high at just over 70%, the same data also show that there is substantial /s/ ʃ/ confusion such that 12% of /s/ were misclassified as /ʃ/ and 26% of /ʃ/ as /s/. However, this confusion would presumably be resolved in combination with the fricative noise, which according to our analyses enabled an almost 95% separation of /s/ from the other two fricative categories. Overall then, the general finding from this study is that the fricative noise provides positive information for the separation of /s/ from /ʃ ç/ and that transitions distinguish /ç/ from /s ʃ/; therefore, the successful identification of /ʃ/ from acoustic data must depend on information both in the noise (to separate it from /s/) and on information in the vowel (to separate it from /ç/). Our study shows that /s/ can be distinguished from the other two fricative categories with reference to information in the noise alone (if the energy in the spectrum of the noise is concentrated in the upper part of the spectrum) and /ç/ can be separated from the other two fricative categories using information in the vowel (if F2 at vowel onset is high). But on the other hand, /ʃ/ requires for its identification two sets of cues to separate it from the other two fricative categories: both in the noise (the energy must be concentrated in the lower part of the spectrum to distinguish it from /ç/) and in the vowel (F2 and F3 must be low to distinguish it from /s/).

Our study also showed in contrast to an earlier analysis in Bukmaier et al. (2014) that rate had no effect either on the fricative noise nor on the vowel transitions. This was not because the speakers did not vary speaking rate: for every one of the speakers, the duration of both the fricative noise and of the vowel was less at the fast than at the slow tempo. We currently have no explanation for the divergent findings between the present study and that of Bukmaier et al. (2014) but can tentatively conclude that rate effects are unlikely to be a synchronic factor involved in diachronic /ʃ/ attrition.

Finally, we consider the issue of whether the retroflex consonant is likely to be the most unstable of the three categories both from a synchronic and diachronic perspective. The instability of the retroflex has been suggested by both Duanmu (2006) and Nowak (2006) independently of this study, who point to the likelihood of the collapse of a three-way to a two-way contrast in many varieties of Mandarin and Polish, typically because of a merger of the dental and retroflex consonants. Similarly, the Taiwan variety of Mandarin lacks the three-way sibilant contrast found in standard Mandarin because the retroflex is frequently substituted by the dental fricative (Chuang & Fon 2010) under the influence of Min (which lacks retroflex consonants). Their study also showed that under prosodic prominence speakers typically only enhanced one of the two /s ʃ/ fricatives, rather than both, and in most cases the enhancement was in /s/. As Ladefoged & Bhaskararao (1983) point out, it is the complexity of gestures involved in the production of retroflex consonants which may explain not only the type of diachronic changes noted above, but also why they are typologically rare, occurring only in languages with large coronal inventories (i.e. there is no known language that has retroflex consonants as the only coronal). To this we would add that it is perhaps not just the articulatory complexity but also the non-linear relationship to acoustics that may make /ʃ/ unstable: that is, whereas in our study retroflex consonants were unambiguously separated from the other two categories on the basis of tongue position, they remained highly confusable with both /ç/ in the fricative noise and with /s/ in the vowel. Thus /ʃ/ may be an example of what in Lindblom's (1998) model is considered to be a high-cost articulation involving complex articulatory maneuvers that nevertheless effect only a limited degree of acoustic or perceptual salience in relation to other fricative categories with which they contrast.

Studies and analyses of child language acquisition also point to the relative instability of /ʃ/. Some studies of Polish have suggested either that /ʃ/ is only acquired after /s ç/ (Łukaszewicz 2006) and/or that the contrast between dental and retroflex places of articulation emerges quite late (Łobacz 1996). Moreover, Nittrouer & Studdert-Kennedy (1987) and Nittrouer (1992, 2002) provide evidence that children rely much more than adults on dynamic than static information for phonetic categorization: for example, young children make far greater use of vowel transitions than the noise for fricative categorization. With increasing age,

this relationship changes so that they progressively take greater advantage of the information that is available in the fricative noise. In terms of the present Polish data, such a model predicts that the /s ʂ/ distinction would be most vulnerable and prone to confusion: this is both because there is, as the present study shows, insufficient information for the clear separation of the /s ʂ/ distinction based on vowel information and because children would be, according to Nitttrouer's model, less able to take advantage of the critical cues for the separation of /ʂ/ from other fricatives in the comparatively much more static noise section. Further empirical analyses of these fricatives need to be conducted in order to test whether the diachronic instability of /ʂ/ has its origins in the greater confusion of the production and perception of /s ʂ/ by children that is predicted by the results of the present study.

