Check for updates

RESEARCH ARTICLE

# REVISED Foreign language acquisition of perceptually similar segments: evidence from Lower Sorbian

[version 2; peer review: 2 approved]

Phil J. Howson

Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin, 10117, Germany

**v2** **First published:** 13 Apr 2023, **3**:56
https://doi.org/10.12688/openreseurope.14895.1
**Latest published:** 15 Feb 2024, **3**:56
https://doi.org/10.12688/openreseurope.14895.2

## Abstract

Lower Sorbian is a moribund language spoken in Eastern Germany that features a three-way sibilant contrast, /s, ʂ, ɕ/. The vast majority of L1 speakers are above eighty years of age and virtually no young Sorbians learn Lower Sorbian as their first language. There are language revitalization programs in place, but this means that virtually all Lower Sorbian speakers are L2 learners whose first language is German. German, as opposed to Lower Sorbian, has a two-way sibilant contrast, /s, ʃ/. So, Lower Sorbian learners need to acquire a perceptually similar sibilant contrast, /ʂ, ɕ/, that commonly assimilates with a single L1 segment, /ʃ/. The two-to-one assimilation makes acquisition difficult. In this project, I examine the acquisition of the three-way sibilant contrast using ultrasound technology. The ultrasound data revealed that learners in the contemporary context do not produce a distinction between /ʂ, ɕ/ and only learners at an advanced level who had significant exposure to L1 speakers have acquired a three-way sibilant distinction. The findings are put into the context of models of L2 acquisition and generalized implications for foreign language acquisition are discussed.

## Plain language summary

Second language acquisition requires that language learners acquire a novel set of speech segments. For young Sorbians who learn Lower Sorbian as a second language, they must acquire two novel sibilant fricative segments (high frequency noisy segments like /s/). Both of these segments are perceptually similar to the German sibilant fricative common represented with *sch* (e.g., Schlange *snake*). This study explores the acquisition of the Lower Sorbian sibilant fricative contrasts using ultrasound technology. Ultrasound records video of tongue contours at a high-frame rate so that statistical analysis of tongue shapes can be performed. In this project, I examine the tongue contours for Lower Sorbian learners at the beginner, intermediate, and advanced levels of acquisition to observe how

**Open Peer Review**

**Approval Status** ✓ ✓

| | 1 | 2 |
|---|---|---|
| **version 2** (revision) 15 Feb 2024 | ✓ view | ✓ view |
| **version 1** 13 Apr 2023 | ? view | ? view |

1. **Claire Nance**, Lancaster University, Lancaster, UK
**Takayuki Nagamine**, Lancaster University, Lancaster, UK

2. **David Bolter**, Denison University, Granville, USA

Any reports and responses or comments on the article can be found at the end of the article.

tongue shapes for sibilant fricatives are acquired. The results implicate that in a language revitalization context where few L1 speakers are available, the input that learners receive should be augmented with pronunciation and perceptual resources to assist in acquisition. Specific recommendations are provided.

Auf Deutsch

Der Zweitspracherwerb setzt voraus, dass die Sprachlernenden eine Reihe neuer Sprachsegmente erwerben. Junge Sorben, die Niedersorbisch als Zweitsprache lernen, müssen zwei neue Zischlaute (hochfrequente, laute Segmente wie /s/) erwerben. Diese beiden Segmente ähneln in der Wahrnehmung dem deutschen Zischlaut, der häufig mit sch dargestellt wird (z.B. Schlange). In dieser Studie wird der Erwerb der niedersorbischen Zischlautkontraste mittels Ultraschalltechnologie untersucht. Ultraschall zeichnet Videos von Zungenkonturen mit einer hohen Bildrate auf, so dass eine statistische Analyse der Zungenformen durchgeführt werden kann. In diesem Projekt untersuche ich die Zungenkonturen niedersorbischer Lerner auf der Anfänger-, Mittelstufen- und Fortgeschrittenenstufe, um zu beobachten, wie Zungenformen für Zischlaute erworben werden. Die Ergebnisse deuten darauf hin, dass in einem Sprachwiederbelebungskontext, in dem nur wenige L1-Sprecher zur Verfügung stehen, der Input, den die Lernenden erhalten, mit Aussprache- und Wahrnehmungsressourcen ergänzt werden sollte, um den Erwerb zu unterstützen. Es werden spezifische Empfehlungen gegeben.

**Keywords**
Lower Sorbian, sibilant fricatives, language acquisition, phonetics, foreign language acquisition, second language acquisition, ultrasound, endangered languages

This article is included in the Marie-Sklodowska-Curie Actions (MSCA) gateway.

This article is included in the Languages and Literature gateway.

This article is included in the Horizon 2020 gateway.



This article is included in the Experimental Psychology and Cognitive Neuroscience collection.

**Corresponding author:** Phil J. Howson (howson@leibniz-zas.de)

**Author roles: Howson PJ**: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101018840.
*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2024 Howson PJ. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Howson PJ. **Foreign language acquisition of perceptually similar segments: evidence from Lower Sorbian [version 2; peer review: 2 approved]** Open Research Europe 2024, **3**:56 https://doi.org/10.12688/openreseurope.14895.2

**First published:** 13 Apr 2023, **3**:56 https://doi.org/10.12688/openreseurope.14895.1

**REVISED** **Amendments from Version 1**

I would like to thank the reviewers for their suggestions and I have incorporated them throughout this manuscript. Changes to the languages used as well as aspects of the interpretation have been changed in accordance with the suggestions of the reviewers. Suggestions for clarification have also been taken into account during this revision.

**Any further responses from the reviewers can be found at the end of the article**

## Introduction

Lower Sorbian is a west Slavic language spoken in Eastern Germany. It is a moribund language (Moseley, 2012) and is spoken near the border of Poland (Stone, 1993). The vast majority of first language Lower Sorbian speakers are above 80 years of age. Additionally complicating the matter, is that the language situation in Lower Sorbian is quite precarious. The majority of first language speakers do not use their mother tongue in daily communication which has led to certain degrees of language attrition. Additionally, nearly every young speaker of Lower Sorbian is a second language learner and acquires the language at school. For example, the *Witaj* program is a kindergarten curriculum which incorporates Lower Sorbian into the students' education. Following that, many students participate in the *Dolnoserbski gymnazium Chóśebuz*, situated in Cottbus (Marti, 2007). The school completes up to grade 12 and includes Lower Sorbian as a mandatory aspect of education. While the education can be beneficial, there is difficulty finding qualified teachers for the school and due to the advanced age of the L1 speakers, teachers are typically second language speakers themselves.

Lower Sorbian has a cross-linguistically uncommon three-way contrast among sibilant fricatives (approximately less than 6% of languages in the world; Maddieson, 1984) that makes contrasts at the dental/alveolar, /s, z/, retroflex, /ʂ, ʐ/, and alveolopalatal, /ɕ, ʑ/, places of articulation, similar to the contrasts observed in Modern Polish (Żygis, 2003). The contrast contains two sibilants, /ʂ, ɕ/, that share acoustic-perceptual similarities to /ʃ/. Many theories of language acquisition, such as the PAM-L2 (Best & Tyler, 2007) and the SLM-r (Flege & Bohn, 2021), have postulated how different aspects of acoustic-perceptual similarities with L1 segments impacts L2 acquisition. Contrasts such as the three-way contrast, under these theories are the most difficult to acquire due to the acoustic similarities between the segments. This makes Lower Sorbian an excellent language to examine foreign language acquisition of sibilant fricatives.

## Second language acquisition

### The PAM-L2

The Perceptual Assimilation Model of L2 Acquisition (PAM-L2; Best & Tyler, 2007) is an extension of the Perceptual Assimilation Model (PAM; Best, 1995) to second language acquisition. The PAM-L2 is a direct realist model, which assumes that perception is related to the perception of distal articulatory events (i.e., changes in vocal tract configurations), not specific acoustic patterns. Under the view of the PAM-L2, perceptual learning can take place on multiple levels, including phonological, phonetic, or gestural. One way in which category acquisition can occur is when there are two L2 segments that assimilate to two separate L1 segments (two-category assimilation). The PAM-L2 predicts good to excellent discrimination in this context. Learners then continue to acquire L2 vocabulary using the assimilated categories. This leads to a common L1-L2 phonological category for each of the L2 segments. However, in the case that there is a perceptible phonetic difference between L1-L2 pairs of segments, then it is possible that this difference becomes perceptibly stronger for the learner with time. If the differences between L1-L2 pairs becomes perceptible enough, then separate L1 and L2 phonetic categories can emerge. However, if the distinction is not strong enough the learner will not develop separate L2 categories (Tyler, 2019). This process is assumed to occur very early in acquisition, although it may strengthen over time.

Best *et al.* (2009) suggest the process of perceptual attunement is tightly related to vocabulary acquisition. Bundgaard-Nielsen *et al.*'s (2012) *Vocabulary-Tuning Model of L2 Rephonologization* posits that an increase in vocabulary size drives perceptual attunement to L2 phonological structure. Support for this position was found by Bundgaard-Nielsen, Best, & Tayler (2011a, 2011b); however, Tyler (2019) suggests that an increase in vocabulary might support the acquisition of more discriminable L1-L2 pairs but could inhibit less discriminable pairs. Thus, Tyler (2019) suggests that the opportunity for phonetic learning is likely before the L2 vocabulary exceeds 50 words. He supports this position by comparing this to cL1 acquisition; children slow their vocabulary up to around 50 words, and then a rapid increase in vocabulary occurs after (e.g., Nazzi & Bertoncini, 2003). For Tyler (2019), after phonetic attunement takes place, vocabulary increase ramps up dramatically. Thus, the effect of learning a large vocabulary prior to phonetic attunement of difficult to perceive contrasts greatly hinders acquisition.

In the case of Lower Sorbian acquisition, there are two segments of interest that are perceptually similar, /ʂ, ɕ/, which are both perceptually similar to the same L1 segment, /ʃ/. According to the PAM-L2, this is single category assimilation and poor discrimination is predicted. Although, there may still be relative goodness-of-fit difference between the two assimilatory segments that allows learners to discriminate between them and thus acquire the L2 segments. However, the PAM-L2 and its predictions focus on learners in an immersion environment (second language acquisition; SLA); learning a second language in the learner's L1 environments with L2 classes (foreign language acquisition; FLA) has differences from immersion learning (Tyler, 2019). Nonetheless, the PAM-L2 can offer potential insights into foreign language acquisition (FLA). Tyler (2019) suggests that single-category assimilations (i.e., two L2 segments assimilating to the same L1 segment) are even more unlikely to be acquired in the classroom setting. The reason for this is because of a reduced access to consistent stimuli and the phonetic

contrasts that distinguish them. Many second language classrooms are also taught by second language speakers, who may or may not consistently produce the language relevant contrasts, and likely produce contrasts differently than the older generation of L1 speakers. Additionally, there is also extensive acoustic-perceptual input from other second language learners, who also may not produce a target contrast. Tyler (2019) also notes that there is an increase in how fast vocabulary is acquired relative to immersion and L1 contexts which could impact perceptual acquisition.

### The Speech Learning Model

The speech learning model (SLM; Flege, 1995) and the revised speech learning model (SLM-r; Flege & Bohn, 2021) have also been frontrunners of second language acquisition theories. The SLM was primarily designed to account for age related differences in language acquisition, while the SLM-r aims at providing an explanation for how reorganization of the phonetic system occurs over the life-span due to naturalistic L2 learning.

The SLM posits that for late acquiring bilinguals (i.e., someone who acquired two languages as a child, but the second language was acquired later than the first), L2 phonetic learning is influenced by acoustic-perceptual similarities between L2 and L1 phonetics. Thus, L1 and L2 segments become perceptually linked together. Specifically, during L2 learning, segments "map onto" perceptually similar L1 sounds. The ability for L2 learners to discern perceptually linked sounds occurs gradually, rather than rapidly; however, when this occurs, formation of a novel phonetic category can occur.

The mechanisms for novel category formation that guide L1 acquisition are believed to be intact and available for L2 learning. In L1 acquisition, this process is slow and begins as a set of equivalence classes (Kuhl, 1983) that involves grouping acoustically similar sounds together. This development continues long after establishing a phonetic inventory (Lee *et al.*, 1999) and extends at least beyond the age of seven years (Bent, 2014). The SLM proposes that L2 learners of any age form acoustic-perceptual equivalence classes from the statistical properties of the input distributions of their exposure to the target L2. However, unlike L1 category formation, which has no previous language exposure and categories to interfere with it, L2 category formation relies on disruption of L2-to-L1 perceptual links through the ability to discern phonetic differences between perceptually similar L2 and L1 segments. Flege & Bohn (2021) suggest that L2 category formation should take at least as long as L1 category formation.

According to the SLM, L2 category formation depends on the degree of acoustic-perceptual similarity between the L2 segment and the closest L1 sound. That is, the more similar it is to an L1 segment, the harder it will be to form a new L2 category. Additionally, age of acquisition plays a significant role, with older learners having lower probabilities of forming new categories.

