

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

¹ Institute 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 /ʂ/ 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 /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 /ʂ/ together with the varying acoustic and articulatory similarities between the sibilants are taken as an indicator for greater instability of /ʂ/. This synchronic 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 explain a neutralization of anterior and non-anterior fricatives that has been observed in various languages 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* /ʂali/, Eng. *scale* (gen.)), and alveopalatal /ɕ/ (e.g. *siali* /ɕali/, Eng. *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]. /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 [3, 5] with transitions being more important for the distinction between the non-anteriors /ʂ/ and /ɕ/ and the steady-state frication part for distinguishing anterior /s/ from the non-anteriors /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 alveopalatal vs. retroflex contrast compared to the earlier contrast of alveopalatal vs. palatalized palatoalveolar sibilants (an argument which may also support the distribution of non-retroflex vs. retroflex sibilants in the world's 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 alveopalatal fricative has the greatest influence on F2 transitions in neighboring vowels.

H2: The alveopalatal 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 alveopalatal 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 ε/ and V=/a e o/) which were embedded in the carrier phrase ‘Ania woła 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 \times 3 vowels \times 10 repetitions \times 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 fronting. Physiological labels included seven different landmarks as can be seen in Fig. 1 [21]. E.g., the beginning and the end of the constriction plateau were interpolated values located at a 20% threshold of two adjacent maxima in the velocity signal.

As the plateau (defined by its on- and offset) of an articulatory gesture is known to be the most stable part in the measurements, the on- and offsets of the TT gesture plateaus,

which are equivalent to the coronal constriction phases, define the time frames in which all articulatory analyses were conducted.

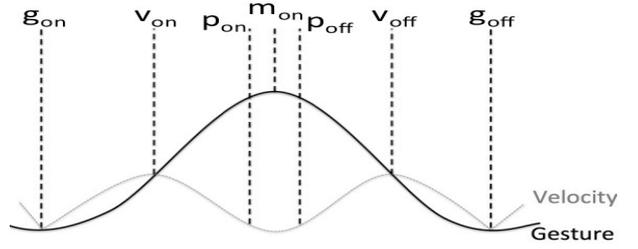


Figure 1: Schematic representation of landmark positions: gestural onset (g_{on}), maximum velocity in gestural onset (v_{on}), onset of constriction plateau (p_{on}), maximum in constriction (m_{on}), offset of constriction plateau (p_{off}), maximum velocity in gestural offset (v_{off}), and gestural offset (g_{off}).

Besides of delivering position data, the TT sensor was also used to determine the differences between the orientations of TT in retroflex vs. alveopalatal 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 alveopalatal 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, m_{xi} , of TT_x 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 E_s and E_e were calculated in the VERTICAL TT ORIENTATION \times HORIZONTAL TT POSITION space separately for each sibilant token. The centroids of the dental and the alveopalatal sibilants in the slow speech rate served as anchors. The log-Euclidean distance ratio d_{sib} was then calculated for each sibilant, from (1):

$$d_{sib} = \log(E_s/E_e) = \log(E_s) - \log(E_e) \quad (1)$$

The log-Euclidean distance ratio d_{sib} was calculated in order to obtain one value per sibilant which is a relative measure: greater positive values denote a closer distance to the alveopalatal 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 alveopalatal 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]. For

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 in (2)

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

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 /ʃ/ as centroids was because [3] and [5] reported an alveopalatal center of gravity that was between /s/ and /ʃ/.

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

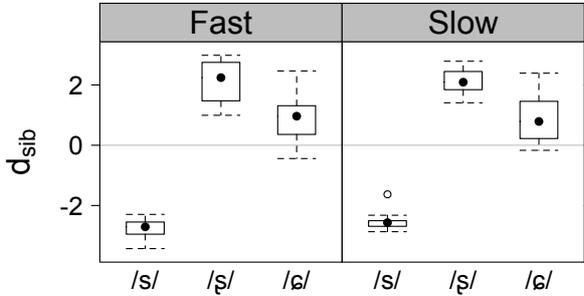


Figure 2: Log-Euclidean distance ratio d_{sib} of the alveopalatal sibilant to the mean positions of the dental and the retroflex sibilants in the C_2 dimension (=curvature of the spectral slice). Each box contains one token per vowel and speaker.