Acknowledgements

This research was supported by ERC grant number 295573 'Sound change and the acquisition of speech'. We are grateful to three anonymous *JIPA* reviewers for their comments on an earlier version of this paper.

Appendix. Calculating spectra using the multitaper methodology

It has been suggested by a reviewer that ensemble-averaging requires special conditions following a multitaper methodology (Jesus & Shadle 2002; Shadle 2006, 2010). We tested whether this was so by calculating spectra using a multitaper approach in Matlab (Percival & Walden 1993). We used the default settings with a value of four for the time-bandwidth product (which in turn determines the number of tapers used); the multitaper analysis also used Thomson's (1982) adaptive nonlinear method for combining the individual spectral estimates. As Figure A1 shows, we obtained almost identical results using our approach (Figure 5 in the main text above) and the multitaper method. Our results are therefore consistent with those in Reidy (2015) showing that, while the multitaper approach may be beneficial to estimating peaks and troughs in the spectrum, its use makes very little difference to parameterizations such as spectral moments (or the types of DCT coefficients we have used here) that are based on the sum of amplitude estimates.

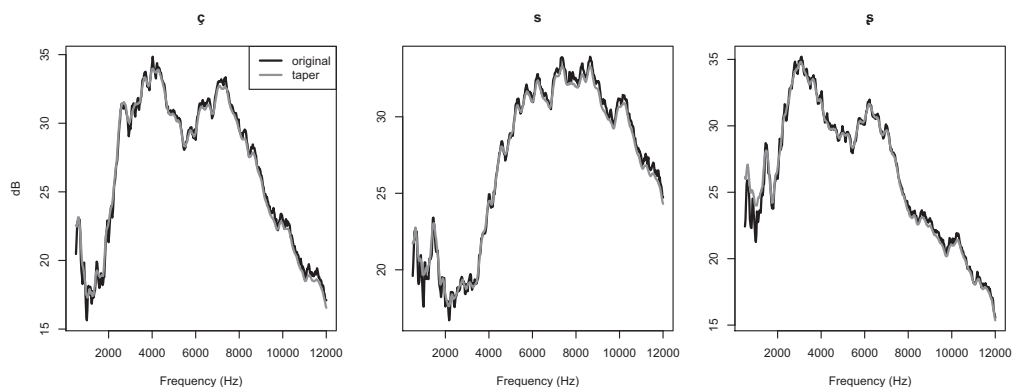


Figure A1 Ensemble-averaged spectra calculated using the methodology in Figure 5 (original) and using the multitaper approach (taper) shown separately for the three fricative categories.

References

- Bladon, Anthony, Christopher Clark & Katrina Mickey. 1987. Production and perception of sibilant fricatives: Shona data. *Journal of the International Phonetic Association* 17, 39–65.

