L2 acquisition of Polish sibilant fricatives by L1 Czech speakers

Phil J. Howson[1] & Irfana Madathodiyil[2]

[1]Leibniz-Zentrum Allgemeine Sprachwissenschaft; [2]All India Institute of Speech and Hearing
howson@leibniz-zas.de; irfana@aiishmysore.in

ABSTRACT

Polish language has an uncommon three-way sibilant contrast /s, ʂ, ɕ/. For learners of Polish who have an L1 system that has two-way sibilant contrast (e.g., /s, ʃ/), this means that they must learn two segments (i.e., /ʂ, ɕ/) that are acoustically and perceptually similar to one L1 segment.

In this study, we examined 5 A-level and 5 C-level L2 leaners of Polish who were L1 Czech speakers. Participants produced sentences in both Czech and Polish containing their L1 and L2 sibilants, while we recorded ultrasound data. Tongue contours were traced and extracted. GAMMs were used to compare contours.

The results revealed that the acquisition process was long and extended at least past the A-level. But that typically, only /ʂ/ was acquired, while /ɕ/ remained articulatorily linked to /ʃ/. The results implicate goodness-of-fit as an important predictor for acquisition and that L2 segments are integrated into the L1 phonetic system.

Keywords: L2 acquisition, Polish, Sibilants, phonetics, Czech

1. INTRODUCTION

Second language acquisition requires that learners develop new articulatory routines and acquire the phonetic and phonological representation for novel segments. Two influential theories for second language acquisition are the Perceptual Assimilation Model L-2 [2] and the Speech Learning Model [3, 4].

2.1. Perceptual Assimilation Model-L2

The perceptual assimilation model-L2 [2], henceforth PAM-L2, is an extension of the perceptual assimilation model [1] to second language learners. In this model, novel L2 segments can either be perceptually assimilated to an L1 segment if they are similar in some phonetic dimension or they can remain uncategorized if they are sufficiently different from L1 segments. In the context of Czech learners of Polish, single category assimilation is expected for /ʂ, ɕ/ because they are perceptually similar to /ʃ/ for non-native listeners [7]. Potential differences in goodness-of-fit to the L1 category, or rather how perceptually similar /ʂ, ɕ/ are to /ʃ/, can vary between the two segments and may potentially inhibit or facilitate learning of one or both segments. In either case, the model predicts that if there are enough perceptual differences between the assimilated L2 segments and the L1 segments, learning can occur.

2.2. Speech Learning Model

The speech learning model [3, 4], henceforth the SLM, posits that L2 segments are mapped onto perceptually similar L1 segments during learning. The SLM additionally posits that the mechanisms involved in L1 acquisition are intact for L2 acquisition and that the learning process, similar to L1 acquisition, takes a significant amount of time. However, for learning to take place, disruption of L2-to-L1 perceptual links is necessary.

2.3. Hypothesis

While both the PAM-L2 and SLM were developed for second language acquisition contexts (SLA, i.e., contexts where learners are immersed in a second language), we believe it has applications in foreign language contexts as well (FLA, i.e., contexts where learners are not immersed and learn in a classroom). We hypothesize that the novel Polish sibilants, /ʂ, ɕ/, will initially develop as both perceptually linked to /ʃ/. However, due to larger differences in formant transitions and spectral cues, /ʂ/ will be easier to acquire as a new category, while /ɕ/ will be perceptually linked to /ʃ/ much longer, delaying, or even blocking acquisition. To test this hypothesis, we performed an ultrasound study of L1 Czech speakers' acquisition of L2 Polish. We recorded ultrasound data for L1 and L2 sibilant inventories.

2. METHODS

2.1. Participants

Ten L1 speakers of Czech learning Polish participated in this experiment (5 female, 5 male; 5 A-, 5 C-level). Participants were all college-educated, aged between 20-24 years, and had no self-reported history of speech or hearing disorders.

2.2 Stimuli & Procedure

Participants read sentences in Polish and Czech. Each language was blocked so that only Czech or only Polish words were produced, and the order of the blocks were counterbalanced. Target words were embedded in carrier phrases, for Polish "powiedz *target* proszę" and for Czech "řekněte *target* ještě jednou." Each of the target segments were placed in three vocalic environments, /e, a, o/. Each target word was produced 8 times for a total of 1,200 token (8 repetitions x 5 segments x 3 environments x 5 learners x 2 groups). Tables 1 and 2 present the target words for Polish and Czech, respectively.

Segment	Target	English
	de**s**ek	planks
/s̪/	ta**s**ak	cleaver
	o**s**o	wasp
	rze**sz**e	multitudes
/ʂ/	pa**sz**a	feed
	o**sz**o	fallen
	le**si**e	forest
/ɕ/	na**si**ał	has clouded
	o**si**oł	donkey

Table 1. Polish stimuli with target segments in bold.