The SLM-r (Flege & Bohn, 2021) maintains that there is no difference in how L2 segments are acquired compared to L1 acquisition. The SLM-r posits that observed differences in L2 acquisition, and subsequently, the production and perception of L2 segments arise because L2 sounds are initially linked to L1 segments and serve as a substitute, especially for early learning. The existing L1 phonetic categories interfere with and can even block the formation of novel categories as a result. Additionally, L2 acquisition typically has a different set of input stimulus, which often includes foreign accented L2 speech.

The SLM-r distinguishes itself from the PAM-L2 in that it posits that the delinking process can be facilitated by growth of an L2 lexicon (Bundgaard-Nielsen *et al.*, 2011a; Bundgaard-Nielsen *et al.*, 2011b). While the PAM-L2 believes that growth of the L2 lexicon (beyond perhaps ~50 words) serves to stagnate L2 category formation, at least in the case of hard to discriminate L1 and L2 segments (Tyler, 2019). In this sense, the SLM-r puts forth that category formation is a much longer and drawn-out process (Flege & Bohn, 2021), while the PAM-L2 suggests it is a quicker process with a narrow opportunity for learners to acquire a new category (Tyler, 2019). Additionally, the PAM-L2 posits that learners attenuating to gestural movements in the vocal tract, while the SLM-r suggests that learners pay attention to acoustic differences in the input signal directly. Thus, under the view of the SLM-r, articulation is a matter of better navigation of what vocal tract shapes produce the target acoustic outputs.

### Hypothesis

Based on both the PAM-L2 and SLM-r, the anticipated patterns of L2 segment assimilation is that learners will assimilate both Lower Sorbian, /ʂ, ɕ/, to German /ʃ/. This is due to the acoustic-perceptual similarities between them. The acoustics between the two segments /ʂ, ɕ/ resemble each other across in COG and skewness, having both a lower COG and higher skewness than /s/. Both values also significantly overlapped with each other for /ʂ, ɕ/. The feature in Lower Sorbian that was found to most strongly distinguish /ʂ, ɕ/ from each other was a much higher transitional F2 into the following vowel for /ɕ/ compared to /ʂ/ (Howson, 2015). The lower COG values observed in Lower Sorbian, tend to match cross-linguistic COG associated with /ʃ/ (Żygis, 2010) and COG and skewness measures associated with German /ʃ/ (Weirich & Simpson, 2015). Thus, I expect that low level (i.e., A-level) learners will share tongue contours for /ʂ, ɕ/ and that they will both resemble /ʃ/. It remains possible that there are still goodness-of-fit (or phonetically discernible) differences between /ʂ, ɕ/ and /ʃ/. More specifically, /ɕ/ has formant transitions and spectral characteristics similar to /ʃ/, while /ʂ/ has similar spectral characteristics, but different formant transitions. Thus, I expect that more advanced learners of Lower Sorbian will initially differentiate /ʂ/ from /ɕ, ʃ/ because of the stronger acoustic-perceptual dissimilarities. In terms of the PAM-L2 (Best & Tyler, 2007), the assumption is that learners are perceiving articulatory gestures and vocal tract changes, not more abstract acoustic characteristics. The implication of this is that as learners become more advanced, they become better at retrieving the articulatory movements necessary to produce a contrast. The expectation is that gradual improvement in the articulation of L2 segments should occur. In terms of the SLM-r (Flege & Bohn, 2021),

there is a similar expectation. As learners' acoustic-perceptual representation improves, so too should articulation. However, because perceptual (and articulatory) dissimilarities may take more time to pick up on (Flege & Bohn, 2021), I predict that only more advanced learners will have acquired these contrasts.

## Methods

### Study design

The study design was an articulatory examination of tongue contours using ultrasound data collection techniques. Participants read sentences in Lower Sorbian with the target segments in them while they were being recorded with ultrasound. Tongue contours were compared using Generalized Additive Mixed Models (GAMMs). Data recording took place from March 27th, 2020 until April 1st, 2020 in Cottbus, Germany for the L2 learners. The advanced L2 learners, C04 and C05, were recorded at the University of Leipzig in Germany from April 4th until April 8th. The L1 speakers were recorded from July 18th, 2022 until July 22nd, 2022 in Cottbus, Germany.

### Participants

As a baseline, 1 bilingual Sorbian/German speaker (male, 24), and 1 late-acquiring bilingual speaker of Sorbian (female, 40; age of first acquisition: 5) were recorded using ultrasound. These participants were chosen for this study because they both had significant input stimuli during the learning process from L1 speakers. Both speakers had input from L1 speaking relatives and additionally the older speaker attended the Sorbian school at a time when L1 Sorbian speaking teachers were active. Additionally, at the time of recording this data, few L1 speakers remain, and the advanced age of potential participants (above 80 years of age) makes ultrasound data especially difficult to record and interpret.

The criteria for language learner selection were that participants attended *Dolnoserbski gymnazium Chóśebuz* in Cottbus and were currently engaged in their language learning program. All participants had a first language of German. Participants were recruited for all three skill levels, A-, B-, and C-level learners based on a scaling system like the CEFR. Their skill level at the time was based on class they attended for Sorbian language at *gymnazium Chóśebuz* at the time of recording. Year learning Lower Sorbian ranged from approximately 6–17 years and was not necessarily reflective of the level of the speaker (i.e., more years did not necessarily reflect higher proficiency). Participant saturation was determined based on typical sample sizes for ultrasound studies. For baseline speakers, participants were selected on the basis that they had early exposure to Lower Sorbian and learned it in a natural setting (i.e., through hearing Lower Sorbian), although both participants also received an education in the Lower Sorbian school system. The Lower Sorbian speaking community is small, especially with respect to L1 speakers and so as many L1 speakers as possible were recruited. The L2 learners consisted of 4 A-level, 6 B-level, and 5 C-level learners. All of these participants were ages 17-18. Two of the C-level speakers were extremely advanced. Their ages were 35 and 56. All participants had no self-reported history of speech or hearing disorders.

### Procedure

All participants read and signed the ethics forms prior to the experiment. They were also verbally informed as to the structure of the experiment and informed of their primary rights as a participant, including that their de-identified data would be shared with other researchers, and that they could refuse data sharing if they wished.

Data for the baseline speakers were recorded in a quiet room in the *Serbski Institut* in Cottbus, Brandenburg. Data for the Lower Sorbian learners were recorded in a quiet room at *Dolnoserbski gymnazium Chóśebuz* in Cottbus, Brandenburg. Ultrasound data were recorded with the Micro system from Articulate Assistant Advanced (AAA). I used the 20mm Radius probe with a 92 degrees field of view (FOV). Data was recorded at an average of 80 frames per second (fps). An ultrasound stabilization headset (Articulate Instruments Ltd., 2008) was also used to prevent movement of the ultrasound probe.

Participant forms were filled out prior to participation, including the questionnaire and consent forms. In order to pseudo-anonymize participant data, participants were assigned a letter and number combination which corresponded to their skill level and the order in which they participated (e.g., C05 = the fifth C-level learner recorded; LS01 = the first L1 Lower Sorbian speaker recorded). Stimuli were presented using the AAA software package. Additionally, audio and video were synchronized and recorded using the AAA software. The full stimuli list is presented in Table 1. Stimuli were presented in a carrier phrase to facilitate more natural production. The carrier phrase was "Grońśo *target* hyšći raz" (please *target* say again). Stimuli were presented in a pseudorandomized order. Each participant produced 6 articulations of each segment in each of the three vocalic environments. This gives a total of 108 tokens for the L1 speakers (2 speakers × 3 segments × 3 vowels × 6 repetitions), 216 tokens for the A-level learners (4 speakers × 3 segments × 3 vowels × 6 repetitions), 324 tokens for the B-level learners (6 speakers × 3 segments × 3 vowels × 6 repetitions), and 270 tokens for the C-level (5 speakers × 3 segments × 3 vowels × 6 repetitions).

### Ethical considerations

Ethical approval was obtained from the Deutsche Gesellschaft für Sprachwissenschaft (DGfS #2021-13-220106) and informed written consent was obtained from all participants for the use and publication of their data.

**Table 1. Stimuli.**

| | *i* | | *a* | | *u* | |
|---|---|---|---|---|---|---|
| s | ćis | *yew tree* | cas | *time* | kus | *bite* |
| ʂ | liš | *excessive* | praš | *leprosy* | duš | *soul* |
| ɕ | biś | *beat* | braś | *take* | duś | *beat* |

### Analysis

Tongue contours were manually traced using AAA software (v220.04.01) at the temporal midpoint of the fricative. The midpoint was identified based on the duration of the fricative, where the onset was measured as the offset of formants and periodic sound waves associated with the preceding vowel and the offset was determined as the reduction in aperiodic noise and dissipation of frication on the spectrogram associated with the fricative. Polar coordinates were then extracted. Tongue contours were then compared using a custom script (Heyne *et al.*, 2019) for GAMM analysis of polar coordinates in R (R Core Team, 2023). GAMMs were performed using the mgcv package (Wood, 2011), which also provides summary statistics. Tongue contours were first compared for L1 speakers to provide a baseline for comparison. Tongue contours were then compared for each language group (A, B, and C). Group C was split into two: C-level and highly advanced C-level. GAMMs were performed with parametric fixed effects for segment (3 levels: /s, ʂ, ɕ/) and environment (3 levels: /i, a, u/). The interaction between segment and environment was also included. A smoothing variable was also included for segment and the interaction between segment and environment. I included a factor smooth (i.e., a random effect) for the interaction between segment and speaker. The dependent variable was r, or the angle of the coordinate from the probe origin, and each smooth included Theta, which is the distance of the coordinate from the probe origin. For all smooths, cubic regression was used. The equation I used is printed in (1).

(1) r ~ Segment * Environment + s(Theta, bs = "cr", k = 25) + s(Theta, by = Segment, bs = "cr", k = 25) + s(Theta, by = Segment : Environment, bs = "cr", k = 25) + s(Theta, by = Segment : Speaker, bs = "fs", k = 25, m = 1)

I also performed an individual analysis for each speaker, which includes a factor smooth for repetition. Because of differences in speaker tongue sizes, k (knots) was set to 20 in order to maintain consistency across all speakers. The equation is printed in (2).

(2) r ~ Segment * Environment + s(Theta, bs = "cr", k = 20) + s(Theta, by = Segment, bs = "cr", k = 20) + s(Theta, by = Segment : Environment, bs = "cr", k = 20) + s(Theta, by = Rep, bs = "fs", k = 20, m = 1)

Data was then visualized with plotly (Sievert, 2020) and a custom script (Heyne *et al.*, 2019) to identify areas of statistical significance.

## Results

### L1 Speakers

Figure 1–Figure 2 below present the GAMM smooths for the L1 speakers of Lower Sorbian and Table 2–Table 3 present the approximate significance for the interaction between theta and segment. For full statistical print-outs, see *Extended data* (Howson, 2023). The adjusted $R^2$ for the models were 0.979 and 0.983.

The GAMMs for the L1 speakers revealed a significant difference between all three segments, /s, ʂ, ɕ/. The tongue dorsum was most retracted for /s, ʂ/ and was more advanced for /ɕ/. The tongue contours for /s, ʂ/ were similar, but the tongue body was more raised for /ʂ/. /ɕ/ had the most raised tongue body, but it was not much more raised than /ʂ/.

### A-Level learners

Figure 3 below presents the GAMM smooths for A-level learners of Lower Sorbian and Table 4 presents the approximate significance for the interaction between theta and segment. For individual plots and full statistical print-out, see *Extended data* (Howson, 2023). The adjusted $R^2$ for the model was 0.946.

