INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Are retroflex-to-dental sibilant substitutions in Polish children's speech an example of a covert contrast? A preliminary acoustic study

Zuzanna Miodonska[1], *Claartje Levelt*[2], *Natalia Mocko*[3], *Michal Krecichwost*[1], *Agata Sage*[1], *Pawel Badura*[1]

[1]Faculty of Biomedical Engineering, Silesian University of Technology, Zabrze, Poland
[2]Leiden University Centre for Linguistics, Leiden University, Leiden, Netherlands
[3]Faculty of Humanities, Institute of Linguistics, University of Silesia, Katowice, Poland

zuzanna.miodonska@polsl.pl

Abstract

The study aimed to investigate whether the dental substitutions of retroflex voiceless fricatives (/ʂ̪/ to [s]) in Polish children's speech are an example of a covert contrast. We analyzed speech samples collected through a picture naming test from 11 children showing this retroflex-to-dental production pattern. The language material included words with /ʂ̪/ and /s/ in various word positions. We extracted a set of spectrum-based acoustic features from the recorded sibilants and conducted the analysis using linear mixed-effect models. The models showed that significant acoustic differences ($p < 0.05$) can be found between realizations of /s/ and /ʂ̪/ substituted by [s]. The main differences were detected in the amplitudes of fricative formants and the energy levels in specific subbands of the frication noise. The study provides preliminary evidence of the existence of covert contrasts in the analyzed substitutions.

Index Terms: child speech, covert contrast, sibilants, Polish

1. Introduction

The acquisition of the phonological system of one's language is a process that happens over years. Researchers of child speech (including speech therapists, pediatricians, and psycholinguists) assume that the process of speech acquisition proceeds in stages, and that the child masters the correct pronunciation around the age of 6;0 [1, 2, 3]. Deviations from the adult pronunciation may be related to insufficient efficiency of articulators, problems with word-form encoding, or to the developmental state of the phonological grammar [4, 5]. Speech sounds that require more precise articulatory movements are initially pronounced as simpler sounds.

One of the examples of this phenomenon can be found in Polish: the realization of the retroflex phonemes /ʂ̪, ʐ̪/ as fricatives with a dental place of articulation. It is often described as a substitution, which suggests that this realization is identical to the realization of the phonemes /s/ or /z/ [6, 7]. However, our hypothesis is that this pronunciation may reflect a covert contrast that has not been explored and described in Polish so far. Covert contrast is a situation in which children in the process of acquisition and improvement of speech realize different phonemes by contrasting them acoustically or articulatory in a way that is not audible to the adult listener – the adult brain interprets these sounds as identical despite the presence of differences detectable in acoustic analysis, electropalatography or ultrasound [8, 9]. This phenomenon occurs as a transitional stage in the child's mastering of phonemic contrasts of the language, serving as an indicator of whether a given segmental contrast already has a representation in the child's mental lexicon. Some studies confirming the presence of covert contrasts in children's speech are available concerning both typically developing children and children with speech disorders (e.g. [10, 11]). Song et al. [12] showed that covert contrasts not only occur during the acquisition of a first language but also in second-language learners. Some authors report that children whose pronunciation showed a covert contrast had a better prognosis in speech therapy than children not realizing any contrast [13], indicating the potential for clinical use of this phenomenon. Two previous studies on this subject analyze sibilants [10, 14], but to date there is no similar research concerning this matter in Polish.

Sibilant sounds – including Polish sibilants – have been investigated for many years. Most of these studies regard adult speakers [15, 16, 17, 18, 19], but there are several studies concerning children [20, 21, 22, 23]. The acoustic features most commonly investigated are spectral moments [18, 24, 25, 26] and spectral peaks (formants of the frication band) [27, 23, 19]; some authors also take into account energies of the noise band [27, 28] or mel-frequency coefficients and their derivatives [16, 29]. Selected acoustic aspects of fricative productions in Polish children have been investigated before [30, 23, 27]. Łobacz & Dobrzanska analyzed the features of Polish fricatives in the speech of 19 typically developing children aged 4–7 [23]. The analysis showed that the fricative formants occur at significantly higher frequencies in dental sibilants than in retroflex fricatives. The researchers also confirmed the relationship between the speaker's age and the gradual approximation of fricatives to the adult-like norm. In one of our previous works, we investigated the acoustics of articulation of Polish preschool children with and without sigmatism [27]. We showed that the acoustic features based on the frication noise band help distinguish different patterns in the pronunciation of retroflex fricatives. However, according to our knowledge, no studies consider the possibility of a covert contrast between perceptually indistinguishable realizations of different phonemes in Polish. Evidence of the existence of covert contrasts in the population of Polish-speaking children would be a significant step toward a better understanding of the process of speech development.