With respect to the C_2 derived from the spectra, there was greater similarity between retroflex and alveopalatal sibilants (cf. Fig. 2). This observation was confirmed by a repeated measures ANOVA with d_{sib} 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 /ʃ/ ($p < 0.001$), as well as between /s/ and /ç/ ($p < 0.05$), but no significant differences between /ʃ/ and /ç/.

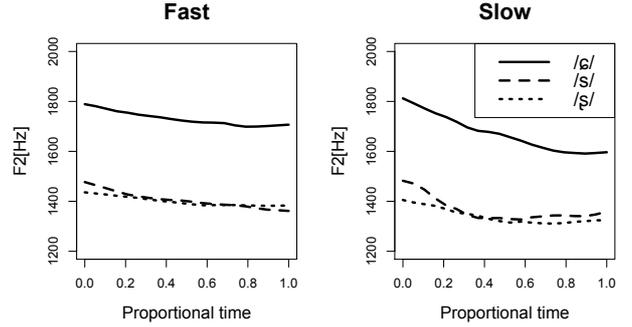


Figure 3: Mean F2 transitions (time normalized) averaged across vowels from vowel onset to the temporal midpoint separately for the dental (dashed), retroflex (dotted) and alveopalatal (solid) sibilant and for fast and slow speech rate.

At both speech rates, dental and retroflex sibilants showed quite similar F2 transitions into the vowels, whereas the F2 transitions of alveopalatals 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, ʃ, ç/) 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 \times 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

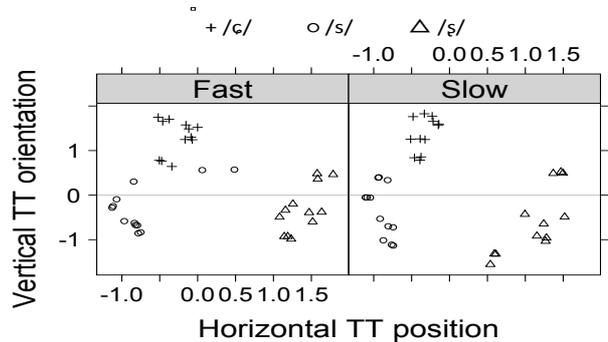


Figure 4: Lobanov-normalized and averaged vertical TT orientation and horizontal TT position at 30% of the constriction plateau duration.

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 alveopalatal 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 alveopalatal 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.

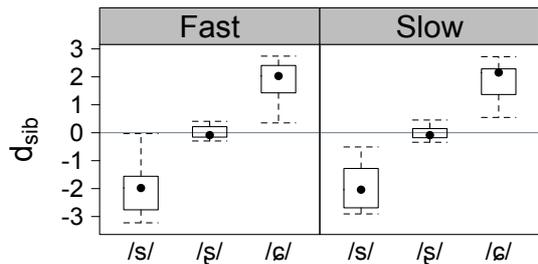


Figure 5: Log-Euclidean distance ratios d_{sib} of the retroflex sibilant to the dental and the alveopalatal sibilant in the VERTICAL TT ORIENTATION \times HORIZONTAL TT POSITION space.

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.

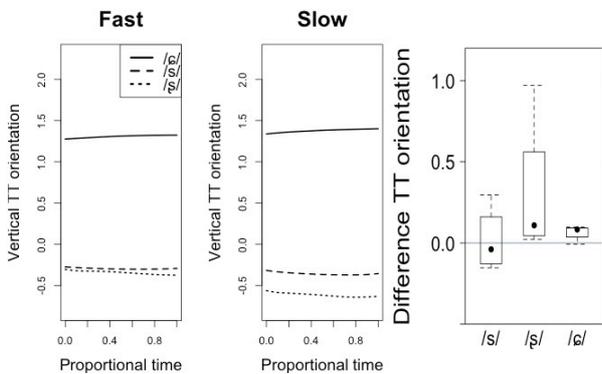


Figure 6: Difference in TT vertical orientation between slow and fast speech rate.