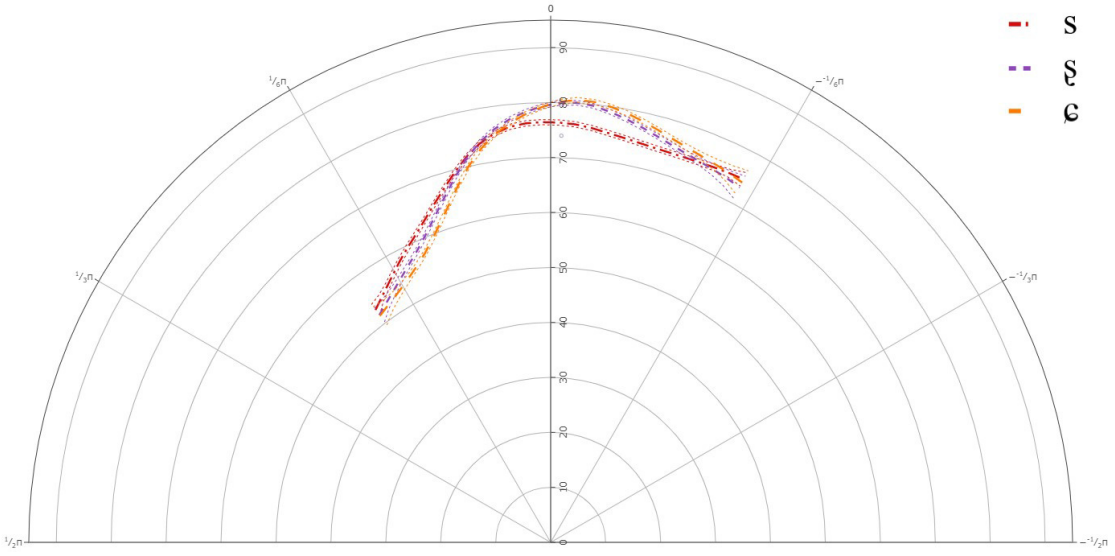


**Figure 1. Tongue contours LS Speaker LS101 for /s/ (red), /ʂ/ (purple), and /ɕ/ (yellow).**

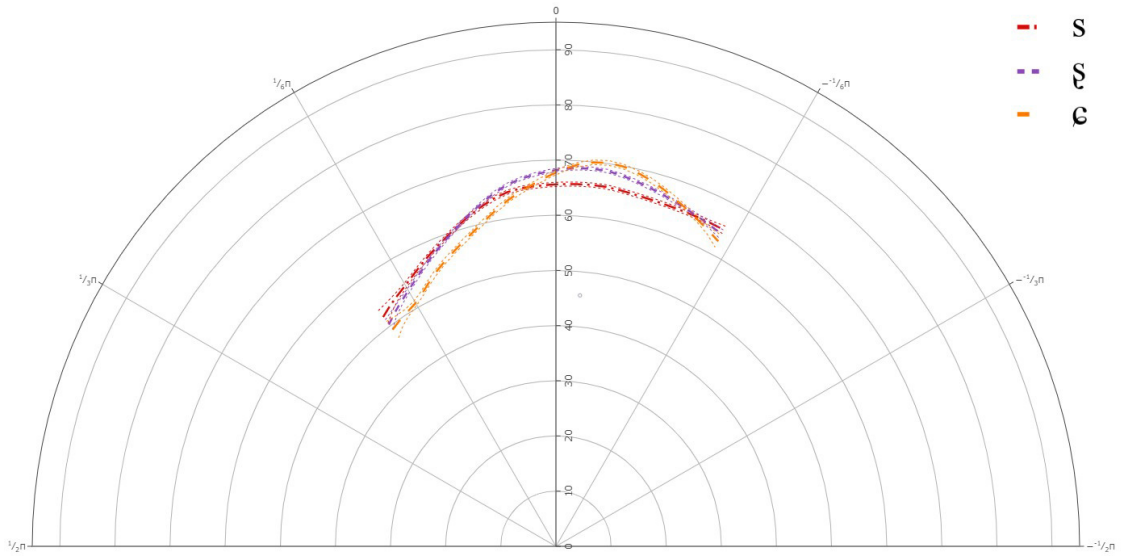


**Figure 2.** **Tongue contours LS Speaker LS102 for /s/ (red), /ʂ/ (purple), and /ɕ/ (yellow).**

**Table 2.** **Approximate significance of smoothing term Theta by Segment for L101.**

| | edf | ref.df | F | p-value |
|---|---|---|---|---|
| s(Theta) | 12.20 | 14.10 | 140.821 | < 0.001 |
| s(Theta): /s/ | 7.38 | 9.09 | 8.039 | < 0.001 |
| s(Theta): /ʂ/ | 1 | 1 | 1.399 | 0.237 |
| s(Theta): /ɕ/ | 3.59 | 4.73 | 3.404 | 0.009 |

**Table 3.** **Approximate significance of smoothing term Theta by Segment for L102 speakers.**

| | edf | ref.df | F | p-value |
|---|---|---|---|---|
| s(Theta) | 9.388 | 10.967 | 239.21 | < 0.001 |
| s(Theta): /s/ | 2.835 | 3.532 | 104.195 | < 0.001 |
| s(Theta): /ʂ/ | 1 | 1 | 109.831 | < 0.001 |
| s(Theta): /ɕ/ | 4.04 | 5.103 | 1.436 | 0.1879 |

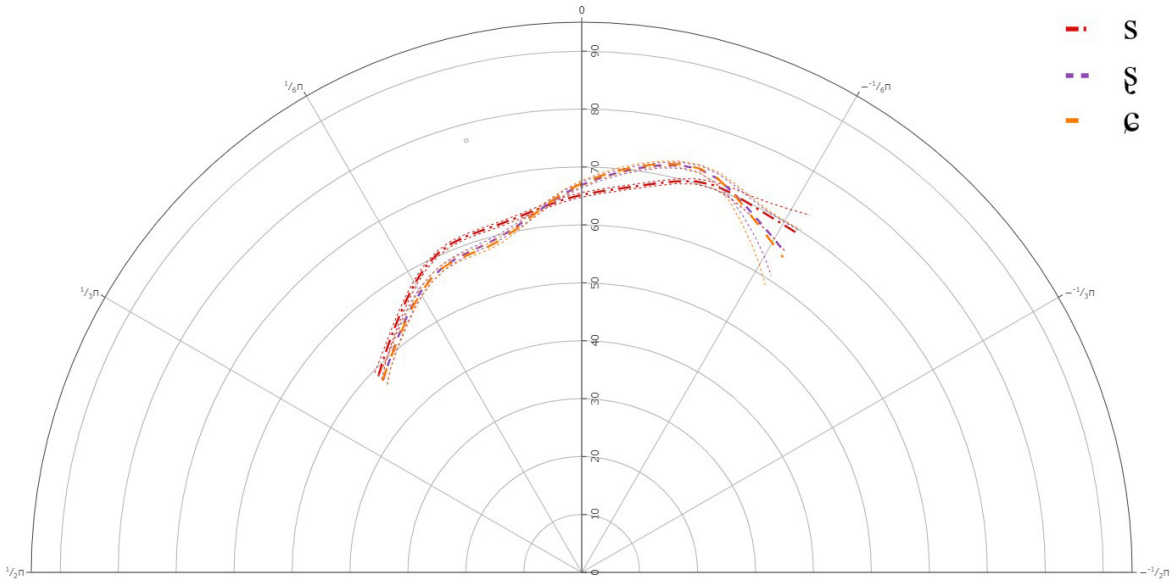


**Figure 3.** **Tongue contours for A-level learners for /s/ (red), /ʂ/ (purple), and /ɕ/ (yellow).**

**Table 4. Approximate significance of smoothing term Theta by Segment for A-level learners.**

| | edf | ref.df | F | p-value |
|---|---|---|---|---|
| s(Theta) | 11.90 | 14.288 | 17.053 | < 0.001 |
| s(Theta): /s/ | 1 | 1 | 8.995 | 0.003 |
| s(Theta): /ʂ/ | 1 | 1 | 0.264 | 0.607 |
| s(Theta): /ɕ/ | 4.29 | 5.45 | 1.829 | 0.094 |

The general results for the A-level learners revealed that there was a significant difference between /s/ and /ʂ, ɕ/, but not between /ʂ/ and /ɕ/. This suggests that learners at the A-level share one tongue contour for their pronunciations of /ʂ, ɕ/. The general tongue contours indicated a more retracted tongue dorsum for /s/, than for /ʂ, ɕ/. The contours for /ʂ, ɕ/ had a slightly more advanced dorsum, with a raised tongue body, resembling /ʃ/, which is present in the L1 German.

Individual results revealed significant deviations in learners' articulation of /ʂ, ɕ/, when compared against the general tongue contour from the group level GAMM. Although it should be noted that none of the individual plots revealed that any of the learners had acquired the three-way contrast, there was significant variation in their articulation of /ʂ, ɕ/.

### B-Level learners

Figure 4 below presents the GAMM smooths for B-level learners of Lower Sorbian and Table 5 presents the approximate significance for the interaction between theta and segment. For individual plots and full statistical print-out, see *Extended data* (Howson, 2023). The adjusted $R^2$ for the model was 0.957.

The GAMM results indicated that there was a significant difference between /s/ and /ʂ, ɕ/, but not between /ʂ/ and /ɕ/. This suggests that like the A-level learners, the B-level learners also have not acquired the three-way contrast between /s, ʂ, ɕ/ with respect to their articulation. The general tongue contours reveal a more retracted tongue dorsum for /s/, with a lower tongue body than /ʂ, ɕ/. The contours for /ʂ, ɕ/ had more rounded tongue shape, with more fronting, and more posterior tongue body raising than for /s/.

Individual results also revealed variation in the articulation of /ʂ, ɕ/, although as with the A-level learners, there were no significant differences between /ʂ, ɕ/. In most cases, the tongue dorsum was more drawn back for /s/ and was more advanced for /ʂ, ɕ/. In some cases, the more anterior part of the tongue body was raised for /ʂ, ɕ/, while for some learners the more posterior part of the tongue body or the entire tongue body for /ʂ, ɕ/ was more raised than /s/. This suggests that learners at the B-level continue to use the same segment in place for both /ʂ, ɕ/, although there was a great deal of variation in its realization.

### C-Level learners

Figure 5 below presents the GAMM smooths for C-level learners of Lower Sorbian and Table 6 presents the approximate significance for the interaction between theta and segment. For individual plots and full statistical print-out, see *Extended data* (Howson, 2023). The adjusted $R^2$ for the model was 0.979.

The GAMM results for C-level learners indicated that there was a significant difference between /s/ and /ʂ, ɕ/, but not between /ʂ/ and /ɕ/. This suggests that learners of Lower Sorbian at all levels have not acquired the three-way contrast. The individual results showed variation in articulation between speakers and in the case of the C-level learners, none of them showed the same backing of the tongue dorsum for /s/ compared to /ʂ/ and /ɕ/.

### Highly advanced C-Level learners

Figure 6 and Figure 7 below presents the GAMM smooths for highly advanced C-level learners of Lower Sorbian and Table 7 and Table 8 presents the approximate significance for the interaction between theta and segment. The adjusted $R^2$ for the models were 0.970 and 0.972, respectively. The *Extended data* (Howson, 2023) presents the full statistical printouts of both models.

In both cases, the learners acquired a three-way contrast for /s, ʂ, ɕ/; however, the realization of /ʂ, ɕ/ varied for both speakers. In both cases, /s/ had the lowest tongue body, accompanied by retracted tongue dorsum. /ʂ/ for C04 had a similar degree of retraction for the tongue dorsum as /s/, with a more raised tongue body. The tongue shape for /ʂ/ was faithful to the L1 pronunciation. /ɕ/ for C04 had a low and advanced tongue dorsum. This shape is likely due to the high degree of anterior tongue body advancement and raising. This tongue shape deviated significantly from the L1 pronunciation for /ɕ/. /ʂ/ for C05 had even more tongue dorsum retraction than /s/, with a raised posterior tongue body that had a downward sloping anterior tongue body. /ɕ/ for C05 had a more advanced tongue dorsum and tongue body. The posterior tongue body was raised, with a downward sloping anterior tongue body.

## Discussion

The analysis revealed that for L1 speakers, there is a 3-way contrast intact, but that for L2 learners, substitution of

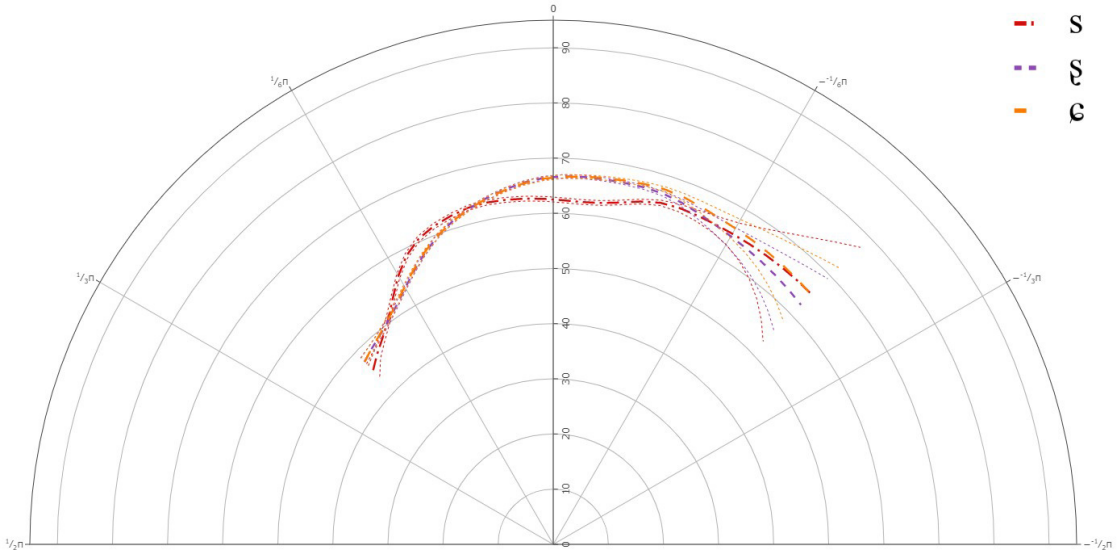


**Figure 4.** **Tongue contours for B-level learners for /s/ (red), /ʂ/ (purple), and /ɕ/ (yellow).**

**Table 5.** **Approximate significance of smoothing term Theta by Segment for B-level learners.**

| | edf | ref.df | F | p-value |
|---|---|---|---|---|
| s(Theta) | 10.54 | 12.59 | 424.179 | < 0.001 |
| s(Theta): /s/ | 0 | 0 | 0.07 | 0.997 |
| s(Theta): /ʂ/ | 1.43 | 1.714 | 0.197 | 0.738 |
| s(Theta): /ɕ/ | 1 | 1 | 0.01 | 0.919 |