The present study aims to investigate whether the dental substitutions of retroflex sibilants in Polish children's speech are an example of a covert contrast. We examined speech samples from eleven Polish children who realized retroflex fricative /ʂ̪/ as dental [s], comparing these phones to productions of phoneme /s/ realized as [s] by the same children. The analysis employed linear mixed-effect models and was based on spectrum-based acoustic features: spectral moments, fricative formants, and frication noise energies. Our research question was whether retroflex-to-dental productions are acoustically different from dental-to-dental productions articulated by the same speakers.

10.21437/Interspeech.2023-2046

Table 1: *Language material used in the picture naming test*

Word	IPA	Phoneme	Following vowel	Word position	Syllable stress	English translation
szafa	[’ʂafa]	/ʂ/	[a]	initial	+	wardrobe
szufelka	[ʂu’fɛlka]	/ʂ/	[u]	initial	-	dustpan
kalosze	[ka’lɔʂɛ]	/ʂ/	[ɛ]	medial	-	wellingtons
koszyk	[’kɔʂɨk]	/ʂ/	[ɨ]	medial	-	basket
lekarz	[’lɛkaʂ]	/ʂ/	-	final	-	doctor
wąż	[’vɔwʂ]	/ʂ/	-	final	+	snake
samolot	[sa’mɔlɔt]	/s/	[a]	initial	-	plane
serce	[’sɛrtsɛ]	/s/	[e]	initial	+	heart
parasol	[pa’rasɔl]	/s/	[o]	medial	-	umbrella
pasek	[’pasɛk]	/s/	[e]	medial	-	belt
lis	[’lis]	/s/	-	final	+	fox
pies	[’pʲɛs]	/s/	-	final	+	dog

2. Materials and Method

2.1. Speech material

The registration of the speech material used in the present study was performed as part of a larger project on the speech of preschool children. We conducted recordings in three urban-localized kindergartens in a group of 5-to-6 years old children. Each participant took a picture naming test consisting of 35 words containing Polish sibilant sounds in different word positions and a speech examination performed by an experienced SLP (speech-language pathologist). The aim of the examination was to determine the articulation patterns of sibilant sounds in the speech of children, i.e., whether sibilants were articulated according to the Polish standard pronunciation or, when this was not the case, whether we could determine pronunciation characteristics that would suggest a speech disorder for these specific sounds. All gathered acoustic material underwent an additional auditory assessment by an independent annotator (a linguist and an SLP) to ensure that the speech diagnostic description matched the recorded signals. The project was approved by the local Bioethical Committee.

In the present study, we used 12 words with a dental fricative /s/ or a retroflex fricative /ʂ/ (Table 1). The recordings were conducted with a sampling frequency of 44.1 kHz and a resolution of 16 bps. We used a Panasonic WM-61 electret microphone with a 20 to 20,000 Hz frequency response. The distance between the microphone and the child’s mouth was about 10 cm.

The children were not preselected in any manner, as we aimed to gather data that would reflect the speech patterns existing in the population of Polish preschoolers. The inclusion criteria were age (5 or 6 years old) and consent for participation in the study (a written one from the legal guardians and an oral one from the participant).

Taking the abovementioned criteria into account, 79 children were found eligible and were included in the speech corpus. As the aim of the presented study was to analyze [s] realizations of the phoneme /ʂ/, we selected all the speakers that were manifesting a normative articulation of /s/ and a dental realization of /ʂ/. Overall, 11 children (8 boys and 3 girls) showed this production pattern and were included in the analysis.

2.2. Signal preprocessing

We performed segmentation and annotation of sibilants based on the spectrograms using a dedicated Matlab application. During the annotation stage, some items were rejected as they were highly distorted when the participant spoke very loudly or laughed. In some cases, the speaker did not know how to name the picture, so some data was missing. The remaining signals were normalized to range 0–1, partitioned into 20 ms frames with a 10 ms overlap, and windowed with a Hamming window. Finally, we calculated a short-time Fourier transform (STFT) with a spectral resolution of 50 Hz.