Fig. 6 (a) and (b) show a very stable TT orientation for dental and alveopalatal fricatives across speech rate conditions, while the retroflex's TT orientation differs from that of /s/ only in slow speech. Nonetheless, a repeated measures ANOVA with TT orientation as dependent variable, CONSONANT and RATE as within-speaker variables and SPEAKER as random factor only showed a significant effect for CONSONANT ($F[2,6]=8.3$, $p<0.05$) but not for RATE. Commensurate with Fig. 6 (c), there was also more variation in retroflex compared to dental and alveopalatal sibilants, again indicating a greater difference between slow and fast speech rate in retroflex sibilants. An RM-ANOVA with TT orientation difference as dependent variable revealed no significant difference. Given that the mean TT orientations of

/s/ and /ʂ/ are almost identical, presumably only a number of speakers seem to show an effect of speech rate on TT orientation.

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 alveopalatal fricatives. On the one hand, /ʂ/ overlaps greatly with the alveopalatal fricative in spectral properties (cf. also [3]) which is partly related to their being closer together in TT position than are /s/ and /ʂ/, though /ɕ/ is nevertheless closer to /s/. The greater acoustic similarity is likely to stem from similar constriction positions between /ʂ/ 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 /ʂ/ 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 /ɕ/ and /s/ appears quite stable in Standard Polish (although [26] reports on a change of /ɕ/ into /s/, the latter then being a potential candidate for a merger) and the greater speech rate dependent variability in retroflex fricatives can be taken as an argument for this particular fricative to neutralize. A neutralization of /ʂ/ as /s/, instead of /ɕ/, seems to be more likely, not just because the former, and not the latter, has been observed in a number of diachronic changes [3, 14] but also for the following reasons: (1) /ʂ/ becomes closer to /s/ in fast speech and (2) /ʂ/ contrasts with /s/ only in fricative noise. Although, in the light of these two findings, /ʂ/ might be perceived as /s/ in fast spontaneous speech, nevertheless fricative noise seems to be a rather stable cue in adult speech.

Children, on the other hand, rely to a far greater extent on transitions in speech perception than adults [27] thus making the /s-ʂ/-contrast hard to acquire since /ʂ/ 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.

5. Acknowledgements

This research was supported by ERC grant number 295573 ‘Sound change and the acquisition of speech’ to Jonathan Harrington.