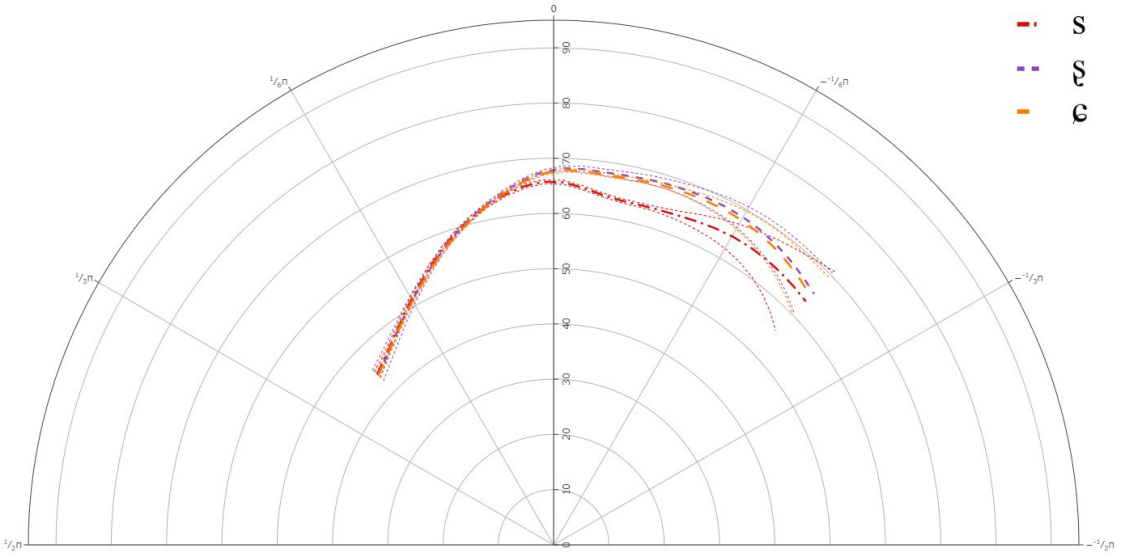


**Figure 5.** **Tongue contours for C-level learners for /s/ (red), /ʂ/ (purple), and /ɕ/ (yellow).**

**Table 6. Approximate significance of smoothing term Theta by Segment for C-level learners.**

| | edf | ref.df | F | p-value |
|---|---|---|---|---|
| s(Theta) | 12.220 | 14.380 | 78.058 | < 0.001 |
| s(Theta): /s/ | 6.516 | 7.923 | 9.604 | < 0.001 |
| s(Theta): /ʂ/ | 0 | 0 | 0.360 | 0.994 |
| s(Theta): /ɕ/ | 1 | 1 | 0.042 | 0.838 |

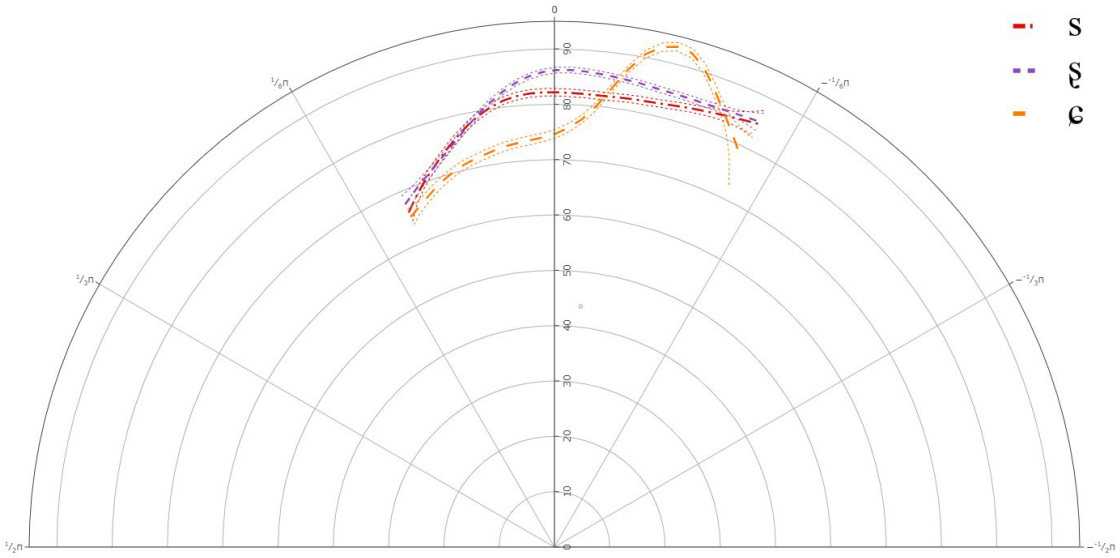


**Figure 6. Tongue contours for C04 for /s/ (red), /ʂ/ (purple), and /ɕ/ (yellow).**

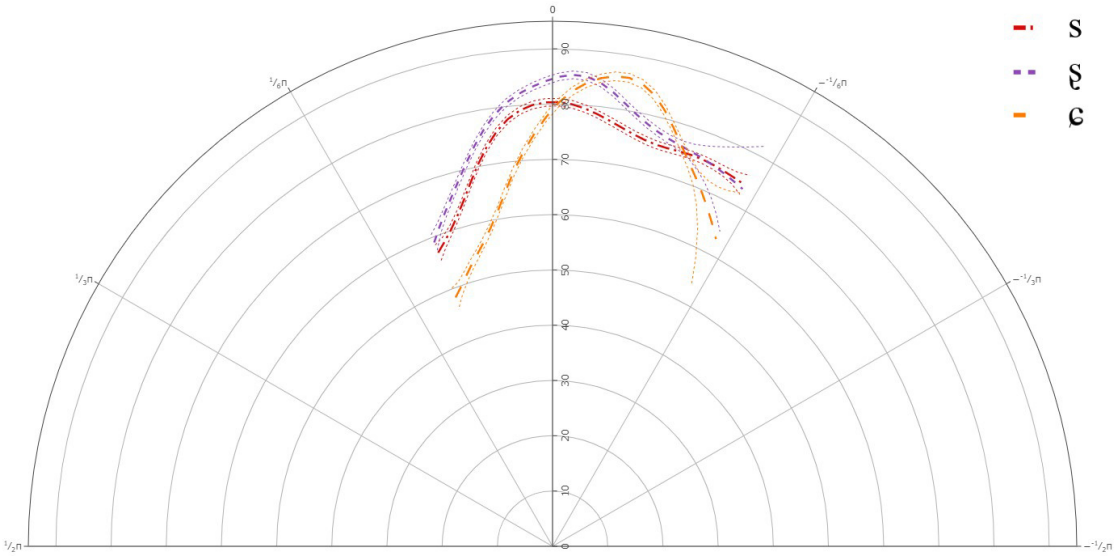


**Figure 7. Tongue contours for C05 for /s/ (red), /ʂ/ (purple), and /ɕ/ (yellow).**

**Table 7. Approximate significance of smoothing term Theta by Segment for C04.**

| | edf | ref.df | F | p-value |
|---|---|---|---|---|
| s(Theta) | 8.78 | 10.45 | 67.019 | < 0.001 |
| s(Theta): /s/ | 0 | 0 | 0.385 | 0.997 |
| s(Theta): /ʂ/ | 1 | 1 | 0.116 | 0.734 |
| s(Theta): /ɕ/ | 11.90 | 13.98 | 47.547 | < 0.001 |

**Table 8. Approximate significance of smoothing term Theta by Segment for C05.**

| | edf | ref.df | F | p-value |
|---|---|---|---|---|
| s(Theta) | 8.618 | 10.029 | 4.493 | < 0.001 |
| s(Theta): /s/ | 5.127 | 6.194 | 1.851 | 0.087 |
| s(Theta): /ʂ/ | 3.495 | 4.184 | 2.407 | 0.0475 |
| s(Theta): /ɕ/ | 8.284 | 10.099 | 8.737 | < 0.001 |

both /ʂ, ɕ/ for /ʃ/ occurred even for learners at the C-level. This was true for all learners except the most highly advanced C-level speakers. Both the PAM-L2 and SLM-r predict that such an assimilation would occur and that the contrast should be difficult to acquire because of the acoustic-perceptual similarity between the two. Nevertheless, in an immersion context, both models predict that it is possible for learners to acquire these contrasts. However, the observed learners were in a foreign-language context and the educators were primarily second language learners themselves. This means that there was likely varied input and the lack of access to L1 input may have greatly hindered their acquisition. However, it should be noted that one limitation of the study is the relatively small wordlist which makes it more difficult to assess category formation.

Learners in this dataset were not given specific pronunciation instructions. What this means is that learners only had access to any existing internal language learning mechanisms. Flege (1995) predicts that the mechanisms involved in L1 acquisition are still available for L2 learners and the evidence presented here does not disprove this but, at the least, it suggests that L1 interference in the acoustic-perceptual space (Kuhl, 1991; Kuhl *et al.*, 1992; Kuhl & Iverson, 1995) significantly interferes with language learning mechanisms if they are still accessible. The result is that the distortion of the perceptual space inhibits perceptual learning of L1 assimilated segments and thus hinders any alteration in articulatory patterns and novel category formation. As a result, learners have a merging between /ʂ, ɕ/ in Lower Sorbian into their L1 German /ʃ/ category. One caveat to note is that the current L2 instructors do not have the level of fluency as the L1 instructors that the two advanced speakers (C04 and C05) had access to. As such, it is difficult to interpret how much input for the three-way sibilant contrast (if any) learners received. It is clear from discussions with the learners that pronunciation lessons are not a regular part of the curriculum. It remains very possible the lack of acquisition of the three-way contrast is predominantly due to lack of the three-way contrast in the input for learners. In short, the development of the language in the context of endangerment, revitalization, and its status as a minority language in the German context has possibly led to an inventory shift away from a three-way contrast to a more typical two-way contrast like the one observed in Upper Sorbian (Howson, 2017). If the desire of the community is to maintain specific speech patterns present in older L1 speakers, then from a practical standpoint, it seems that additional resources need to be committed to this achieve this goal. This is at least true in the foreign language context but would undoubtably assist in immersion contexts as well. Idealistically, this would involve perceptual training that would cater to the speaker's L1 segments and assist in training the learner in distinguishing their existing L1 categories and L2 categories. This would also be accompanied by specific instructions on how the target segments are produced. Ultrasound technology has been used in this context both for direct visualization of how the learner produces the contrast themselves and how they should produce the contrasts (Antolík *et al.*, 2019) as well as providing visual instruction guides for learners (Bliss *et al.*, 2018). This indicates that in language learning and preservations efforts, a multitude of resources should be employed to assist second language learners in acquisition of L2 segments.

There is also the case of the two highly advanced speakers who have acquired a three-way contrast in their L2 speech. First and foremost, the speakers are much older, and as a result had a significant amount of input from L1 speakers during their acquisition processes. The increased access to authentic speech could have contributed to the eventual formation of novel categories. However, it is also important to note that Lower Sorbian /ɕ/ for both speakers appears to have been assimilated into the German /ʃ/ category. In terms of the PAM-L2, this would suggest a better goodness-of-fit match between /ɕ/ and /ʃ/. While, /ʂ/ has similar spectral qualities, the formant transitions are much more similar between /ɕ/ and /ʃ/, while also having similar spectral qualities. This suggests that at least a certain degree of perceptual dissimilarity must be present for the acquisition process to take place. When a segment is "good enough," rather than forming a novel category, the L1 category becomes linked (in SLM terms). Whether or not L1 phonological patterns are imported into L2 or if L2 influences L1 phonological patterns is unclear. Additionally, it remains unclear if phonetic linking occurs with a decoupling of phonological behaviour. As a result, the interaction in phonological patterning and effects between L1 and L2 linked segments needs to be explored further.

## Data availability

### Underlying data

OSF: L2 Lower Sorbian. https://doi.org/10.17605/OSF.IO/DAURS. (Howson, 2023)

This project contains the following underlying data:

- lower_sorbian_dataset.xlsx (data used in the statistical analyses.)

- participant_data.pdf (data from the participant questionnaires.)

## Extended data

This project contains the following extended data:

- extended_data_for_Howson_2023.pdf (full statistical print outs and plots for all the models presented in this paper.)


## Acknowledgements

I would like to thank the *Dolnoserbski gymnazium Chóśebuz*, *the Serbski Institut* in Cottbus, and the *Institut für Sorabistik* in Leipzig for their tremendous help in scheduling participants and providing recording space for data collection. Additionally, I would like to thank members of the *Institut für Sorabistik* for advice on an appropriate stimuli set.