2.3. Feature extraction

The signals were then used to calculate literature-based acoustic features of the sibilants: spectral moments, fricative formants and noise band energies. The calculations were performed in accordance to the procedure described by Miodonska et al. [27].

2.3.1. Spectral moments

The four linear spectral moments were calculated for the spectral representation of the signal frame as:

$$\mathrm{L}_1 = \sum_{n=0}^{N_{\mathrm{DFT}}-1} p(n) \cdot f_n, \tag{1}$$

$$\mathrm{L}_m = \sum_{n=0}^{N_{\mathrm{DFT}}-1} p(n) \cdot (f_n - \mathrm{L}_1)^m, \quad 2 \le m \le 4, \tag{2}$$

where $p(n)$ is the normalized power spectrum derived from the N_{DFT}-point DFT for n-th frequency band with a central frequency f_n.

We employed spectral moments in a form that corresponds to descriptive statistics of the spectrum:

- mean, center of gravity: $\mathrm{CoG} = \mathrm{L}_1$,
- standard deviation: $\mathrm{M}_2 = \sqrt{\mathrm{L}_2}$,
- skewness: $\mathrm{M}_3 = \mathrm{L}_3/(\mathrm{L}_2)^{3/2}$,
- and kurtosis: $\mathrm{M}_4 = \mathrm{L}_4/(\mathrm{L}_2)^2 - 3$.

Further analysis employed six spectral moments: CoG, M_2–M_4, and non-normalized versions of third and fourth spectral moments: L_3 and L_4.

2.3.2. Fricative formants

The fricative formants are understood as maxima of the spectrum envelope (spectral peaks) in the noise band accompanying the sibilant sounds. For fricative formant detection, we calculated linear predictive coding coefficients for each signal frame [27]. Peak frequencies were then selected by the rejection of the lower-amplitude peak of each of two maxima located too close to each other ($F_{n+1} - F_n < 500$ Hz). Finally, four lowest-frequency maxima above the threshold 1900 Hz were selected as the frequencies of the four fricative formants (FF_1, FF_2, FF_3, FF_4). The amplitudes (FFL_i) of fricative formants were determined as the mean power spectrum of the FF_i frequency and its two nearest neighbors.

Apart from FF_1–FF_4 and FFL_1–FFL_4, we employed measures concerning the highest-amplitude peak in the middle-frequency band: a maximum of a power spectrum in the range of 2-7 kHz (peak amplitude, PA) and its corresponding frequency (peak frequency, PF).

2.3.3. Frication noise energies

The noise energies NE_k were determined for individual subbands from 2000 to 7000 Hz:

$$NE_k = \sum_{n=n_{l,k}}^{n_{u,k}} |DFT_n|^2, k = 0, 1, 2, ..., K - 1, \qquad (3)$$

where $n_{l,k}$ and $n_{u,k}$ indicate the spectrum bins related to cut-off frequencies of the k-th DFT subband, and $K = 10$ is the number of subbands with a width of 500 Hz.

2.4. Statistical analysis

We used linear mixed-effects (LME) models to evaluate acoustic differences between dental and retroflex sibilants assessed as having dental realization by the SLPs. In the collected corpus, the sibilants occurred in different phonetic contexts (different preceding and following vowels), so we decided to analyze 40% of the frames from the middle of each sibilant, thus reducing the coarticulation impact on the results. Overall, the analysis included 664 observations (301 [s] realizations of /ʂ/ and 363 [s] realizations of /s/).

For each response variable (26 acoustic features described in Section 2.3), we tested several LMEMs with different random structures. Then, we used ANOVA to compare the models, starting with models best fit to the data (highest values of log-likelihood). Finally, our models included *Phoneme*={/s/, /ʂ/} as a fixed effect and a maximized random structure: *Speaker* and *Word* as random intercepts, by-speaker random slopes for *Word position*, *Syllable stress*, and *Following vowel*, and by-word random slope for *Speaker*.

All analyses were performed in MATLAB R2021a at $p = 0.05$.