Segment	Target	English
	ve**s**el	oars
/s/	já**s**at	to rejoice
	po**s**olit	to salt
	ve**š**el	went in
/ʃ/	Dá**š**a	Dasha (name)
	po**š**oupat	to push/move

Table 2. Czech stimuli with target segments in bold.

Data were recorded in a quiet room in Prague using the AAA suite and a Micro ultrasound with a 20mm probe that had 92-degree field of view. The average frame rate was 90 frames per second.

2.2 Analysis

We annotated the data for each of the relevant segments based on the spectrogram. The onset was marked as the end of the formant structure of the preceding vowel and the onset of aperiodic noise, while the offset was marked as the end of the aperiodic noise and the onset of formant structure for the following vowel. DeepCutLab for Speech [13] was then used to trace tongue contours at the midpoint of each segment. DeepCutLab [5] is a machine learning algorithm for automatic feature extraction and has been trained on ultrasound data to detect edges for use with AAA. x- and y-coordinates were then extracted, normalized, and compared with generalized additive mixed models (GAMMs) [9] using the mgcv package [12] in R [6]. Posthoc comparisons were performed with tidymv [8]. We generated models for the Czech sibilant system and the Polish sibilant system for a total of 4 models (2 groups x 2 models). We included smoothing terms for the x-axis and x-axis by Segment. A factor smooth for x-axis and the interaction between Segment, Environment, and Participant was also included. Data were visualized using ggplot2 [13].

3. RESULTS

3.1 A-Level Learners

The statistical analysis for A-level learners of Polish revealed significant effects for the smoothing term for the interaction between the x-axis and /s̪/ [F(4.5, 4.7) = 98; p < 0.01], /ʂ/ [F(1.6, 1.7) = 8; p = 0.01], and /ɕ/ [F(3.3, 3.5) = 30; p < 0.01]. The R^2 for the model was 0.78. Figure 1 presents the GAMM contours for /s̪, ʂ, ɕ/. Figure 2 presents the estimated differences between /ʂ, ɕ/.

The GAMM contours revealed A-level learners distinguished /s̪/ from /ʂ, ɕ/, and that there was a slightly lower body for /ʂ/ than /ɕ/ (Figure 2). The results further indicated that /s̪/ had a more retracted tongue dorsum, compared to /ʂ, ɕ/ and a lower tongue body/blade. /ʂ, ɕ/ had a more advanced tongue root and a more raised tongue body.

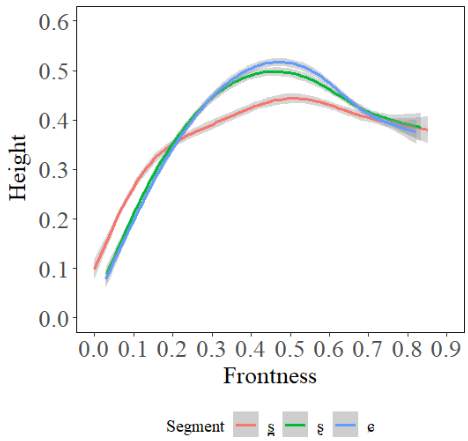

Figure 1: GAMM tongue contours for /s̪/ (red), /ʂ/ (green), /ɕ/ (blue) for A-level learners. Tongue tip is on the right.

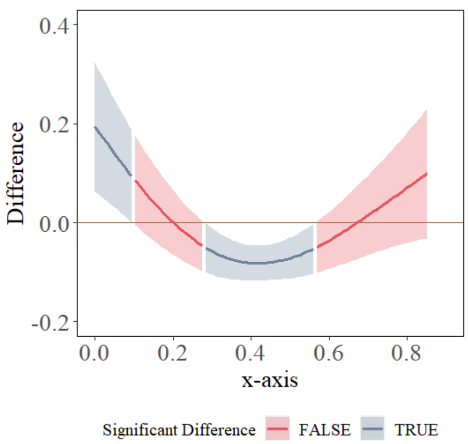

Figure 2: Estimated difference in contours between /ʂ/ and /ɕ/ for A-level learners. Areas with red indicate no significant differences between the two contours.

The comparison of Czech /ʃ/ and Polish /ʂ, ɕ/ revealed a significant effect for interaction between the x-axis and /ʃ/ [F(1.8, 1.9) = 23; $p < 0.01$], /ɕ/ [F(2.3, 2.5) = 18; $p < 0.01$], and /ʂ/ [F(3.1, 3.4) = 40; $p < 0.01$]. The R^2 was 0.83. The observable difference was a slightly lowered tongue body for /ʂ, ɕ/, although the overall tongue contour was comparable. Figure 3 presents the comparison of /ʃ, ʂ, ɕ/.