## References


Antolík TK, Pillot-Loiseau C, Kamiyama T: **The effectiveness of real-time ultrasound visual feedback on tongue movements in L2 pronunciation training Japanese learners' progress on the French vowel contrast /y/-/u/.** *J Second Lang Pronunciation.* 2019; **5**(1): 72–97.
**Publisher Full Text**

Articulate Instruments Ltd: **Ultrasound Stabilisation Headset Users Manual: Revision 1.4**. Edinburgh, UK: Articulate Instruments Ltd, 2008.
**Reference Source**

Bent T: **Children's perception of foreign-accented words.** *J Child Lang.* 2014; **41**(6): 1334–1355.
**PubMed Abstract** | **Publisher Full Text**

Best CT: **A direct realist view of cross-language speech perception**. In: Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research*. Baltimore: York Press, 1995; 171–204.
**Reference Source**

Best CT, Tyler MD: **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*. Amsterdam: John Benjamins, 2007; 13–34.
**Reference Source**

Best CT, Tyler MD, Gooding TN, *et al.*: **Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words.** *Psychol Sci.* 2009; **20**(5): 539–542.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Bliss H, Bird S, Cooper PA, *et al.*: **Seeing Speech: Ultrasound-based Multimedia Resources for Pronunciation Learning in Indigenous Languages.** *Lang Doc Conserv.* 2018; **12**: 318–338.
**Reference Source**

Bundgaard-Nielsen RL, Best CT, Kroos C, *et al.*: **Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility.** *Appl Psycholinguist.* 2012; **33**(3): 643–664.
**Publisher Full Text**

Bundgaard-Nielsen RL, Best CT, Tyler MD: **Vocabulary size is associated with second-language vowel perception performance in adult learners.** *Stud Second Lang Acquis.* 2011a; **33**(3): 433–461.
**Publisher Full Text**

Bundgaard-Nielsen RL, Best CT, Tyler MD: **Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories.** *Appl Psycholinguist.* 2011b; **32**(1): 51–67.
**Publisher Full Text**

Flege JE: **Second language speech learning: Theory, findings, and problems**. In: Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Baltimore: York Press, 1995; 233–276.
**Reference Source**

Flege JE, Bohn OS: **The revised speech learning model (SLM-r)**. In: Wayland, R. (ed.), *Second Language Speech Learning: Theoretical and Empirical Progress*. Cambridge University Press, 2021; 3–83.
**Reference Source**

Heyne M, Derrick D, Al-Tamimi J: **Native language influence on brass instrument performance: An application of generalized additive mixed models (GAMMs) to midsagittal ultrasound images of the tongue.** *Front Psychol.* 2019; **10**: 2597.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Howson P: **An acoustic examination of the three-way sibilant contrast in Lower Sorbian.** *Interspeech.* Dresden, Germany, 2015; 2670–2674.
**Publisher Full Text**

Howson P: **Upper Sorbian.** *J Int Phon Assoc.* 2017; **47**(3): 359–367.
**Publisher Full Text**

Howson PJ: **L2 Lower Sorbian**. 2023.
**http://www.doi.org/10.17605/OSF.IO/DAURS**

Kuhl PK: **Perception of auditory equivalence classes for speech in early infancy.** *Infant Behav Dev.* 1983; **6**(2–3): 263–285.
**Publisher Full Text**

Kuhl P: **Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not.** *Percept Psychophys.* 1991; **50**(2): 93–107.
**PubMed Abstract** | **Publisher Full Text**

Kuhl PK, Williams KA, Lacerda F, *et al.*: **Linguistic experience alters phonetic perception in infants by 6 months of age.** *Science.* 1992; **255**(5044): 606–608.
**PubMed Abstract** | **Publisher Full Text**

Kuhl P, Iverson P: **Linguistic experience and the "perceptual magnet effect"**. In: Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press, 1995; 121–154.
**Reference Source**

Lee S, Poramianos A, Narayanan S: **Acoustics of children's speech: Developmental changes of temporal and spectral parameters.** *J Acoust Soc Am.* 1999; **105**(3): 1455–1468.
**PubMed Abstract** | **Publisher Full Text**

Maddieson I: **Patterns of sounds.** Cambridge University Press, 1984.
**Reference Source**

Marti R: **Lower Sorbian — twice a minority language.** *Int J Sociol Lang.* 2007; **2007**(183): 31–51.
**Publisher Full Text**

Moseley C: **The UNESCO atlas of the world's languages in danger: Context and process**. (World Oral Literature Project Occasional Paper 5). Cambridge: University of Cambridge, 2012.
**Reference Source**

Nazzi T, Bertoncini J: **Before and after the vocabulary spurt: Two modes of word acquisition?** *Dev Sci.* 2003; **6**(2): 136–142.
**Publisher Full Text**

R Core Team: **R: A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna, Austria, 2023.
**Reference Source**

Sievert C: **Interactive Web-Based Data Visualization with R, plotly, and shiny**. Chapman and Hall/CRC Florida, 2020.
**Reference Source**

Stone G: **Sorbian (Upper and Lower)**. In: Comrie, B. & Corbett, G. G. (eds.), *The Slavonic languages*. London & New York: Routledge, 1993; 759–794.

Tyler MD: **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*. Dept. of English, School of Communication & Culture, Aarhus University, 2019; 607–630.
**Reference Source**

Weirich M, Simpson A: **Gender-specific differences in sibilant contrast realizations in English and German.** In: *Proceedings of the 18th International Congress of Phonetic Sciences.* The University of Glasgow, Glasgow, UK, 2015; 1–4.
**Reference Source**

Wood SN: **Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models.** *J R Stat Soc (B).* 2011; **73**(1): 3–36.
**Publisher Full Text**

Żygis M: **Phonetic and phonological aspects of Slavic sibilant fricatives**. In: Tracy Hall, A. & Hamann, S. (eds.), *Papers in Phonology and Phonetics*. (ZAS Papers in Linguistics), Berlin: ZAS, 2003; **32**: 175–213.
**Publisher Full Text**

Żygis M: **On changes in Slavic sibilant systems and their perceptual motivation.** In: Recasens, D., Sánchez Miret, F., & Wireback, K. J. (eds.). *Experimental Phonetics and Sound Change*. München: Lincom. 2010; 115–138.
**Reference Source**

# Open Peer Review

## Current Peer Review Status: ✔ ✔

**Version 2**

Reviewer Report 25 March 2024

https://doi.org/10.21956/openreseurope.18664.r38172

© **2024 Nance C.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Claire Nance**
Linguistics and English Language, Lancaster University, Lancaster, England, UK

The author has now addressed our main points.

I would still recommend making the following minor changes:

I still find the distinction between 'L1 speakers' and 'baseline speakers' confusing I'm afraid. It is clear that it would not be possible to ultrasound 80+ year old L1 speakers. My understanding is that the 'baseline' speakers in this study are sequential bilinguals who have achieved very high proficiency in Sorbian, with lots of L1 Sorbian input. That's completely fine. But it's confusing to refer to them as 'L1 speakers'. Suggest changing to 'Baseline speakers' or similar consistently throughout.

- 2nd paragraph of Participants
- 3rd paragraph of Procedure
- 1st paragraph of Analysis
- Results
- Discussion

Without this change in wording, it's a bit confusing to read the study I'm afraid.

The response document says 'L2 learners' has been changed to 'L2 users'. This doesn't seem consistent through the manuscript. For me, although this is a terminological point, it represents a change in focus. These speakers are the future of Sorbian and are using the language, some of them as prominent community figures by the sound of it. It seems unfair to still call them 'learners', especially at C2 level. This distinction has been discussed for some time in SLA e.g. Cook (1999; https://doi.org/10.2307/3587717) and also applied to language revitalisation contexts e.g. O'Rourke & Ramallo (2013; https://doi.org/10.1017/S0047404513000249) and Ó hIfernáin (2015; https://doi.org/10.1515/ijsl-2014-0031).

I would consider this an essential change for describing at least the highly advanced C-level

speakers.

I appreciate the clarifications made to the description of the statistics. For the significance testing, I can see that the model summary from the mgcv package has been used for the approximate significance values. It would be good to explain this approach over, for example, model comparison (Sóskuthy 2017; http://eprints.whiterose.ac.uk/113858/). I'm not suggesting that the models should be redone. Indeed this might not be practical or appropriate for the current study. But it would be good to know why this strategy for significance testing was chosen. Perhaps I am not understanding correctly.

The word list is quite short, which has implications for the interpretation of the results. It is hard to know whether speakers' productions relate to individual words, or relate to production of a system. This could be clearly highlighted in the methods section as well as a short sentence in the Conclusion.

Ultrasound data were not rotated to the occlusal plane. I accept that not everything can be done, and this can't be changed after data collection! But it should be mentioned in the text.

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Bilingualism, articulatory phonetics, ultrasound and acoustic analysis, sociolinguistics of language revitalisation (Scottish Gaelic in particular).

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 25 March 2024

https://doi.org/10.21956/openreseurope.18664.r38173

**© 2024 Bolter D.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**David Bolter**
Denison University, Granville, Ohio, USA

As I noted in my original review, this article contains an articulatory study of the acquisition of the (voiceless) sibilant system of Lower Sorbian, including /s ʂ ɕ/, by speakers whose L1 or dominant language is German and whose baseline (voiceless) sibilant system includes /s ʃ/. In general, those speakers, whatever level they have on the Common European Framework of Reference, do not produce the unfamiliar /ʂ ɕ/ with distinct tongue profiles, as evidenced by the ultra-sound imaging undertaken by this author. Although most learners at the A, B and C did not produce distinct tongue contours for /ʂ ɕ/, two advanced C-level learners did show evidence of distinct tongue contours for /ʂ ɕ/.

In assessing the changes to the version of the paper, I think that the author did a sufficient job

revising the paper such that I would revise my recommendation to “Accept” with minimal changes necessary.

I appreciated that the author added more discussion about the fact we are dealing with articulatory data showing no contrast between /ʂ/ and /ɕ/ for the majority of L2 learners. However, it is correct to point out, as Nance & Nagamine do in their review, that this does not necessarily mean that they do not have two categories at the acoustic or perceptual level. I feel, however, that this issue is best addressed with future research projects. It sounds like the author does have audio and video data from the same research project that could be used to further evaluate Sorbian learners’ ability to acquire /ʂ/ and /ɕ/.

I also feel that a future study could investigate these same individual’s productions in their German. Are they producing German /ʃ/ in the same way as they produce this merged /ʂ ~ ɕ/ category?

I also very much appreciated the addition of a German language summary at the top. I’m wondering whether a Sorbian-language could be added and might be even more important to the community.

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Phonetics, Phonology, German Dialectology, Historical Linguistics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

**Version 1**

Reviewer Report 31 August 2023

https://doi.org/10.21956/openreseurope.16097.r34213

**© 2023 Bolter D.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

**David Bolter**
[1] Denison University, Granville, Ohio, USA
[2] Denison University, Granville, Ohio, USA

This article presents an interesting analysis of the degree to which German learners of Lower Sorbian can acquire the contrast between /s ʂ ɕ/ in that language, coming from a system that only contrasts /s ʃ/. The study uses ultra-sound technology and finds that A-Level, B-Level and C-Level learners do not show a contrast in their articulation of (target) /ɕ/ and /ʂ/. Only highly advanced C-level learners demonstrate a clear contrast between (target) /ɕ/ and /ʂ/. However, these highly

advanced speakers did not necessarily realize that in a fully target-like manner.

This is an interesting article that has interesting implications for the kinds of language contact situations where the smaller language is moribund. In my review, I concentrate my energies on questions regarding the type of German that the participants speak (do they contrast /ʃ/ and /ç/?) and the consequences this may have for their acquisition of Sorbian as well as the larger context of language contact situations where the sibilant inventories differ. On the latter point, I wonder specifically whether the authors might be able to offer any predictions on what might happen in some of the cases that I mention.

My recommendation is that the article be approved pending some revision, with the changes overall being mostly rather minor. I think most of the paper is very clear, but there are some paragraphs that I find confusing specifically in the “Methods” section.

I further clarify that I have not evaluated the statistical methods undertaken in this article and I encourage further reviewers to address this aspect of the paper.

Typographic / Clarity suggestions:

(p. 4-5): The language in the participants section is confusing to me. I’m having a hard time understanding the speakers’ linguistic backgrounds. For example, the authors write “All participants had a first language of German” (p. 5), but then a few sentences later, they discuss the data collected with L1 Sorbian speakers. Perhaps I am not understanding how the authors are L1 speakers, but this seems contradictory to me. Also, what exactly is meant by a “late-acquiring bilingual speaker of Sorbian” (p. 4)?

(p. 6): “The contours for /ʂ, ɕ/ were more rounded, fronted, and raised than for /s/.”
In this sentence, my assumption is that “rounded” is referring to rounding or arching of the tongue body, rather than to lip rounding. Is this correct?

(p. 9) “learners“ is misspelled as “leaners” in the second line under “Discussion”.

General comments:

I’m wondering about the system of fricatives used by the participants (and the surrounding area) when speaking (Standard?) German. For example, many speakers in Saxony (cf. AdA: https://www.atlas-alltagssprache.de/runde-2/f25c/) do not contrast between /ʃ/ and /ç/. I suppose Cottbus is a bit outside of this area, but nonetheless establishing their ‘starting point’ in German could be useful in assessing their acquisition. If any speakers are not distinguishing /ʃ/ and /ç/ when speaking Standard German, then I imagine this might affect their ability to learn a contrast between /ʂ/ and /ɕ/.