3. Results & Discussion

From the 26 acoustic features that were analyzed, 10 proved to be significantly different between dental-to-dental and retroflex-to-dental realizations. The significant differences between phonemes were obtained for spectral skewness (M_3), spectral energy in the bands 2000-2500 Hz (NE_0), 5500-7000 Hz (NE_7–NE_9), the second fricative formant (FF_2), and all the calculated formant levels (FFL_1–FFL_4). The statistically significant results are presented in Table 2.

Most differences were found in spectral energies and formant amplitudes, and not in the frequency patterns. Formant frequency structure proved to be similar between the retroflex-to-dental and dental-to-dental productions. That could explain the aural resemblance, as differences in spectral amplitudes are, in general, less evident to perceive. The exception was FF_2,

Table 2: *The summary of linear-mixed effect models for acoustic features (responses) with statistically significant findings. Dental articulation of phoneme /ʂ/ (/ʂ/ as [s]) was a baseline and is presented as the Intercept. Estimates for normative (dental) /s/ are given in reference to the Intercept value*

	Effect	**Est**	**SE**	**t**	**p**
M_3	Intercept	3.762	0.066	57.187	0
	/s/	0.235	0.106	2.222	0.027
NE_0	Intercept	-403.01	5.226	-77.114	0
	/s/	-16.895	6.975	-2.422	0.016
NE_7	Intercept	-272.29	4.741	-57.431	0
	/s/	-17.096	6.731	-2.54	0.011
NE_8	Intercept	-260.21	6.797	-38.284	0
	/s/	-30.696	10.312	-2.977	0.003
NE_9	Intercept	-274.34	7.095	-38.666	0
	/s/	-25.548	9.695	-2.635	0.009
FF_2	Intercept	4.552	0.046	98.122	0
	/s/	0.306	0.041	7.481	0
FFL_1	Intercept	-36.201	0.794	-45.569	0
	/s/	-2.258	0.815	-2.77	0.006
FFL_2	Intercept	-21.782	0.474	-45.951	0
	/s/	-1.666	0.581	-2.868	0.004
FFL_3	Intercept	-21.234	0.357	-59.563	0
	/s/	-1.279	0.463	-2.763	0.006
FFL_4	Intercept	-22.456	0.442	-50.796	0
	/s/	-1.774	0.504	-3.518	0

which was significantly higher in dental phoneme productions than in the retroflex-to-dental substitutions. This is in line with the typical formant patterns found in sibilant sounds: FF_2 is known to be generally lower in retroflex than in dental fricatives [19, 23, 27].

We also performed an exploratory within-speaker analysis considering the spectrum's shape in the band 2–7 kHz. We averaged the power spectral density over all signal frames provided by individual speakers to see whether the differences in noise energies detected by LME models (NE_0, NE_7–NE_9) would be visible for particular participants. We found three basic patterns that occur in the speakers' spectra: *(1)* no contrast between the productions, *(2)* the contrast expressed in the altered formants structure, and *(3)* the contrast in the frication noise energy. Some speakers did not produce any significant acoustic contrast concerning spectral energies or fricative formant structure in the investigated realizations. The example of speaker M8's productions' spectra is presented in Fig. 1. In this case, the shape of the spectral envelope in the analyzed band and the amplitudes of specific frequency components are very similar for productions of both phonemes. A different pattern can be observed in the sibilants produced by speaker F3 (Fig. 2). The averaged spectra for investigated sounds differed significantly, mostly in the amplitudes of the spectrum components: the spectral energies in retroflex-to-dental productions were consistently lower than in dental-to-dental. In the case of speaker M6 (Fig. 3), we observe a different structure of the spectral peaks: the noise subband with the highest energy occurs in lower frequencies for /ʂ/ realizations than in /s/. This is compliant with the differences