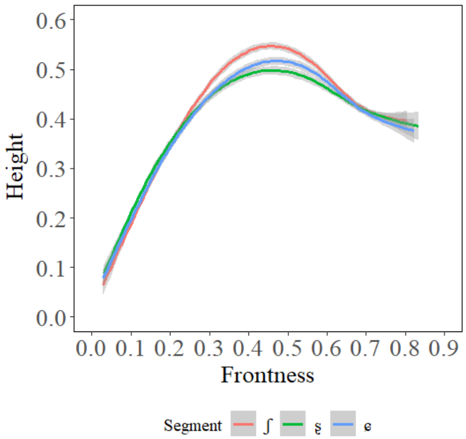

Figure 3: GAMM tongue contours for /ʃ/ (red), /ʂ/ (green), /ɕ/ (blue) for A-level learners of Polish. Tongue tip is on the right.

3.2 C-Level Learners

The analysis for C-level learners of Polish revealed significant effects for the smoothing terms for the interaction between the x-axis and /ʂ̪/ [F(5, 5) = 10; $p < 0.01$], /ʂ/ [F(5, 5) = 267; $p < 0.01$], and /ɕ/ [F(3.9, 4.3) = 191; $p < 0.01$]. The R^2 value for the model was 0.67. Figure 4 presents the GAMM contours for /ʂ̪, ʂ, ɕ/. Figure 5 presents the estimated differences between /ʂ, ɕ/.

The GAMM contours for C-level learners revealed a distinct three-way sibilant contrast. /ʂ̪/ had a retracted tongue dorsum compared to /ɕ/, and a relatively low and flat tongue body/blade compared to /ʂ, ɕ/. /ʂ/ had the most retracted tongue dorsum, with a tongue body/blade raised in between /s, ɕ/. Finally, /ɕ/ had the most advanced tongue dorsum and the most fronted and raised tongue body/blade.

The comparison of Polish /ʂ, ɕ/ and Czech /ʃ/ revealed a significant effect for the interaction between the x-axis and /ʃ/ [F(5, 5) = 193; $p < 0.01$], /ʂ/ [F(3.3, 3.8) = 201; $p < 0.01$], and /ɕ/ [F(3.6, 3.1) = 189; $p < 0.01$]. The R^2 was 0.70. The comparison of /ʃ, ʂ/ revealed differences in the tongue dorsum and the tongue body/blade, with /ʂ/ having a more retracted tongue dorsum and a lowered tongue body. The only area of significant difference between /ʃ, ɕ/ was in the most posterior regions of the tongue dorsum, indicating a slightly more retracted dorsum for /ʃ/.

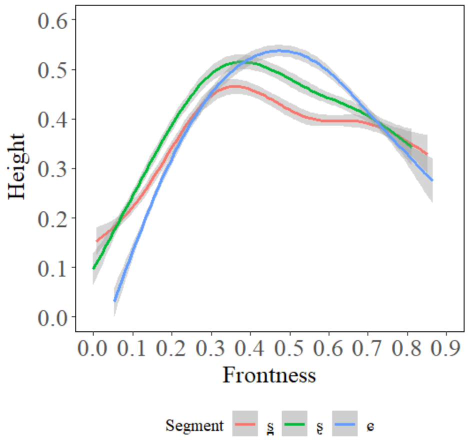

Figure 4: GAMM tongue contours for /ʂ̪/ (red), /ʂ/ (green), /ɕ/ (blue) for C-level learners of Polish. Tongue tip is on the right.

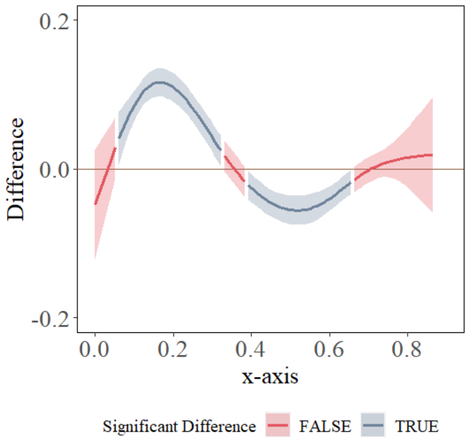

Figure 5: Estimated difference in contours between /ʂ/ and /ɕ/ for C-level learners. Areas with red indicate no significant differences between the two contours.

Figure 6 presents the comparison of Czech /ʃ/ and Polish /ʂ, ɕ/.