Either way, it is sometimes said that Standard German /ʃ/ is labialized [ʃʷ] (Krech et al. 2009: 81-83), perhaps to amplify the contrast with /ç/ (?). If this is so, then a labialized [ʃʷ] might not be all that different from a Slavic [ʂ]. Especially, since some researchers dispute the label of “retroflex” being applied to a language like Polish on the grounds that Polish /ʂ/ has a relatively flat tongue profile (see the debate between Hamann 2004, Żygis et al. 2012 and Ćavar & Lulich 2020).

I believe that the article could benefit from some mention of this problem, specifically with regards to the hypothesis section. As an ancillary question, I'm wondering whether the data in question could be used to test the goodness of fit of L2 /ɕ/ and /ʂ/ to other L1 sounds (mostly L2 /ɕ/ to L1 /ç/, but a similar problem arises with /ʂ/ and other adjacent sounds in the German inventory).

On a very different note, I'm wondering if the authors have any thoughts on other types of language contact situations where the fricative inventories differ. An interesting comparandum (in more ways than one) that the author may wish to consider now or for future work, would be the sibilant systems of Basque in contact with (Peninsular) Spanish (cf. Beristain 2022). Here, the fricative systems in contact are different (Basque /s̻ s̻̪ ʃ/ vs. Peninsular Spanish /θ s̪/), but the /s̪/ may merge in either direction in different Basque varieties and this may have implications on how Spanish is spoken by those individuals.

Another language contact situation that could provide an interesting comparison would be the varieties of Sinitic Languages in China. My understanding is that Sinitic languages of the north like Mandarin general contrast /s ʂ ɕ/ , whereas Sinitic languages of the Central and Southern have fewer sibilants. For example, Chen & Gussenhoven (2015) describe Shanghai Chinese as contrasting /s/-series and /ɕ/-series and Hong Kong Cantonese (Zee 1991) has only an /s/-series. My understanding is that L1 speakers of such varieties merge the /s/ and /ʂ/-series when speaking Mandarin.

All that said, I'm wondering if the author has any predictions regarding how language contact situations with differing sibilant systems such as those referenced in the preceding paragraphs might behave.

**References**

1. Beristain A: Spectral properties of anterior sibilant fricatives in Northern Peninsular Spanish and sibilant-merging and non-merging varieties of Basque. *Journal of the International Phonetic Association*. 2022; **52** (3): 421-452 Publisher Full Text
2. Ćavar ME, Rudman EM, Lulich SM: Palatalization in coronal consonants of Polish: A three-/four-dimensional ultrasound study.*J Acoust Soc Am*. 2020; **148** (6): EL447 PubMed Abstract | Publisher Full Text
3. Elspaß S, Möller R: Atlas zur deutschen Alltagssprache (AdA). 2003.
4. Hamann S: Retroflex fricatives in Slavic languages. *Journal of the International Phonetic Association*. 2004; **34** (1): 53-67 Publisher Full Text
5. Krech EM, Stock E, Hirschfeld U, Anders LC: Deutsches Aussprachewörterbuch. *Walter de Gruyter.* 2009.
6. Chen Y, Gussenhoven C: Shanghai Chinese. *Journal of the International Phonetic Association*. 2015; **45** (3): 321-337 Publisher Full Text
7. Zee E: Chinese (Hong Kong Cantonese). *Journal of the International Phonetic Association*. 1991; **21** (1): 46-48 Publisher Full Text
8. Żygis M, Pape D, Jesus L: (Non-)retroflex Slavic affricates and their motivation: Evidence from Czech and Polish. *Journal of the International Phonetic Association*. 2012; **42** (3): 281-329 Publisher Full Text

**Is the work clearly and accurately presented and does it engage with the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**Are all the source data and materials underlying the results available?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
I cannot comment. A qualified statistician is required.

**Are the conclusions drawn adequately supported by the results?**
Yes

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Phonetics, Phonology, German Dialectology, Historical Linguistics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 02 Feb 2024

**Phil Howson**

This article presents an interesting analysis of the degree to which German learners of Lower Sorbian can acquire the contrast between /s ʂ ɕ/ in that language, coming from a system that only contrasts /s ʃ/. The study uses ultra-sound technology and finds that A-Level, B-Level and C-Level learners do not show a contrast in their articulation of (target) /ɕ/ and /ʂ/. Only highly advanced C-level learners demonstrate a clear contrast between (target) /ɕ/ and /ʂ/. However, these highly advanced speakers did not necessarily realize that in a fully target-like manner.

This is an interesting article that has interesting implications for the kinds of language contact situations where the smaller language is moribund. In my review, I concentrate my energies on questions regarding the type of German that the participants speak (do they contrast /ʃ/ and /ç/?) and the consequences this may have for their acquisition of Sorbian as well as the larger context of language contact situations where the sibilant inventories differ. On the latter point, I wonder specifically whether the authors might be able to offer any predictions on what might happen in some of the cases that I mention.

My recommendation is that the article be approved pending some revision, with the changes overall being mostly rather minor. I think most of the paper is very clear, but there

are some paragraphs that I find confusing specifically in the “Methods” section.

I further clarify that I have not evaluated the statistical methods undertaken in this article and I encourage further reviewers to address this aspect of the paper. Typographic / Clarity suggestions: (p. 4-5): The language in the participants section is confusing to me. I’m having a hard time understanding the speakers’ linguistic backgrounds. For example, the authors write “All participants had a first language of German” (p. 5), but then a few sentences later, they discuss the data collected with L1 Sorbian speakers. Perhaps I am not understanding how the authors are L1 speakers, but this seems contradictory to me. Also, what exactly is meant by a “late-acquiring bilingual speaker of Sorbian” (p. 4)?

**Response:** I changed L1 to “baseline speakers” to be consistent throughout. The two baseline speakers are also described in more detail: “These participants were chosen for this study because they both had significant input stimuli during the learning process from L1 speakers. Both speakers had input from L1 speaking relatives and additionally the older speaker attended the Sorbian school at a time when L1 Sorbian speaking teachers were active. Additionally, at the time of recording this data, few L1 speakers remain, and the advanced age of potential participants (above 80 years of age) makes ultrasound data especially difficult to record and interpret.”

(p. 6): “The contours for /ʂ, ɕ/ were more rounded, fronted, and raised than for /s/.” In this sentence, my assumption is that “rounded” is referring to rounding or arching of the tongue body, rather than to lip rounding. Is this correct?

**Response:** Now reads: “The contours for /ʂ, ɕ/ had more rounded tongue shape, with more fronting, and more posterior tongue body raising than for /s/.”

(p. 9) “learners“ is misspelled as “leaners” in the second line under “Discussion”.

**Response:** Corrected.

General comments: I’m wondering about the system of fricatives used by the participants (and the surrounding area) when speaking (Standard?) German. For example, many speakers in Saxony (cf. AdA: https://www.atlas-alltagssprache.de/runde-2/f25c/) do not contrast between /ʃ/ and /ç/. I suppose Cottbus is a bit outside of this area, but nonetheless establishing their ‘starting point’ in German could be useful in assessing their acquisition. If any speakers are not distinguishing /ʃ/ and /ç/ when speaking Standard German, then I imagine this might affect their ability to learn a contrast between /ʂ/ and /ɕ/. Either way, it is sometimes said that Standard German /ʃ/ is labialized [ʃʷ] (Krech et al. 2009: 81-83), perhaps to amplify the contrast with /ç/ (?). If this is so, then a labialized [ʃʷ] might not be all that different from a Slavic [ʂ]. Especially, since some researchers dispute the label of “retroflex” being applied to a language like Polish on the grounds that Polish /ʂ/ has a relatively flat tongue profile (see the debate between Hamann 2004, Żygis et al. 2012 and Ćavar & Lulich 2020).

**Response:** This is difficult to say with the current data set. What can be ascertained from the data provided here is that there is tremendous variation in the way individual speakers

produce /ʂ, ɕ/ and that overwhelmingly the tongue contours overlap in a way as to suggest no significant differences in the way they are articulated.

I believe that the article could benefit from some mention of this problem, specifically with regards to the hypothesis section. As an ancillary question, I'm wondering whether the data in question could be used to test the goodness of fit of L2 /ɕ/ and /ʂ/ to other L1 sounds (mostly L2 /ɕ/ to L1 /ç/, but a similar problem arises with /ʂ/ and other adjacent sounds in the German inventory).

**Response:** With the current dataset, it's not really possible to do this type of comparison. It is an interesting suggestion and examinations of German L1 learners of a language with the three-way sibilant contrast would benefit strongly from incorporating these data.

On a very different note, I'm wondering if the authors have any thoughts on other types of language contact situations where the fricative inventories differ. An interesting comparandum (in more ways than one) that the author may wish to consider now or for future work, would be the sibilant systems of Basque in contact with (Peninsular) Spanish (cf. Beristain 2022). Here, the fricative systems in contact are different (Basque /s̻ ʂ̻ ʃ/ vs. Peninsular Spanish /θ ʂ̺/), but the /ʂ̺/ may merge in either direction in different Basque varieties and this may have implications on how Spanish is spoken by those individuals. Another language contact situation that could provide an interesting comparison would be the varieties of Sinitic Languages in China. My understanding is that Sinitic languages of the north like Mandarin general contrast /s ʂ ɕ/ , whereas Sinitic languages of the Central and Southern have fewer sibilants. For example, Chen & Gussenhoven (2015) describe Shanghai Chinese as contrasting /s/-series and /ɕ/-series and Hong Kong Cantonese (Zee 1991) has only an /s/-series. My understanding is that L1 speakers of such varieties merge the /s/ and /ʂ/-series when speaking Mandarin. All that said, I'm wondering if the author has any predictions regarding how language contact situations with differing sibilant systems such as those referenced in the preceding paragraphs might behave.

**Response:** I think language contact situations can differ significantly based on the communities. Certainly, there are theoretical considerations along with those, such that we may assume or observe tendencies towards certain patterns of acquisition and assimilation. In keeping with the theoretical claims in this paper, the most likely assimilatory patterns are likely those that adhere to the assimilation of segments with the greatest acoustic-perceptual overlap.

***Competing Interests:*** No competing interests were disclosed.

Reviewer Report 25 August 2023

https://doi.org/10.21956/openreseurope.16097.r33691

**© 2023 Nance C et al.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Claire Nance**
[1] Linguistics and English Language, Lancaster University, Lancaster, England, UK
[2] Linguistics and English Language, Lancaster University, Lancaster, England, UK
**Takayuki Nagamine**
[1] Linguistics and English Language, Lancaster University, Lancaster, England, UK
[2] Linguistics and English Language, Lancaster University, Lancaster, England, UK

**Summary**

This is an articulatory study which considers L1 German speakers' production of fricatives in L2 Lower Sorbian speech. The study analyses production of Lower Sorbian /s ʂ ɕ/, hypothesising that L1 German - L2 Lower Sorbian speakers would produce Lower Sorbian /ʂ ɕ/ similarly to German /ʃ/ due to their acoustic-perceptual similarity. Effect of L2 proficiency is also considered.

The author analyses ultrasound tongue imaging data obtained from Lower Sorbian speakers at different proficiency levels. Tongue splines from Sorbian fricative midpoints were extracted at the acoustically defined fricative midpoint and analysed with polar GAMMs.

L1 speakers showed three-way contrasts in Lower Sorbian /s ʂ ɕ/ whereas the majority of L2 speakers exhibited two-way contrasts between /s/ and /ʂ ɕ/, arguably due to influence of L1 German. Only advanced L2 users of Lower Sorbian produced the three-way fricative contrast.

**Strengths of this study**

It is wonderful to see more articulatory work on Sorbian! It's so important to have this kind of language documentation, as well as a detailed investigation of speech production in a language revitalisation setting.

We haven't previously seen analysis of language revitalisation users at different proficiency levels, so this is an important contribution of the work.

There is very little articulatory work carried out with bilinguals/L2 users, so this work is a significant addition to the field.

We suggest some revisions to the final version of the paper. This mainly concerns some rephrasing of the framing, more details on aspects of the methods and analysis, and greater acknowledgement of what can and can't be concluded from this dataset.

**Major points**

Sociolinguistically, this is a context of extreme language endangerment and some revitalisation. It is to be expected that speakers using the language as an L2 won't sound exactly the same as older people who acquired it in a completely different social context. Suggest some rephrasing and acknowledgement of this throughout. For example, removing references to speakers not producing sounds 'properly' etc. Replace with them sounding 'different' to Sorbian acquired two generations ago in an L1 setting.

We appreciate that sociolinguistics is not the primary framing of the work here, but if you wanted to consider this aspect in more detail, you could investigate work conducted in the new speaker framework. It seems that 'new speakers' would be a very appropriate way to describe these Sorbian L2 users. Some papers we found helpful:
O'Rourke & Ramallo. (2011). https://doi.org/10.1075/lplp.35.2.03oro
Jaffe. (2015). https://doi.org/10.1515/ijsl-2014-0030

**Hypothesis:**
This section gives predictions for the study based on the PAM-L2 and SLM(-r). However, the current study is all about articulation with no examination of the acoustic data. Some thinking needs to go into this about the targets that speakers are aiming for. Are they aiming for an acoustic target or an articulatory target in acquiring Sorbian?