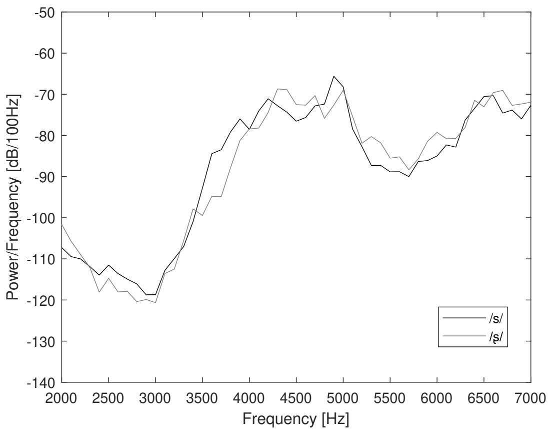

Figure 1: *The averaged power spectrum density of /s/ (black line) and /ʂ/ articulated as [s] (gray line) by speaker M8. No significant contrast between phoneme productions is detectable in the analyzed frequency band. The negative values result from a logarithmic representation of near-zero magnitude levels*

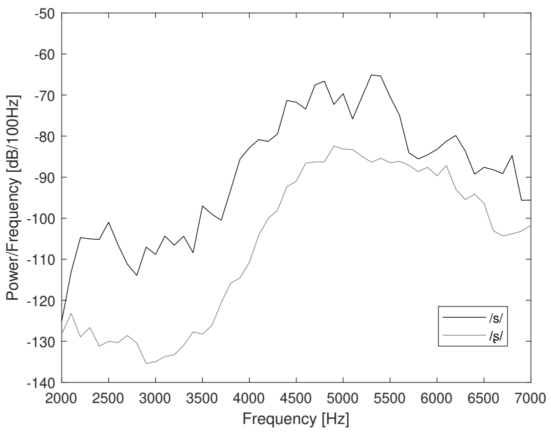

Figure 2: *The averaged power spectrum density of /s/ (black line) and /ʂ/ articulated as [s] (gray line) by speaker F3. The spectrum amplitudes in the analyzed band are consistently lower in /ʂ/ productions than in /s/. The negative values result from a logarithmic representation of near-zero magnitude levels*

between normative retroflex and dental sibilants reported in the literature [19, 27, 23].

The contrasts visible in the realizations produced by speakers M6 and F3 can be two different examples of introducing transitional forms of a retroflex fricative into the developing articulation. Not all the speakers produced the contrast in the investigated phoneme set, which can probably be justified by differences in the level of development of the phonological system and motor control in the participants; this issue could be analyzed in a longitudinal study, that would reveal whether the contrast evolves over time depending on a speech therapy intervention.

One of the significant limitations of our study was set by the available language material. We recorded the speech corpus as part of a larger project, and the vocabulary was not explicitly selected to ascertain uniform phonetic contexts. Therefore,

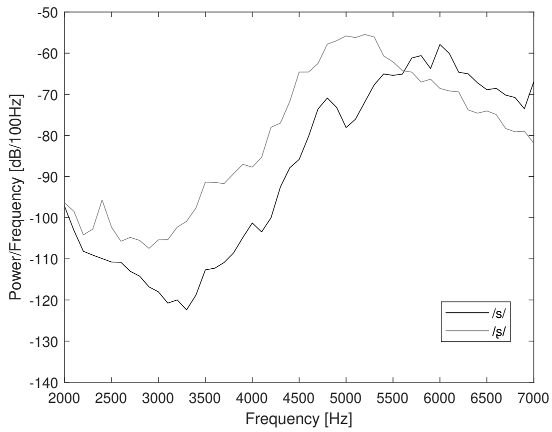

Figure 3: *The averaged power spectrum density of /s/ (black line) and /ʂ/ articulated as [s] (gray line) by speaker M6. The spectral envelope in /ʂ/ productions is shifted toward lower frequencies than in /s/. The negative values result from a logarithmic representation of near-zero magnitude levels*

retroflex and dental sibilants in the analyzed words were adjacent to different vowels. To address this issue, we ignored onsets and offsets of analyzed fricatives and processed only the middle parts of the segments. Moreover, using linear mixed-effect models enabled us to introduce corrections on the diversity found in the speech material: the maximized random effect structure included random slopes for the stress, word position, and following vowel, as well as speaker- and word-based grouping. We plan to extend the database in the future: collect samples from additional speakers and supplement the word set with phrases providing a more unified phonetic context. That would allow us to perform a more detailed analysis and investigate intra-subject data more extensively. The selection of acoustic features indicating the presence or absence of the contrast in a particular child could be used as a numerical indicator for estimating the risk of difficulties at the level of phonological coding and the risk of developing a speech sound disorder.