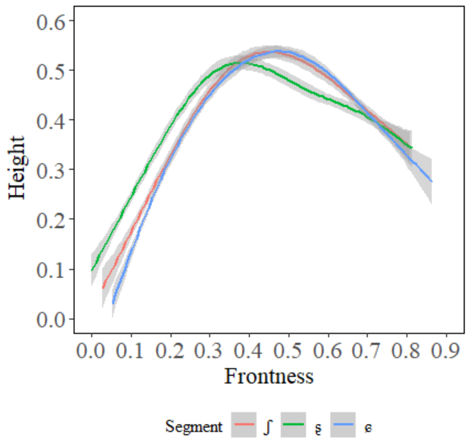

Figure 6: GAMM tongue contours for /ʃ/ (red), /ʂ/ (green), /ɕ/ (blue) for C-level learners of Polish. Tongue tip is on the right.

4. DISCUSSION

In this project, we hypothesized that /ʂ, ɕ/ would initially develop as linked to the /ʃ/ category, and that with experience novel category formation for /ʂ, ɕ/ would occur. We found evidence for early learning, even at the A-level, with differences in tongue body height and tongue dorsum retraction for /ʂ, ɕ/ compared to /ʃ/. Additionally, we found small differences between /ʂ, ɕ/. C-level learners had successfully split the articulation of /ʂ, ɕ/, although we found that /ɕ/ had the same posture as /ʃ/, aside from a slightly more advanced tongue dorsum.

The data when taken together implicates that acquisition and learning for perceptually linked segments happens early [10], but that it is a slow process [4]. It takes at least past the A-level to acquire the three-way sibilant contrast in Polish. Additionally, /ɕ, ʃ/ merge later in the acquisition process, not at the onset. This indicates that even small acoustic-perceptual differences lead to differences in articulation, no matter how minor. The learning trajectory observed here suggests that perceptually similar segments are used to "piggyback" novel articulatory routines. L2 leaners make use of learned patterns and constriction locations and modify them during the acquisition process to eventually develop novel articulatory representations. However, in cases of segments with high goodness-of-fit, it appears that they become articulatorily linked with time [3], which suggests merging of L1 and L2 phonetic spaces.

The fact that our experiment examined FLA acquisition and not SLA acquisition does mean it does not have a direct bearing on the theoretical implications of models like the PAM-L2. However, Tyler [10] does discuss the implications for the PAM-L2 in an FLA context. He suggests that single-category assimilation (i.e., two L2 segments assimilating to the same L1 segment) will yield an even lower probability of acquisition due to increased poverty of stimulus. However, our research finds that this is only partially correct. In single-category assimilation, the segment with the lower goodness-of-fit can and will be acquired in the FLA context, but that the segment high goodness-of-fit will remain articulatory and potentially cognitively linked to the L1 segment. Variation observed in L1 articulation of /ʃ/ may partially account for this effect, as the internalized acoustic-perceptual representation of /ʃ/ may intersect with those same properties for /ɕ/.

7. ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 101018840.

8. REFERENCES

[1] Best, C. T. (1995). A direct realist view of cross-language speech perception. In Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research*, pp. 171-204. Baltimore: York Press.

[2] Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In Munro, M. J. & Bohn, O. S., (eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, pp. 13-34. Amsterdam: John Benjamins.

[3] Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, pp. 233-276. Baltimore: York Press.

[4] Flege, J. E. & Bohn, O.S. (2021). The revised speech learning model (SLM-r). In Wayland, R. (ed.), *Second Language Speech Learning: Theoretical and Empirical Progress*, pp. 3-83. Cambridge University Press.

[5] Mathis, A., Mamidanna, P., Cury, K. M., Abe, Taiga, Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 1281-1289.

[6] R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

[7] Shoemaker, E. (2016). The development of perceptual sensitivity to Polish sibilants at first exposure. In Albright, A. & Fullwood, M. A. (eds.), *Proceedings of the 2014 Annual Meeting on Phonology*. Washington, DC: Linguistic Society of America.

[8] Stefano, C. (2022). *tidymv: Tidy Model Visualisation for Generalised Additive Models*. R package.

[9] Stroup, W. W. (2012). *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. New York: CRC Press.

[10] Tyler, M. D. (2019). PAM-L2 and phonological category acquisition in the foreign language classroom. In Nyvad, A. M., Hejná, M., Højen, A., Jespersen, A. B. & Sørensen, M. H. (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn*, pp. 607-630. Dept. of English, School of Communication & Culture, Aarhus University.

[11] Wickham. H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

[12] Wood, S. N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1): 3-36.

[13] Wrench, A. & Balch-Tomes, J. (2022). Beyond the edge: markerless pose estimation of speech articulators from ultrasound and camera images using DeepLabCut. *Sensors*, 1133.