6. References

- [1] Toda, M., Maeda, S., and Honda, K., "Formant-cavity affiliation in sibilant fricatives", in S. Fuchs, M. Toda, and M. Żygis, [Eds], *Turbulent sounds-an interdisciplinary guide*, Berlin, New York: De Gruyter Mouton, 343–374, 2010.
- [2] Jassem, W., "The acoustic parameters of Polish voiceless fricatives: An analysis of variance", *Phonetica*, 52: 252–258, 1995.
- [3] Nowak, P. M., "The role of vowel transitions and friction noise in the perception of Polish sibilants", *Journal of Phonetics*, 34(2): 139 – 152, 2006.
- [4] Żygis, M. and Hamann, S., "Perceptual and acoustic cues of Polish coronal fricatives", in *Proceedings of the 15th International Conference of Phonetic Sciences*, 395–398, 2003.
- [5] Lisker, L., "Hearing the Polish sibilants [s š ś]: Phonetic and auditory judgements", *Travaux du Cercle Linguistique de Copenhague XXXI. To honour Eli Fischer-Jørgensen*, 226–238, 2001.
- [6] Żygis, M. and Padgett, J., "A perceptual study of Polish fricatives, and its implications for historical sound change", *Journal of Phonetics*, 38(2): 207–226, 2010.
- [7] Żygis, M., "The role of perception in Slavic sibilant systems," in Peter Kosta, Joanna Bl aszczak, Jens Frasek, Ljudmila Geist, and Marzena Zygis [Eds], *Investigations into Formal Slavic Linguistics: Contributions of the Fourth European Conference on Formal Description of Slavic Languages*. Frankfurt: Peter Lang, 137–153, 2003.
- [8] Nowak, P. M., "The role of vocalic context in the perception of Polish sibilants", in *Proceedings of the 15th International Congress of the Phonetic Sciences*, 3: 2309–2312, 2003.
- [9] Whalen, D. H., "Perception of the English/s-/j/distinction relies on fricative noises and transitions, not on brief spectral slices", *The Journal of the Acoustical Society of America*, 90(4): 1776–1785, 1991.
- [10] Stieber, Z., "The phonological development of Polish", *Ann Arbor: University of Michigan, Department of Slavic Languages and Literatures*, 1968.
- [11] Padgett, J. and Żygis, M., "The Evolution of Sibilants in Polish and Russian", *Journal of Slavic linguistics*, 15(2), 2007.
- [12] Żygis M., "Contrast optimisation in Slavic sibilant systems", *Habilitation*, Berlin, 2006.
- [13] Ladefoged, P., "Vowels and Consonants", Malden, Mass.: Blackwell, 2001.
- [14] Duanmu, S., "The phonology of standard Chinese", Oxford University Press, 2002.
- [15] de Jong, K., Beckman, M. E., and Edwards, J., "The Interplay Between Prosodic Structure and Coarticulation", *Language and Speech*, 36(2–3): 197–212, 1993.
- [16] Lindblom, B., Agwuele, A., Sussman, H. M., and Cortes, E. E., "The effect of emphatic stress on consonant vowel coarticulation", *The Journal of the Acoustical Society of America*, 121(6): 3802–3813, 2007.
- [17] Agwuele, A., Sussman, H. M. and Lindblom, B., "The Effect of Speaking Rate on Consonant Vowel Coarticulation", *Phonetica*, 65(4): 194–209, 2008.
- [18] Beckman, M. E., Edwards, J., and Fletcher, J., "Prosodic structure and tempo in a sonority model of articulatory dynamics", *Papers in laboratory phonology II*, 68–86, 1992.
- [19] Hoole, P., Zierdt A. and Geng C., "Beyond 2D in articulatory data acquisition and analysis", in *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 265–268, 2003.
- [20] Harrington, J., "Phonetic Analysis of Speech Corpora", Chichester: Wiley-Blackwell, 2010.
- [21] Bombien, L., "Segmental and prosodic aspects in the production of consonant clusters --- On the goodness of clusters", Thesis at: LMU München: Fakultät für Sprach- und Literaturwissenschaften. LMU München: Fakultät für Sprach- und Literaturwissenschaften, 2011.
- [22] Lobanov, B., "Classification of Russian Vowels Spoken by Different Speakers", *The Journal of the Acoustical Society of America*, 49: 606–608, 1971.
- [23] Kleber, F., Harrington, J. and Reubold, U., "The relationship between the perception and production of coarticulation during a sound change in progress", *Language and speech*, 55(3): 383–405, 2012.
- [24] Harrington, J., Kleber, F., and Reubold, U., "Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study", *The Journal of the Acoustical Society of America*, 123: 2825–2835, 2008.
- [25] Schiel, F., "MAUS goes iterative," in *Proc. of the IV. International Conference on Language Resources and Evaluation*, Lisbon, Portugal, 1015-1018, 2004.
- [26] Żygis, M., Pape, D. and Czaplicki, B., "New developments in Polish sibilant system?!", in *Proceedings of the 2nd International Workshop on Sound Change*, Kloster Seeon, Munich (Germany), 2012.
- [27] Nittrouer, S. and Studdert-Kennedy, M., "The role of coarticulatory effects in the perception of fricatives by children and adults", *Journal of Speech and Hearing Research*, 30: 319–329, 1987.