4. Conclusions

According to our knowledge, the present study is the first reported attempt to investigate the acoustic nature of sibilant substitutions from the perspective of a possible covert contrast. Our results show that significant acoustic differences in the features of the frication band spectrum can be found between realizations of /s/ and /ʂ/ substituted by [s]. This provides preliminary evidence of the existence of covert contrasts in the retroflex-to-dental substitutions in the pronunciation of Polish preschool children.

5. Acknowledgements

This work was partially supported by the Polish National Science Centre, Poland under Grant 2018/30/E/ST7/00525 (Hybrid System for Acquisition and Processing of Multimodal Signal in the Analysis of Sigmatism in Children) and partially by the Foundation for Polish Science (FNP). We want to express our thanks to all study participants and to the speech-language pathologists who performed the speech examinations.

6. References

[1] S. M. B. Frances L. Ilg, Louise Bates Ames, *Child Behavior*. HarperCollins, 1981. [Online]. Available: https://books.google.pl/books?id=wnPoCDULf9IC

[2] J. Porayski-Pomsta, *Our child talks: on the speech development in children, (PL) Nasze dziecko mówi. Rozważania o rozwoju mowy dziecka*. Wydawnictwo Harmonia, 2023.

[3] J. Bruner, "Child's talk: Learning to use language," *Child Language Teaching and Therapy*, vol. 1, no. 1, pp. 111–114, 1985. [Online]. Available: https://doi.org/10.1177/026565908500100113

[4] J. G. Foy and V. Mann, "Changes in letter sound knowledge are associated with development of phonological awareness in pre-school children," *Journal of Research in Reading*, vol. 29, no. 2, pp. 143–161, 2006.

[5] M. Łuszczuk, "Evaluation of the structure and motility of speech organs, (PL) Ocena budowy i motoryki narządów mowy," in *Preschool and early school speech therapy*, A. Domagała and U. Mirecka, Eds. Wydawnictwo Harmonia, 2021.

[6] D. Emiluta-Rozya and D. Lipiec, "Articulation disorders - causes, symptomatology, classifications, (PL) Zaburzenia artykulacji – przyczyny, symptomatologia, klasyfikacje," in *Preschool and early school speech therapy*, A. Domagała and U. Mirecka, Eds. Wydawnictwo Harmonia, 2021.

[7] J. Trzaskalik, "Diagnosis and therapy of sigmatism, (PL) Diagnoza i terapia sygmatyzmu," in *Preschool and early school speech therapy*, A. Domagała and U. Mirecka, Eds. Wydawnictwo Harmonia, 2021.

[8] J. Scobbie, F. Gibbon, W. Hardcastle, and P. Fletcher, "Covert contrast as a stage in the acquisition of phonetics and phonology," in *Papers in laboratory phonology V: Language acquisition and the lexicon*, . J. P. M. Broe, Ed. Cambridge: Cambridge University Press, 2000, p. 194–203.

[9] F. E. Gibbon and A. Lee, "Preface to the special issue on covert contrasts," *Clinical Linguistics & Phonetics*, vol. 31, no. 1, pp. 1–3, 2017, pMID: 27925482. [Online]. Available: https://doi.org/10.1080/02699206.2016.1254684

[10] F. Li, J. Edwards, and M. E. Beckman, "Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers," *Journal of Phonetics*, vol. 37, no. 1, pp. 111–124, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0095447008000582

[11] F. E. Gibbon and A. Lee, "Electropalatographic (EPG) evidence of covert contrasts in disordered speech," *Clinical Linguistics & Phonetics*, vol. 31, no. 1, pp. 4–20, 2017, pMID: 27267128. [Online]. Available: https://doi.org/10.1080/02699206.2016.1174739

[12] J. Y. Song and F. Eckman, "Covert contrasts in the acquisition of English high front vowels by native speakers of Korean, Portuguese, and Spanish," *Language Acquisition*, vol. 26, no. 4, pp. 436–456, 2019. [Online]. Available: https://doi.org/10.1080/10489223.2019.1593415

[13] A. A. Tyler, G. R. Figurski, and T. Langsdale, "Relationships between acoustically determined knowledge of stop place and voicing contrasts and phonological treatment progress," *Journal of Speech, Language, and Hearing Research*, vol. 36, no. 4, pp. 746–759, 1993. [Online]. Available: https://pubs.asha.org/doi/abs/10.1044/jshr.3604.746

[14] S. R. Baum and J. C. McNutt, "An acoustic analysis of frontal misarticulation of /s/ in children," *Journal of Phonetics*, vol. 18, no. 1, pp. 51–63, 1990. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0095447019303584

[15] T. Yoshinaga, K. Nozaki, and S. Wada, "Aeroacoustic analysis on individual characteristics in sibilant fricative production," *The Journal of the Acoustical Society of America*, vol. 146, no. 2, pp. 1239–1251, 2019.