- Brown, Keith (ed.). 2006. *Encyclopedia of language and linguistics*, 2nd edn., 351–355. Oxford: Elsevier.
- Bukmaier, Véronique, Jonathan Harrington, Ulrich Reubold & Felicitas Kleber. 2014. Synchronic variation in the articulation and the acoustics of the Polish place distinction in sibilants and its implications for diachronic change. *15th Annual Conference of the International Speech Communication Association (Interspeech 2014)*, Singapore.
- Cheon, Sang Y. & Victoria B. Anderson. 2008. Acoustic and perceptual similarities between English and Korean sibilants: Implications for second language acquisition. *Korean Linguistics* 14, 41–64.
- Chuang, Yu-Ying & Janice Fon. 2010. The effect of dental and retroflex sibilants in Taiwan Mandarin spontaneous speech. *Proceedings of the 5th International Conference on Speech Prosody*, No. 100414.
- Delattre, Pierre C., Alvin M. Liberman & Franklin S. Cooper. 1962. Formant transitions and loci as acoustic correlates of place of articulation in American fricatives. *Studia Linguistica* 16, 104–121.
- Duanmu, San. 2006. Chinese (Mandarin): Phonology. In Brown (ed.), 351–355.
- Evers, Vincent, Henning Reetz & Aditi Lahiri. 1998. Crosslinguistic acoustic categorization of sibilants independent of phonological status. *Journal of Phonetics* 26, 345–370.
- Fant, Gunnar. 1960. *Acoustic theory of speech production*. The Hague: Mouton.
- Forrest, Karen, Gary Weismer, Paul Milenkovic & Ronald D. Dougall. 1988. Statistical analysis of word-initial voiceless obstruents: Preliminary data. *The Journal of the Acoustic Society of America* 84, 115–123.
- Fujisaki, Hiroya & Osamu Kunisaki. 1978. Analysis, recognition, and perception of voiceless fricative consonants in Japanese. *IEEE Transactions (ASSP)* 26, 21–27.
- Gordon, Matthew, Paul Barthmaier & Kathy Sands. 2002. A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* 32, 141–174.
- Gussmann, Edmund. 2007. *The phonology of Polish*. Oxford: Oxford University Press.
- Halle, Morris & Kenneth N. Stevens. 1997. The postalveolar fricative of Polish. In Shigeru Kiritani, Hajime Hirose & Hiroshi Fujisaki (eds.), *Speech production and language: In honour of Osamu Fujimura*, 177–193. Berlin: Mouton de Gruyter.
- Hamann, Silke. 2002a. Retroflexion and retraction revised. In Tracy A. Hall, Bernd Pompino-Marschall & Marzena Rochoń (eds.), *Papers on phonetics and phonology: The articulation, acoustics and perception of consonants (ZAS Papers in Linguistics 28)*, 13–26. Berlin: Zentrum für Allgemeine Sprachwissenschaft (ZAS).
- Hamann, Silke. 2002b. Postalveolar fricatives in Slavic languages as retroflexes. In Sergio Baauw, Mike Huiskes & Maaike Schoorlemmer (eds.), *OTS yearbook 2002*, 105–127. Utrecht: Utrecht Institute of Linguistics.
- Harrington, Jonathan. 2010. *Phonetic analysis of speech corpora*. Chichester: Wiley-Blackwell.
- Harris, Kathrine S. 1958. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech* 1, 1–7.
- Heinz, Jeff & Kenneth N. Stevens. 1961. On the properties of fricative consonants. *Journal of the Acoustical Society of America* 33, 589–593.
- Hoole, Phil, Lasse Bombien, Barbara Kühnert & Christine Mooshammer. 2009. Intrinsic and prosodic effects on articulatory coordination in initial consonant clusters. In Gunnar Fant, Hiroshi Fujisaki & Jiaxuan Shen (eds.), *Frontiers in phonetics and speech science: Festschrift for Wu Zongji*, 275–286. Beijing: Commercial Press.
- Hoole, P. & Zierdt Andreas. 2010. Five-dimensional articulography. In Ben Maassen & Pascal H. H. M. van Lieshout (eds.), *Speech motor control: New developments in basic and applied research*, 331–349. Oxford: Oxford University Press.
- Hoole, Phil & Christine Mooshammer. 2002. Articulatory analysis of the German vowel system. In Peter Auer, Peter Gilles & Helmut Spiekemann (eds.), *Silbenschnitt und Tonakzente*, 129–152.
- Hu, Fang. 2008. The three sibilants in Standard Chinese. *Proceedings of the 8th International Seminar on Speech Production (ISSP 2008)*, 105–108.
- Jassem, Wiktor. 1995. The acoustic parameters of Polish voiceless fricatives: Analysis of variance. *Phonetica* 52, 252–258.
- Jesus, Luis M. T. & Christine H. Shadle. 2002. A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics* 30, 437–464.