Our understanding is that while the PAM-L2 has a prediction about articulation (that speakers directly perceive gestures and acquire these), the SLM-r is less specific about articulation. Clarify this distinction somewhere in the lit review as it is important for the hypothesis here.

In the hypothesis paragraph, you mention quite a lot about formant transitions and spectral characteristics of the fricatives. This can't be tested in the current design, which is purely articulatory. We suggest rethinking about how this material fits into the articulatory message of this paper, and the links between acoustics and articulation.

This is important for the rest of the study. For example, on p9 you refer to speakers not acquiring the contrast. The data here refer to midsagittal tongue splines at fricative midpoint only. It is possible that speakers are making an acoustic contrast in some other way. Unlikely perhaps, but some thought needs to go into what we can and can't ascertain from these data.

**Discussion:**
There is an argument here and in the Results that Sorbian speakers are substituting German /ʃ/ for two of the Sorbian fricatives. This seems impossible to know from the current data, since German /ʃ/ isn't analysed. It is possible that Sorbian learners are doing something that is different from L1 Sorbian and also different from German. See Moore et al. (2018) for a similar example where Japanese L1 English L2 speakers produce a sound which is not like L1-English /l/ and /ɹ/, but is also different from their Japanese [ɾ].
Moore et al. (2018). https://doi.org/10.1250/ast.39.75

**Open Science:**
We are grateful to the author for providing the datasheets used. In order for future researchers to fully replicate this study and understand all the statistics carried out, it would be very helpful to have the code used as well.

**Minor points**

**Abstract:**
This reads more like the first page of the Introduction. It would be helpful to make this more of an 'advert' for the study. Include more of the findings here. A major contribution of the work, and a major feature of the study design, is comparing the different proficiency levels for speakers. This should be mentioned in the Abstract.

**Plain language summary:**
This is still quite technical for non-linguists to read, for example references to IPA terminology. Suggest some rephrasing such that this summary would be useful to Sorbian community members (e.g. give example words for IPA). Would it be possible to translate into German and/or Sorbian as well for maximum readability by Sorbian users?

**Introduction:**
2nd paragraph: Is it possible to quantify how uncommon a three-way sibilant contrast is?
For example, look through Maddieson (1984) *Patterns of sounds*, or the P-database:
https://pbase.phon.chass.ncsu.edu

A lot in this analysis hangs on /ʂ/ and /ɕ/ being more acoustically/perceptually similar to /ʃ/ than /s/. Is there any way you can demonstrate this?

**Second language acquisition:**
The material in this section comes a bit abruptly after the Introduction. Could it be linked in a little more?

**PAM-L2:**
P3, final paragraph, column 2: Suggest rephrasing 'poverty of stimulus'. We're talking about a highly endangered language where people are doing their very best to revitalise and transmit the language however they can. A classroom will be different to a home setting, but 'poverty of stimulus' sounds very negative, perhaps unnecessarily so.

Similarly, the reference to teachers 'not properly producing' target segments. Maybe rephrase to acknowledge that L2 users will speak differently to L1 users, but this is what we would expect anyway.

Could the discussion about lexical frequency and word learning vs. category learning (Tyler 2019) be linked into the current study? This material seems perhaps less relevant for a word list task such as the current work. See also comments further down about the stimuli.

**Speech Learning Model:**
Second paragraph tells us that the target population are late-acquiring bilinguals. Mention this earlier.
It would be good to define what is meant by 'late-acquiring bilingual'.

**Methods:**
**Study design:**
No need to mention the quiet room as this is repeated later on.

**Participants:**
We were a bit confused about the baseline participants. Are they the L1 speakers? You mentioned earlier that most L1 speakers are elderly.

As one of them acquired Sorbian from age 5 (if we understand correctly), could you give a little more information about their language acquisition trajectory? How were they selected as

representative of L1 Sorbian?

**Figure 1:**
Could you give a bit more information about what is shown in this figure? Are the arrows relevant to a particular bit of the waveform? What do the spectrograms show here which informs the reader beyond IPA symbols and arrows? Suggest changing to a schematised IPA diagram instead unless the spectrograms illustrate something specific.

**Participants p5:**
Could you specify the age and gender of participants in the main text instead of the supplementary materials, as this seems like key information. Explain a little bit more about language learning trajectories and how some participants have spent a long time at A-level compared to others who have reached B-level comparatively faster. In a context of language revitalisation learning, this isn't entirely surprising, but might be less evident to readers coming from a majority-language teaching background.

How was the level of the learners assessed? Is this through a standardised test or teacher perceptions?

Make it clear that these scores refer to CEFR scales (we assume), otherwise this can be a bit confusing. For example, as a British reader, it is hard not to read 'A-Level' as referring to the standard tests taken by 18-year-old school leavers in most of the UK. These are referred to as 'A-Levels'.

How was the level of the advanced C2 speakers ascertained?
Is it still appropriate to refer to them as 'learners' if they are so advanced? Suggest replacing with 'L2 users' or 'L2 speakers'.

'one of which has achieved a near-native level of fluency' – how was this ascertained?
Suggest replacing with 'L1-like level of fluency'.

**Procedure:**
Second paragraph at the bottom of p5: 'Data for the bilingual speakers' – aren't all of them bilingual?

**Stimuli:**
This is quite a small list for a study to make claims about category learning. As there is only one word per context, it is not possible to disentangle category learning, and learning of individual words. Some of the words are presumably more frequent than others, which will likely affect acquisition for the less advanced speakers. This needs to be acknowledged in the study text.

**Analysis:**
It would be helpful to have a little more information about the ultrasound data here:
Was any manual correction carried out for the tongue splines obtained from AAA?
Were the data rotated to the occlusal plane?
Which version of AAA was used for the analysis?

Could you give a bit more information about how significance testing was carried out? Via model

comparison?

**Results:**
**A-Level learners:**
'A-Level learners share one phoneme' – rephrase. They share one tongue shape. We don't know about the perceptual categories needed for phoneme-hood.
Final paragraph of this section: is the 'significant variation in their articulation of /ʃ/' a typo?

**B-Level learners:**
'have not acquired the three-way contrast' – make it clear that this refers to articulation, as we don't know about their acoustics/perception from this study.

First paragraph on p9: 'B-level learners continue to use German /ʃ/'. It's not possible to draw this conclusion from these data as they don't include analysis of German /ʃ/.

/ʃ/ often involves significant lip rounding. This study doesn't consider lip rounding analysis. This should be taken into account when drawing conclusions about likely German influence.

**Highly advanced C-Level learners:**
It seems there is some inconsistency in labelling these participants between the main text, and the supplementary materials. Consider using 'C04' and 'C05' throughout instead of 'C201' and 'C202'.

**Tongue spline figures:**
Consider changing the red/orange colour scheme to maximise contrast for readers.

**Discussion:**
It is suggested here that L2 Sorbian learners experience 'L1 interference in the acoustic-perceptual space'. But at the same time, the text says that Sorbian teachers might not produce this contrast themselves. So, it is possible that learners are doing a really amazing job of repeating exactly what their teachers produce and aren't experiencing L1 influence directly themselves at all. Could you explain how these things can be disentangled, or what is the most likely interpretation of the data?

'contrasts with difficult to perceive differences require specific training to acquire'. Could you give some evidence from the L2 pronunciation literature to back this up? This isn't my exact area of specialism, but my impression is that L2 pronunciation training is extremely difficult and might not be 'successful' in creating L1-like pronunciations.

At the same time, thought needs to be given to what is a realistic target for L2 users and language revitalisation speakers. In applied circles, it is now usually considered more appropriate to aim for intelligible and comprehensible L2 speech, rather than an L1 target for pronunciation.

You conclude that intervention and training is needed here and in the plain language summary. But the advanced C-Level speakers did produce this contrast without having had special training. It could therefore be argued that more language exposure and use is required instead.

**References**
1. O'Rourke B, Ramallo F: The native-non-native dichotomy in minority language contexts.

*Language Problems and Language Planning*. 2011; **35** (2): 139-159 Publisher Full Text
2. Jaffe A: Defining the new speaker: theoretical perspectives and learner trajectories. *International Journal of the Sociology of Language*. 2015; **2015** (231): 21-44 Publisher Full Text
3. Moore J, Shaw J, Kawahara S, Arai T: Articulation strategies for English liquids used by Japanese speakers. *Acoustical Science and Technology*. 2018; **39** (2): 75-83 Publisher Full Text
4. Maddieson I: Patterns of sounds. *Cambridge University Press*. 1984.

**Is the work clearly and accurately presented and does it engage with the current literature?**
Yes

**Is the study design appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**
Partly

**Are all the source data and materials underlying the results available?**
Partly

**If applicable, is the statistical analysis and its interpretation appropriate?**
Partly

**Are the conclusions drawn adequately supported by the results?**
Yes

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Claire Nance and Takayuki Nagamine reviewed this article. We work in bilingualism and articulatory phonetics. We use ultrasound and acoustic analysis as research methods. Claire has expertise in sociolinguistics of language revitalisation (Scottish Gaelic in particular). Takayuki has expertise in language teaching and L2 pronunciation (Japanese-English bilinguals and English teaching).

**We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.**

Author Response 02 Feb 2024

**Phil Howson**

This is an articulatory study which considers L1 German speakers' production of fricatives in L2 Lower Sorbian speech. The study analyses production of Lower Sorbian /s ʂ ɕ/, hypothesising that L1 German - L2 Lower Sorbian speakers would produce Lower Sorbian /ʂ ɕ/ similarly to German /ʃ/ due to their acoustic-perceptual similarity. Effect of L2 proficiency is also considered. The author analyses ultrasound tongue imaging data obtained from Lower Sorbian speakers at different proficiency levels. Tongue splines from Sorbian fricative

midpoints were extracted at the acoustically defined fricative midpoint and analysed with polar GAMMs. L1 speakers showed three-way contrasts in Lower Sorbian /s ʂ ɕ/ whereas the majority of L2 speakers exhibited two-way contrasts between /s/ and /ʂ ɕ/, arguably due to influence of L1 German. Only advanced L2 users of Lower Sorbian produced the three-way fricative contrast. Strengths of this study It is wonderful to see more articulatory work on Sorbian! It's so important to have this kind of language documentation, as well as a detailed investigation of speech production in a language revitalisation setting. We haven't previously seen analysis of language revitalisation users at different proficiency levels, so this is an important contribution of the work. There is very little articulatory work carried out with bilinguals/L2 users, so this work is a significant addition to the field. We suggest some revisions to the final version of the paper. This mainly concerns some rephrasing of the framing, more details on aspects of the methods and analysis, and greater acknowledgement of what can and can't be concluded from this dataset. *Major points* Sociolinguistically, this is a context of extreme language endangerment and some revitalisation. It is to be expected that speakers using the language as an L2 won't sound exactly the same as older people who acquired it in a completely different social context. Suggest some rephrasing and acknowledgement of this throughout. For example, removing references to speakers not producing sounds 'properly' etc. Replace with them sounding 'different' to Sorbian acquired two generations ago in an L1 setting.

**Response:** Thank you for this, I think it is an important change to make. I have changed the language throughout.

We appreciate that sociolinguistics is not the primary framing of the work here, but if you wanted to consider this aspect in more detail, you could investigate work conducted in the new speaker framework. It seems that 'new speakers' would be a very appropriate way to describe these Sorbian L2 users. Some papers we found helpful: O'Rourke & Ramallo. (2011). https://doi.org/10.1075/lplp.35.2.03oro Jaffe. (2015). https://doi.org/10.1515/ijsl-2014-0030 Hypothesis: This section gives predictions for the study based on the PAM-L2 and SLM(-r). However, the current study is all about articulation with no examination of the acoustic data. Some thinking needs to go into this about the targets that speakers are aiming for. Are they aiming for an acoustic target or an articulatory target in acquiring Sorbian?

**Response:** I have added the following text: "In terms of the PAM-L2 (Best & Tyler, 2007), the assumption is that learners are perceiving articulatory gestures and vocal tract changes, not more abstract acoustic characteristics. The implication of this is that as learners become more advanced, they become better at retrieving the articulatory movements necessary to produce a contrast. The expectation is that gradual improvement in the articulation of L2 segments should occur."

Our understanding is that while the PAM-L2 has a prediction about articulation (that speakers directly perceive gestures and acquire these), the SLM-r is less specific about articulation. Clarify this distinction somewhere in the lit review as it is important for the hypothesis here.

**Response:** I have added the following text: "The PAM-L2 is a direct realist model, which

assumes that perception is related to the perception of distal articulatory events (i.e., changes in vocal tract configurations), not specific acoustic patterns." and: "Additionally, the PAM-L2 posits that learners attenuating to gestural movements in the vocal tract, while the SLM-r suggests that learners pay attention to acoustic differences in the input signal directly. Thus, under the view of the SLM-r, articulation is a matter of better navigation of what vocal tract shapes produce the target acoustic outputs."