[16] L. Spinu, A. Kochetov, and J. Lilley, "Acoustic classification of Russian plain and palatalized sibilant fricatives: Spectral vs. cepstral measures," *Speech Communication*, vol. 100, pp. 41–45, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167639317303680

[17] A. Kochetov, "Acoustics of Russian voiceless sibilant fricatives," *Journal of the International Phonetic Association*, vol. 47, no. 3, p. 321–348, 2017.

[18] L. Spinu and J. Lilley, "A comparison of cepstral coefficients and spectral moments in the classification of Romanian fricatives," *Journal of Phonetics*, vol. 57, pp. 40–58, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0095447016300109

[19] M. Toda, S. Maeda, and K. Honda, "Formant-cavity affiliation in sibilant fricatives," in *Turbulent Sounds*, ser. Interface Explorations, S. Fuchs, M. Toda, and M. Zygis, Eds. Berlin: De Gruyter Mouton, 2010, pp. 341–371.

[20] N. Zharkova, W. J. Hardcastle, and F. E. Gibbon, "The dynamics of voiceless sibilant fricative production in children between 7 and 13 years old: An ultrasound and acoustic study," *J Acoust Soc Am*, vol. 144, no. 3, p. 1454, 09 2018.

[21] S. L. Nissen and R. A. Fox, "Acoustic and spectral characteristics of young children's fricative productions: a developmental perspective," *J Acoust Soc Am*, vol. 118, no. 4, pp. 2570–2578, Oct 2005.

[22] F. Li and B. Munson, "The development of voiceless sibilant fricatives in Putonghua-speaking children," *Journal of Speech, Language, and Hearing Research*, vol. 59, no. 4, pp. 699–712, 2016.

[23] P. Łobacz and K. Dobrzańska, "Acoustic description of sibilant sounds in the pronunciation of preschool children, (PL) Opis akustyczny glosek sybilantnych w wymowie dzieci przedszkolnych," *Audiofonologia*, vol. 15, pp. 7–26, 1999.

[24] K. L. Haley, E. Seelinger, K. C. Mandulak, and D. J. Zajac, "Evaluating the spectral distinction between sibilant fricatives through a speaker-centered approach," *Journal of Phonetics*, vol. 38, no. 4, pp. 548 – 554, 2010.

[25] P. M. Nowak, "The role of vowel transitions and frication noise in the perception of Polish sibilants," *Journal of Phonetics*, vol. 34, no. 2, pp. 139 – 152, 2006.

[26] C. Patgiri, M. Sarma, and K. K. Sarma, "A class of neuro-computational methods for Assamese fricative classification," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 5, no. 1, pp. 59–70, 2015.

[27] Z. Miodonska, P. Badura, and N. Mocko, "Noise-based acoustic features of Polish retroflex fricatives in children with normal pronunciation and speech disorder," *Journal of Phonetics*, vol. 92, p. 101149, 2022. [Online]. Available: https://doi.org/10.1016/j.wocn.2022.101149

[28] V. Evers, H. Reetz, and A. Lahiri, "Crosslinguistic acoustic categorization of sibilants independent of phonological status," *Journal of Phonetics*, vol. 26, no. 4, pp. 345 – 370, 1998.

[29] I. Anjos, N. Cavalheiro Marques, M. Grilo, I. Guimarães, J. Magalhães, and S. Cavaco, "Sibilant consonants classification comparison with multi- and single-class neural networks," *Expert Systems*, vol. 37, no. 6, p. e12620, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.12620

[30] P. Łobacz, *Polish children's phonology. Phonetic and acoustic studies, (PL) Polska fonologia dziecieca. Studia fonetyczno-akustyczne*. Wydawnictwo Energeia, 1996.