- Jongman, Allard. 1989. Duration of fricative noise required for identification of English fricatives. *Journal of the Acoustical Society of America* 85, 1718–1725.
- Jongman, Allard, Ratrete Wayland & Serena Wong. 2000. Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America* 108(3), 1252–1263.
- Ladefoged, Peter & Peri Bhaskararao. 1983. Non-quantal aspects of consonant production: A study of retroflex consonants. *Journal of Phonetics* 11, 291–302.
- Ladefoged, Peter. 2001. *Vowels and consonants*. Malden, MA: Blackwell.
- LaRivière, Conrad, Harris Winitz & Eve Herriman. 1975. The distribution of perceptual cues in English prevocalic fricatives. *Journal of Speech and Hearing Research* 18, 613–622.
- Li, Fangfang, Jan Edwards & Mary E. Beckman. 2007. Spectral measures for sibilant fricatives of English, Japanese, and Mandarin Chinese. *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS XVI)*, vol. 4, 917–920.
- Li, Fangfang, Ben Munson, Jan Edwards, Kiyoko Yoneyama & Kathleen C. Hall. 2011. Language specificity in the perception of voiceless sibilant fricatives in Japanese and English: Implications for cross-language differences in speech-sound development. *Journal of the Acoustical Society of America* 129, 999–1011.
- Lindblom, Björn. 1998. Systemic constraints and adaptive change in the formation of sound structure. In James R. Hurford, Michael Studdert-Kennedy & Knight Chris (eds.), *Approaches to the evolution of language: Social and cognitive bases*, 242–264. Cambridge: Cambridge University Press.
- Lisker, Leigh. 2001. Hearing the Polish sibilants [s š ś]: Phonetic and auditory judgements. In Nina Grønnum & Jørgen Rischel (eds.), *To honour Eli Fischer-Jørgensen* (Travaux du Cercle Linguistique de Copenhague XXX), 226–238.
- Łobacz, Piotra. 1996. *Polska fonologia dziecięca* [Polish child phonology]. Warszawa: Energeia.
- Lobanov, Boris. 1971. Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America* 49, 606–608.
- Łukaszewicz, Beata. 2006. Extrasyllabicity, transparency and prosodic constituency in the acquisition of Polish. *Lingua* 116, 1–30.
- Łukaszewicz, Beata. 2007. Reduction in syllable onsets in the acquisition of Polish: Deletion coalescence, metathesis, and gemination. *Journal of the Child Language* 34(1), 52–82.
- Mandera, Paweł, Emmanuel Keuleers, Zofia Wodniecka & Marc Brysbaert. 2014. Subtlex-pl: Subtitle-based word frequency estimates for Polish. *Behavior Research Methods* 47(2), 471–483.
- Maniwa, Kazumi, Allard Jongman & Thomas Wade. 2009. Acoustic characteristics of clearly spoken English fricatives. *Journal of the Acoustic Society of America* 125(6), 3962–3973.
- Mann, Virginia & Bruno Repp. 1980. Influence of vocalic context on the perception of the [ʃ]–[s] distinction. *Perception & Psychophysics* 28, 213–228.
- McGuire, Grant. 2007. English listeners' perception of Polish alveopalatal and retroflex voiceless sibilants: A pilot study (UC Berkeley Phonology Lab Annual Report).
- Narayanan, Shrikanth, Abeer Alwan & Katherine Haker. 1995. An articulatory study of fricative consonants using magnetic resonance imaging. *Journal of the Acoustic Society of America* 98(3), 1325–1347.
- Nittrouer, Susan. 1992. Age-related differences in perceptual effect of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics* 20, 1–32.
- Nittrouer, Susan. 2002. Learning to perceive speech: How fricative perception changes, and how it stays the same. *Journal of the Acoustical Society of America* 112, 711–719.
- Nittrouer, Susan & Michael Studdert-Kennedy. 1987. The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech and Hearing Research* 30, 319–329.
- Nowak, Paweł M. 2006. The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of Speech and Hearing Research* 34, 139–152.
- Percival, Donald B. & Andrew T. Walden. 1993. *Spectral analysis for physical applications: Multitaper and conventional univariate techniques*. Cambridge: Cambridge University Press.
- Proctor, Michael, Li H. Lu, Yinghua Zhu, Louis Goldstein & Shrikanth Narayanan. 2012. Articulation of Mandarin sibilants: A multi-plane realtime MRI study. *Proceedings of the 14th Australasian International Conference on Speech Science and Technology*, Sydney, Australia.

