Synchronic variation in the articulation and the acoustics of the Polish three-way place distinction in sibilants and its implications for diachronic change

Véronique Bukmaier¹, Jonathan Harrington¹, Ulrich Reubold¹, Felicitas Kleber¹

¹Institute of Phonetics and Speech Processing, University of Munich, Germany

[bukmaier|jmh|reubold|kleber]@phonetik.uni-muenchen.de

Abstract

The aim of the present study was to relate articulatory properties of the Polish sibilants /s ʂ ɕ/ to a potential neutralization of /ʂ/ as either /s/ or /ɕ/, the former having occurred in a number of Polish dialects. For this purpose tongue tip (TT) movement data was obtained together with acoustic data using electromagnetic articulography. The sibilants, that were always followed by either /a e o/, were produced by four L1-Polish speakers at fast and slow speech rates. While /s/ /ʂ/ had almost identical transitions, they differed greatly in the spectral characteristics with /s/ being closer to /ɕ/. In order to capture differences in tongue position as well as shape both TT position and TT orientation data were analyzed. The vertical TT orientation showed similarities in /ʂ/ and /s/ production, but the two sibilants were clearly separated in TT position, with /s/ /ʂ/ being produced far more back than /s/ and /ɕ/, and the latter two being very similar. The tendentially greater effect of speech rate on /s/ /ʂ/ together with the varying acoustic and articulatory similarities between the sibilants are taken as an indicator for greater instability of /s/. This synchronous instability is discussed in terms of potential diachronic mergers.

Index Terms: Electromagnetic articulography, three-way place distinction in Polish sibilants, synchronic variation, diachronic change, instability

1. Introduction

The aim of the present study was to relate articulatory properties of the Polish sibilants /s ʂ ɕ/ to a potential neutralization of /ʂ/ as either /s/ or /ɕ/, the former having occurred in a number of Polish dialects. For this purpose tongue tip (TT) movement data was obtained together with acoustic data using an articulatory analysis of the Polish sibilants /s ʂ ɕ/. Standard Polish is one of the very few languages that distinguishes lexically between one anterior and two non-anterior sibilants: dental /s/ (e.g. sali /sali/, Eng. room (gen.)), retroflex /ʂ/ (e.g. szali /sali/, Engl. scale (gen.)), and alveolopalatal /ɕ/ (e.g. siali /cali/, Engl. sown).

As the descriptive terms suggest, the three sibilants differ articulatory not only in place of articulation, but also in tongue shape, as has been shown by MRI data in [1]. /s/ /ʂ/ and /ɕ/ can both be described as sharing a postalveolar place of articulation, and the resulting fricative noise has been reported to be rather similar by showing overlapping centers of gravity [2, 3, 4]; however, both non-anterior sibilants do differ in tongue-shape [1], leading to very different coarticulatory influences on neighboring segments, i.e. to very pronounced acoustic differences in formant transitions. Perception experiments have shown that the three Polish sibilants are distinguished both by spectral properties and by formant transitions into the following vowel /e/ with transitions being more important for the distinction between the non-antenners /s/ /ʂ/ and /ɕ/ and the steady-state friction part for distinguishing anterior /s/ from the non-antenners /ʂ/ /ɕ/. The /s/ /ɕ/ contrast is encoded by both cues and is therefore perceptually robust. The distinction between the retroflex and the other two fricatives depends on only one of the two cues. In particular, as the three fricatives frequently occur in non-prevocalic position in Polish complex onset clusters, the transition cue may be perceptually masked, thus diminishing considerably its perceptual role [3, 4, 5, 6, 7, 8, 9]. Nevertheless these perceptual results together with those from articulatory and acoustic studies suggest a rather stable three-way contrast in Polish sibilants.

The retroflex sibilants in Polish have been claimed to be results of a historical sound change during the 16th century, in which palatalized palatoalveolars depalatalized and became retroflex [10, 11]. [7] reasoned that such a sound change could have come about because of the greater perceptual stability of the alveolopalatal vs. retroflex contrast compared to the earlier contrast of alveolopalatal vs. palatalized palatoalveolar sibilants (an argument which may also support the distribution of non-retroflex vs. retroflex sibilants in the worlds’ languages [12]). Yet, although there is some evidence for a good deal of stability in the three-way contrast in Polish sibilants, comparably crowded sibilant systems are still not only rare in the world’s languages [13], but may be unstable: e.g. most non-standard varieties of Polish have already merged dental and retroflex sibilants [3], and the same sound change has been reported for the very similar three-way distinction of sibilants in Mandarin [14].