- Reidy, Patrick F. 2015. A comparison of spectral estimation methods for the analysis of sibilant fricatives. *JASA Express Letters* 137(4), EL248–EL254.
- Sawicka, Irena. 1995. Fonologia [Phonology]. In Henryk Wróbel (ed.), *Gramatyka współczesnego języka polskiego: Fonetyka i fonologia* [A grammar of contemporary Polish: Phonetics and phonology], 105–194. Warszawa: Wydawnictwo Instytutu Języka Polskiego.
- Schiel, Florian. 2004. MAUS goes iterative. *Proceedings of the IVth International Conference on Language Resources and Evaluation*, Lisbon, Portugal, 1015–1018.
- Shadle, Christine H. 2006. Acoustic phonetics. In Brown (ed.), 442–460.
- Shadle, Christine H. 2010. The aerodynamics of speech. In William J. Hardcastle & John Laver (eds.) *The handbook of phonetic sciences*, 39–80, Blackwell: Oxford.
- Shadle, Christine H. 2012. The acoustic and aerodynamics of fricatives. In Abigail C. Cohn, Cécile Fougeron & Marie K. Huffman (eds.), *The Oxford handbook of laboratory phonology*, 511–526. Oxford: Oxford University Press.
- Shadle, Christine H. & Sheila J. Mair. 1996. Quantifying spectral characteristics of fricatives. *Proceedings of the International Conference on Spoken Language Processing (ICSLP 96)*, Philadelphia, PA, 1517–1520.
- Soli, Sigfrid D. 1981. Second formants in fricatives: Acoustic consequences of fricative vowel coarticulation. *Journal of the Acoustic Society of America* 70, 976–984.
- Srivastava, Anuj, Ian Jermyn & Shantanu Joshi. 2007. Riemannian analysis of probability density functions with applications in vision. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2007)*, 1–8.
- Stevens, Kenneth N. 1998. *Acoustic phonetics*. Cambridge, MA: MIT Press.
- Thomson, David J. 1982. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE* 70(9), 1055–1096.
- Toda, Martine, Shinji Maeda & Kiyoshi Honda. 2010. Formant-cavity affiliation in sibilant fricatives. In Susanne Fuchs, Martine Toda & Marzena Żygis (eds.), *Turbulent sounds: An interdisciplinary guide*, 343–374. Berlin & New York: De Gruyter Mouton.
- van Heuven, Walter J. B., Paweł Mandera, Emmanuel Keuleers & Marc Brysbaert. 2014. SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology* 67(6), 1176–1190.
- Wagner, Anita, Mirjam Ernestus & Anne Cutler. 2006. Formant transitions in fricative identification. *Journal of the Acoustical Society of America* 120(4), 2267–2277.
- Whalen, Douglas H. 1991. Perception of the English /s-/ʃ/ distinction relies on fricative noises and transitions, not on brief spectral slices. *The Journal of the Acoustical Society of America* 90(4), 1776–1785.
- Żygis, Marzena & Silke Hamann. 2003. Perceptual and acoustic cues of Polish coronal fricatives. *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS XV)*, Barcelona, 3–9 August, 395–398.
- Żygis, Marzena, Daniel Pape & Luis M. T. Jesus. 2012a. (Non)retroflex Slavic affricates and their motivation: Evidence from Czech and Polish. *Journal of the International Phonetic Association* 42(3), 281–329.
- Żygis, Marzena, Daniel Pape & Bartłomiej Czaplicki. 2012b. *Dynamics in sibilant systems: Standard Polish and its dialects* (Phonetik & Phonologie 8). Jena: Friedrich-Schiller-Universität Jena.
- Żygis, Marzena, Daniel Pape, Luis M. T. Jesus & Marek Jaskuła. 2014a. Intended intonation of statements and polar questions in whispered, semi-whispered and normal speech modes. *Proceedings of Speech Prosody*, Dublin, 678–682.
- Żygis, Marzena, Daniel Pape, Luis M. T. Jesus & Marek Jaskuła. 2014b. How do voiceless consonant clusters contribute to intended intonation? A comparison of normal, whispered, and semi-whispered speech. *Proceedings of the 10th International Seminar on Speech Production*, Cologne, 476–480.