Given that coarticulation allows for a reasonably robust perception of differences between the three sibilants in Polish, robustness of perception may diminish in conditions in which the amount of coarticulation may be influenced, as in prosodically weak constituents [15, 16] or at higher speaking rates [17]. Conditions such as these are known to be possible triggers of historical sound changes [18]. One of the main motivations for the present study is to draw a connection between the production of the three sibilants /s, ʂ, ɕ/ and a potential diachronic collapse of the three-way to a binary contrast (most probably by a dental-retroflex merger). A comparison of the acoustic and articulatory Polish sibilant data is a good test case for quantifying a link between place differences and coarticulatory influences in a possible collapse. This is the first experimental study that investigates the production of the three sibilants in terms of electromagnetic articulographic (EMA) measurements of tongue movement and whose focus is the relation between their acoustic and articulatory properties at different speaking rates. The following three hypotheses were tested:

H1: The alveolopalatal fricative has the greatest influence on F2 transitions in neighboring vowels.

H2: The alveolopalatal fricative differs from dental and retroflex fricatives mainly in tongue shape and less so in position.

H3: The relative distance of the retroflex fricatives between dental and alveolopalatal diminishes in fast speech towards the dental fricative.
2. Method

2.1. Data collection and participants

Acoustic and articulatory movement data were collected using electromagnetic articulometry at the IPS in Munich (AG501, Carstens Medizinelektronik; [18]) from four Polish L1-speakers (two male, two female) aged between 19 and 28. The speakers were born in Poland, but lived in Munich, Germany, though no longer than two years at the time of recording.

Two sensors were placed on the tongue: one on the midline 1 cm behind the tongue tip (TT) and the other on a level with the molar teeth at the tongue back (TB). Two sensors were placed on the upper and lower lip. Four additional sensors were fixed to the maxilla, the nose bridge, as well as to the left and right mastoid bones: these served as reference sensors to correct for head movement.

2.2. Speech material

The participants were asked to produce symmetrical CVCV (e.g. /sasa/) non-words (in which C=/s ʂ ɕ/ and V=/a e o/) which were embedded in the carrier phrase ‘Ania wola CVCV aktualnie’ (literally ‘Ania shouted CVCV currently’). In this study, only the initial CV-sequence was analyzed. The speech material was produced at a slow and a fast speech rate. Each carrier phrase was produced with a nuclear pitch accent on the target word, with participants repeating the sentence in case of producing it with an incorrect prosody.

2.3. Experimental set-up

The recording session consisted of ten blocks, alternating between slow and fast speech rates. In order to determine the individual speech rate for each speaker and to adjust the corresponding recording time, each participant was asked to read examples of the speech material at self-selected fast and slow speech rates prior to the actual recording. To ensure consistent within speaker speech rate per condition, the display was enhanced with 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 indicated the time frame for each token. For each block, the carrier sentence containing the target words appeared in random order. In total, each participant produced 360 sentences (3 places of articulation × 3 vowels × 10 repetitions × 4 speakers).

2.4. Analysis of articulatory data

After post-processing the physiological raw data semi-automatically in Matlab, labeling and subsequent analyses of physiological data were conducted using EMU/R [20].

The physiological annotation of the three sibilants was based on the vertical movement of the TT in millimeters and the TT tangential velocity in millimeters per second. The tangential velocity is of importance in detecting TT landmarks because coronal constrictions can include TT raising as well as TT velocity is of importance in detecting TT landmarks because tangential velocity in millimeters per second. The tangential onset of constriction plateau (p on), maximum in constriction (m on), offset of constriction plateau (p off), maximum velocity in constriction (v on), velocity in gestural offset (v off), and gestural offset (g off).

Besides of delivering position data, the TT sensor was also used to determine the differences between the orientations of TT in retroflex vs. alveolarpalatal fricatives. The curled anterior tongue shape is predicted to cause the TT sensor to point upwards for the retroflex, while the lowered anterior part of the tongue in the alveolarpalatal fricative should cause the TT sensor to be oriented downwards [1, 13].

In order to reduce as far as possible speaker differences for further analyses the articulatory data were Lobanov normalized [22]. As to do so, for each utterance the mean value, , of TT was calculated across all of the TT orientation and position values separately between the starting point of the constriction plateau and the endpoint of the constriction plateau of the i-th utterance produced by the speaker.

To quantify the articulatory distance between the three sibilants, the Euclidean distances and were calculated in the vertical TT orientation X horizontal TT position space separately for each sibilant token. The centroids of the dental and the alveolarpalatal sibilants in the slow speech rate served as anchors. The log-Euclidean distance was then calculated for each sibilant, from (1):

\[
d_{iab} = \log(E_i/E_a) = \log(E_i) - \log(E_a) \tag{1}\]

The log-Euclidean distance was calculated in order to obtain one value per sibilant which is a relative measure: greater positive values denote a closer distance to the alveolarpalatal centroid, whereas greater negative values are associated with distances to the dental centroid, while a value of zero denotes that a given sibilant is equidistant in this articulatory space between the dental and the alveolarpalatal centroids (see e.g. [23, 24] for a similar methodology).

2.5. Analysis of acoustic data

The synchronized acoustic data was digitized at 16 kHz and automatically segmented and labeled using forced alignment (Munich Automatic Segmentation tool, [25]). Calculations of spectra (256 point discrete Fourier transform with a 40 Hz frequency resolution, 5 ms Blackmann window, and a frame shift of 5 ms), of formant frequencies (F1-F4; pre-emphasis of -0.8, 20 ms Blackman window with a frame shift of 5 ms), and all further analyses were conducted in EMU/R [20].
acoustic analyses, 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), i.e. 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 Hz, the $m$th DCT-coefficient $C_m$ ($m = 0, 1, 2$) was calculated with the formula (2)

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

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 [20]. Since sibilants are well distinguishable by using only $C_2$ (i.e. the curvature of the spectral slice), all further quantifications of the sibilants were based on this coefficient.

To quantify the acoustic distance between the three sibilants, for each sibilant token the Euclidean distances were calculated in the $C_2$ dimension following formula (1), but with $E_s$ instead of $E_c$. The reason for choosing slow /s/ and /ʂ/ as centroids was because [3] and [5] reported an alveolopalatal center of gravity that was between /s/ and /ʂ/.

To quantify coarticulatory effects, the F2 transitions and the linear slopes (specified by the second DCT coefficient) were calculated for the second formant trajectories (from the onset of the vowel to its temporal midpoint) after applying the discrete cosine transformation (2) to the F2 trajectory (from the onset of the vowel to its temporal midpoint). The acoustic data was again Lobanov normalized [22].

3. Results

3.1. Spectral data and F2 transitions

![Figure 2: Log-Euclidean distance ratio $d_{sv}$ of the alveolopalatal sibilant to the mean positions of the dental and the retroflex sibilants in the $C_2$ dimension (=curvature of the spectral slice). Each box contains one token per vowel and speaker.](image)

With respect to the $C_2$ derived from the spectra, there was greater similarity between retroflex and alveolopalatal sibilants (cf. Fig. 2). This observation was confirmed by a repeated measures ANOVA with $d_{sv}$ as dependent variable and CONSONANT, VOWEL and SPEECH RATE as independent factors: the results showed a significant influence of CONSONANT ($F[2,6] = 85.7$, $p < 0.001$) but no significant influence of VOWEL or SPEECH RATE. In order to test separately the three levels of CONSONANT, post-hoc Bonferroni-adjusted t-tests were carried out, showing significant influences between /s/ and /ʂ/ ($p < 0.001$), as well as between /s/ and /ɕ/ ($p < 0.05$), but no significant differences between /ʂ/ and /ɕ/.

![Figure 3: Mean F2 transitions (time normalized) averaged across vowels from vowel onset to the temporal midpoint separately for the dental (dashed), retroflex (dotted) and alveolopalatal (solid) sibilant and for fast and slow speech rate.](image)

At both speech rates, dental and retroflex sibilants showed quite similar F2 transitions into the vowels, whereas the F2 transitions of alveolopalatal sibilants were shown to differ from those of the other sibilants (cf. Fig. 3). In addition, there was more undershoot in fast than in slow speech. A repeated measures ANOVA with F2 (averaged over the transition from onset to temporal midpoint) as the dependent variable, VOWEL (three levels: /a, e, o/), SIBILANT (three levels /s, ʂ, ɕ/) and SPEECH RATE (two levels: slow, fast) as within-speaker factors was calculated in order to test the observations from Fig. 3. Apart from the significant influence of SIBILANT ($F[2,6] = 46.2$, $p < 0.001$) on the acoustic parameter, there was a predictable significant influence of VOWEL ($F[2,6] = 29.0$, $p < 0.001$) and of SPEECH RATE ($F[1,3] = 11.0$, $p < 0.05$) and a significant VOWEL X SPEECH RATE interaction ($F[2,6] = 6.1$, $p < 0.05$).

In order to test whether there was a difference in the slope of the F2 transitions a repeated measures ANOVA with slope (encoded by the second DCT coefficient) as dependent variable and VOWEL, SIBILANT and SPEECH RATE as independent variables was calculated. In this case, there was a predictable significant influence of VOWEL ($F[2,6] = 5.2$, $p < 0.05$) but no influence of SIBILANT and SPEECH RATE.

3.2. Articulatory analyses: tongue tip (TT) orientation data

![Figure 4: Lobanov-normalized and averaged vertical TT orientation and horizontal TT position at 30% of the constriction plateau duration.](image)
While dental and retroflex sibilants resemble each other in vertical TT orientation (both show a slight upward TT orientation indicated by negative TT orientation values), the alveolopalatal sibilant differs from the two other sibilants in showing a downward TT orientation (indicated by positive TT orientation values). The TT position data shows that dental and alveolopalatal sibilants are fronted (indicated by negative TT position values) compared to the retroflex, which is located further back (indicated by positive TT position; cf. Fig. 4). The log-Euclidean distance ratios $d_{sib}$ in Fig. 5 show no difference between the sibilants. This observation was confirmed by a repeated measures ANOVA with $d_{sib}$ as dependent variable and VOWEL and SPEECH RATE as independent factors showing no significances.

Finally, in order to quantify the influence of speech rate and therefore as well as the influence of speaking style on the articulatory distribution and the articulatory stability of the sibilants, the difference of the TT vertical orientation between the slow and the fast speech rate was calculated.

Several implications for diachronic change can be drawn from these findings for synchronic variation. Because of the acoustic and articulatory differences as well as the availability of two perceptual cues, the distinction between /ɕ/ and /s/ appears quite stable in Standard Polish (although [26] reports of two perceptual cues, the distinction between /ɕ/ and /s/ is nevertheless closer to /s/. The greater acoustic similarity is likely to stem from similar constriction positions between /ɡ/ and /ɕ/, presumably resulting in two cavities with similar resonance frequencies (cf. [1]). On the other hand, the formant transitions into the following vowel are almost identical for retroflex and dental fricatives. The third finding from this study was the greater effect of speech rate showing more variability in TT orientation in /ɡ/ than in /s/ and /ɕ/ production.

Children, on the other hand, rely to a far greater extent on transitions in speech perception than adults [27] thus making the /s-ɕ/-contrast hard to acquire since /ɡ/ and /s/ have almost identical transitions. It is in this respect that a diachronic change from a three-way place distinction to a two-way contrast may also be related to perceptual mergers during language acquisition.

### 4. Discussion & Conclusion

Three major findings arise from the present study that aimed at explaining a potential neutralization of anterior and non-anterior fricatives using an articulatory analysis of the Polish sibilants /s ʂ ɕ/. The first one is that the three-way place contrast in Polish sibilants is maintained articulatorily in terms of different tongue shapes and positions indicating a synchronic stability. Our second finding, however, revealed that the retroflex shows – commensurate with previous acoustic findings – considerable acoustic similarities with both dental and alveolopalatal fricatives. On the one hand, /ɡ/ overlaps greatly with the alveolopalatal fricative in spectral properties (cf. also [3]) which is partly related to their being closer together in TT position than are /s/ and /ɡ/, though /ɡ/ is nevertheless closer to /s/. The greater acoustic similarity is likely to stem from similar constriction positions between /ɡ/ and /ɕ/, presumably resulting in two cavities with similar resonance frequencies (cf. [1]). On the other hand, the formant transitions in the following vowel are almost identical for retroflex and dental fricatives. The third finding from this study was the greater effect of speech rate showing more variability in TT orientation in /ɡ/ than in /s/ and /ɕ/ production.

Three major findings arise from the present study that aimed at explaining a potential neutralization of anterior and non-anterior fricatives using an articulatory analysis of the Polish sibilants /s ʂ ɕ/. The first one is that the three-way place contrast in Polish sibilants is maintained articulatorily in terms of different tongue shapes and positions indicating a synchronic stability. Our second finding, however, revealed that the retroflex shows – commensurate with previous acoustic findings – considerable acoustic similarities with both dental and alveolopalatal fricatives. On the one hand, /ɡ/ overlaps greatly with the alveolopalatal fricative in spectral properties (cf. also [3]) which is partly related to their being closer together in TT position than are /s/ and /ɡ/, though /ɡ/ is nevertheless closer to /s/. The greater acoustic similarity is likely to stem from similar constriction positions between /ɡ/ and /ɕ/, presumably resulting in two cavities with similar resonance frequencies (cf. [1]). On the other hand, the formant transitions in the following vowel are almost identical for retroflex and dental fricatives. The third finding from this study was the greater effect of speech rate showing more variability in TT orientation in /ɡ/ than in /s/ and /ɕ/ production.

Several implications for diachronic change can be drawn from these findings for synchronic variation. Because of the acoustic and articulatory differences as well as the availability of two perceptual cues, the distinction between /ɡ/ and /s/ appears quite stable in Standard Polish (although [26] reports of two perceptual cues, the distinction between /ɡ/ and /s/ is nevertheless closer to /s/. The greater acoustic similarity is likely to stem from similar constriction positions between /ɡ/ and /ɕ/, presumably resulting in two cavities with similar resonance frequencies (cf. [1]). On the other hand, the formant transitions in the following vowel are almost identical for retroflex and dental fricatives. The third finding from this study was the greater effect of speech rate showing more variability in TT orientation in /ɡ/ than in /s/ and /ɕ/ production.

Several implications for diachronic change can be drawn from these findings for synchronic variation. Because of the acoustic and articulatory differences as well as the availability of two perceptual cues, the distinction between /ɡ/ and /s/ appears quite stable in Standard Polish (although [26] reports of two perceptual cues, the distinction between /ɡ/ and /s/ is nevertheless closer to /s/. The greater acoustic similarity is likely to stem from similar constriction positions between /ɡ/ and /ɕ/, presumably resulting in two cavities with similar resonance frequencies (cf. [1]). On the other hand, the formant transitions in the following vowel are almost identical for retroflex and dental fricatives. The third finding from this study was the greater effect of speech rate showing more variability in TT orientation in /ɡ/ than in /s/ and /ɕ/ production.

Several implications for diachronic change can be drawn from these findings for synchronic variation. Because of the acoustic and articulatory differences as well as the availability of two perceptual cues, the distinction between /ɡ/ and /s/ appears quite stable in Standard Polish (although [26] reports of two perceptual cues, the distinction between /ɡ/ and /s/ is nevertheless closer to /s/. The greater acoustic similarity is likely to stem from similar constriction positions between /ɡ/ and /ɕ/, presumably resulting in two cavities with similar resonance frequencies (cf. [1]). On the other hand, the formant transitions in the following vowel are almost identical for retroflex and dental fricatives. The third finding from this study was the greater effect of speech rate showing more variability in TT orientation in /ɡ/ than in /s/ and /ɕ/ production.

Several implications for diachronic change can be drawn from these findings for synchronic variation. Because of the acoustic and articulatory differences as well as the availability of two perceptual cues, the distinction between /ɡ/ and /s/ appears quite stable in Standard Polish (although [26] reports of two perceptual cues, the distinction between /ɡ/ and /s/ is nevertheless closer to /s/. The greater acoustic similarity is likely to stem from similar constriction positions between /ɡ/ and /ɕ/, presumably resulting in two cavities with similar resonance frequencies (cf. [1]). On the other hand, the formant transitions in the following vowel are almost identical for retroflex and dental fricatives. The third finding from this study was the greater effect of speech rate showing more variability in TT orientation in /ɡ/ than in /s/ and /ɕ/ production.
6. References