In the hypothesis paragraph, you mention quite a lot about formant transitions and spectral characteristics of the fricatives. This can't be tested in the current design, which is purely articulatory. We suggest rethinking about how this material fits into the articulatory message of this paper, and the links between acoustics and articulation.

**Response:** I have added the following text: "In terms of the PAM-L2 (Best & Tyler, 2007), the assumption is that learners are perceiving articulatory gestures and vocal tract changes, not more abstract acoustic characteristics. The implication of this is that as learners become more advanced, they become better at retrieving the articulatory movements necessary to produce a contrast. The expectation is that gradual improvement in the articulation of L2 segments should occur. In terms of the SLM-r (Flege & Bohn, 2021), there is a similar expectation. As learners' acoustic-perceptual representation improves, so too should articulation."

This is important for the rest of the study. For example, on p9 you refer to speakers not acquiring the contrast. The data here refer to midsagittal tongue splines at fricative midpoint only. It is possible that speakers are making an acoustic contrast in some other way. Unlikely perhaps, but some thought needs to go into what we can and can't ascertain from these data. Discussion: There is an argument here and in the Results that Sorbian speakers are substituting German /ʃ/ for two of the Sorbian fricatives. This seems impossible to know from the current data, since German /ʃ/ isn't analysed. It is possible that Sorbian learners are doing something that is different from L1 Sorbian and also different from German. See Moore et al. (2018) for a similar example where Japanese L1 English L2 speakers produce a sound which is not like L1-English /l/ and /ɹ/, but is also different from their Japanese [ɾ]. Moore et al. (2018). https://doi.org/10.1250/ast.39.75 Open Science: We are grateful to the author for providing the datasheets used. In order for future researchers to fully replicate this study and understand all the statistics carried out, it would be very helpful to have the code used as well.

**Response:** The code is included: " (1) r ~ Segment * Environment + s(Theta, bs = "cr", k = 25) + s(Theta, by = Segment, bs = "cr", k = 25) + s(Theta, by = Segment : Environment, bs = "cr", k = 25) + s(Theta, by = Segment : Speaker, bs = "fs", k = 25, m = 1)" "(2) r ~ Segment * Environment + s(Theta, bs = "cr", k = 20) + s(Theta, by = Segment, bs = "cr", k = 20) + s(Theta, by = Segment : Environment, bs = "cr", k = 20) + s(Theta, by = Rep, bs = "fs", k = 20, m = 1)" These models are clearly written such that they are easily producible with datasheet given.

Minor points Abstract: This reads more like the first page of the Introduction. It would be helpful to make this more of an 'advert' for the study. Include more of the findings here. A major contribution of the work, and a major feature of the study design, is comparing the different proficiency levels for speakers. This should be mentioned in the Abstract.

**Response:** I have added the following: “The ultrasound data revealed that learners in the contemporary context do not produce a distinction between /ʂ, ɕ/ and only learners at an advanced level who had significant exposure to L1 speakers have acquired a three-way sibilant distinction.”

Plain language summary: This is still quite technical for non-linguists to read, for example references to IPA terminology. Suggest some rephrasing such that this summary would be useful to Sorbian community members (e.g. give example words for IPA). Would it be possible to translate into German and/or Sorbian as well for maximum readability by Sorbian users?

**Response:** I have added the following example: “Both of these segments are perceptually similar to the German sibilant fricative common represented with sch (e.g., Schlange snake).” I have also changed the following text to read: “The results implicate that in a language revitalization context where few L1 speakers are available, the input that learners receive should be augmented with pronunciation and perceptual resources to assist in acquisition. Specific recommendations are provided.” A German translation is also provided: “Auf Deutsch Der Zweitspracherwerb setzt voraus, dass die Sprachlernenden eine Reihe neuer Sprachsegmente erwerben. Junge Sorben, die Niedersorbisch als Zweitsprache lernen, müssen zwei neue Zischlaute (hochfrequente, laute Segmente wie /s/) erwerben. Diese beiden Segmente ähneln in der Wahrnehmung dem deutschen Zischlaut, der häufig mit sch dargestellt wird (z.B. Schlange). In dieser Studie wird der Erwerb der niedersorbischen Zischlautkontraste mittels Ultraschalltechnologie untersucht. Ultraschall zeichnet Videos von Zungenkonturen mit einer hohen Bildrate auf, so dass eine statistische Analyse der Zungenformen durchgeführt werden kann. In diesem Projekt untersuche ich die Zungenkonturen niedersorbischer Lerner auf der Anfänger-, Mittelstufen- und Fortgeschrittenenstufe, um zu beobachten, wie Zungenformen für Zischlaute erworben werden. Die Ergebnisse deuten darauf hin, dass in einem Sprachwiederbelebungskontext, in dem nur wenige L1-Sprecher zur Verfügung stehen, der Input, den die Lernenden erhalten, mit Aussprache- und Wahrnehmungsressourcen ergänzt werden sollte, um den Erwerb zu unterstützen. Es werden spezifische Empfehlungen gegeben.”

Introduction: 2nd paragraph: Is it possible to quantify how uncommon a three-way sibilant contrast is? For example, look through Maddieson (1984) Patterns of sounds, or the P-database: https://pbase.phon.chass.ncsu.edu

**Response:** From what I can derive from Maddieson (1984), specific numbers on the three-way sibilant contrast are not provided. Numbers for languages with one or more fricative are provided, but reference to which segment that is are not. This is more complex when voiced and voiceless pairs are considered 2 segments and non-sibilant segments are included. Thus, minimally, we would expect that any of the numbers for 6+ fricatives could apply, but the numbers then vary. I could estimate, that in all likelihood, of the languages examined by Maddieson (1984), less than 6% have a three-way sibilant contrast. That being said, this would be an estimation based on the available data.

A lot in this analysis hangs on /ʂ/ and /ɕ/ being more acoustically/perceptually similar to /ʃ/

than /s/. Is there any way you can demonstrate this?

**Response:** I have added the following to page ?: "The acoustics between the two segments /ʂ, ɕ/ resemble each other across in COG and skewness, having both a lower COG and higher skewness than /s/. Both values also significantly overlapped with each other for /ʂ, ɕ/. The feature in Lower Sorbian that was found to most strongly distinguish /ʂ, ɕ/ from each other was a much higher transitional F2 into the following vowel for /ɕ/ compared to /ʂ/ (Howson, 2015). The lower COG values observed in Lower Sorbian, tend to match cross-linguistic COG associated with /ʃ/ (Żygis, 2010) and COG and skewness measures associated with German /ʃ/ (Weirich & Simpson, 2015)."

Second language acquisition: The material in this section comes a bit abruptly after the Introduction. Could it be linked in a little more?

**Response:** I have added the following: "Many theories of language acquisition, such as the PAM-L2 (Best & Tyler, 2007) and the SLM-r (Flege & Bohn, 2021), have postulated how different aspects of acoustic-perceptual similarities with L1 segments impacts L2 acquisition. Contrasts such as the three-way contrast, under these theories are the most difficult to acquire due to the acoustic similarities between the segments. This makes Lower Sorbian an excellent language to examine foreign language acquisition of sibilant fricatives."

PAM-L2: P3, final paragraph, column 2: Suggest rephrasing 'poverty of stimulus'. We're talking about a highly endangered language where people are doing their very best to revitalise and transmit the language however they can. A classroom will be different to a home setting, but 'poverty of stimulus' sounds very negative, perhaps unnecessarily so.

**Response:** I have changed this to: "The reason for this is because of a reduced access to consistent stimuli and the phonetic contrasts that distinguish them."

Similarly, the reference to teachers 'not properly producing' target segments. Maybe rephrase to acknowledge that L2 users will speak differently to L1 users, but this is what we would expect anyway.

**Response:** I have changed this to read the following: "Many second language classrooms are also taught by second language speakers, who may or may not consistently produce the language relevant contrasts, and likely produce contrasts differently than the older generation of L1 speakers."

Could the discussion about lexical frequency and word learning vs. category learning (Tyler 2019) be linked into the current study? This material seems perhaps less relevant for a word list task such as the current work. See also comments further down about the stimuli.
Speech Learning Model: Second paragraph tells us that the target population are late-acquiring bilinguals. Mention this earlier. It would be good to define what is meant by 'late-acquiring bilingual'.

**Response:** I have added the following: "The SLM posits that for late acquiring bilinguals (i.e.,

someone who acquired two languages as a child, but the second language was acquired later than the first)"

Methods: Study design: No need to mention the quiet room as this is repeated later on.

**Response:** Removed.

Participants: We were a bit confused about the baseline participants. Are they the L1 speakers? You mentioned earlier that most L1 speakers are elderly. As one of them acquired Sorbian from age 5 (if we understand correctly), could you give a little more information about their language acquisition trajectory? How were they selected as representative of L1 Sorbian?

**Response:** Both speakers acquired the languages in a bilingual context and as a minority language. That is, German was the primarily language of the household, but they both had access to first language speakers through family and friends and through earlier integration into the Sorbian school system when L1 teachers were still available. There are two reasons L1 speakers were not included. The first, there are very few and they are very old. This means that most L1 speakers have no real interest in participating in any linguistic (or other) studies on Lower Sorbian. The second reason is that for speakers at such an advanced age, it is extremely difficult to obtain a quality recording. This is one of the draw backs of ultrasound in general and is something that has to be considered. I have added the following text: "These participants were chosen for this study because they both had significant input stimuli during the learning process from L1 speakers. Both speakers had input from L1 speaking relatives and additionally the older speaker attended the Sorbian school at a time when L1 Sorbian speaking teachers were active. Additionally, at the time of recording this data, few L1 speakers remain, and the advanced age of potential participants (above 80 years of age) makes ultrasound data especially difficult to record and interpret."

Figure 1: Could you give a bit more information about what is shown in this figure? Are the arrows relevant to a particular bit of the waveform? What do the spectrograms show here which informs the reader beyond IPA symbols and arrows? Suggest changing to a schematised IPA diagram instead unless the spectrograms illustrate something specific.

**Response:** The figure simply illustrated some observable similarities in center of gravity, however, above I now provide citations and values that compare these segments (see above for specific text added) and it is perhaps no longer necessary. I have removed this figure and renumbered the other figures accordingly.

Participants p5: Could you specify the age and gender of participants in the main text instead of the supplementary materials, as this seems like key information. Explain a little bit more about language learning trajectories and how some participants have spent a long time at A-level compared to others who have reached B-level comparatively faster. In a context of language revitalisation learning, this isn't entirely surprising, but might be less evident to readers coming from a majority-language teaching background.

How was the level of the learners assessed? Is this through a standardised test or teacher

perceptions?

**Response:** There are no real standardized tests for assessing Sorbian abilities. It had only to do with their level within the Sorbian school system. I have added the following text: "Participants were recruited for all three skill levels, A-, B-, and C-level learners based on a scaling system like the CEFR. Their skill level at the time was based on class they attended for Sorbian language at gymnazium Chóśebuz at the time of recording."

Make it clear that these scores refer to CEFR scales (we assume), otherwise this can be a bit confusing. For example, as a British reader, it is hard not to read 'A-Level' as referring to the standard tests taken by 18-year-old school leavers in most of the UK. These are referred to as 'A-Levels'.

**Response:** See above for additional text.

How was the level of the advanced C2 speakers ascertained?

**Response:** Both speakers are well known in the Sorbian community (I prefer not to provide too many identifying details to maintain anonymity). But I can say that both are well known speakers.

Is it still appropriate to refer to them as 'learners' if they are so advanced? Suggest replacing with 'L2 users' or 'L2 speakers'.

**Response:** Changed.

'one of which has achieved a near-native level of fluency' – how was this ascertained?

**Response:** Again, this is generally accepted within the community, and they are well known for this. However, perhaps this information is too much as people who know them might easily figure out who they are. I have removed it.

Suggest replacing with 'L1-like level of fluency'.

**Response:** Removed the reference.

Procedure: Second paragraph at the bottom of p5: 'Data for the bilingual speakers' – aren't all of them bilingual?

**Response:** Changed this to "baseline speakers."

Stimuli: This is quite a small list for a study to make claims about category learning. As there is only one word per context, it is not possible to disentangle category learning, and learning of individual words. Some of the words are presumably more frequent than others, which will likely affect acquisition for the less advanced speakers. This needs to be acknowledged in the study text.

**Response:** The following has been added to the conclusions: “However, it should be noted that one limitation of the study is the relatively small wordlist which makes it more difficult to assess category formation.”

Analysis: It would be helpful to have a little more information about the ultrasound data here: Was any manual correction carried out for the tongue splines obtained from AAA?

**Response:** Tongue traces were all manual. I have added that in the text. This was largely done before I had DeepLabCut installed and operational.

Were the data rotated to the occlusal plane?

**Response:** No.

Which version of AAA was used for the analysis?

**Response:** R was used for the analysis and is now included in the citations. Version data for AAA also included in the text.

Could you give a bit more information about how significance testing was carried out? Via model comparison?

**Response:** The following text was included: “GAMMs were performed using the mgcv package (Wood, 2011), which also provides summary statistics.”

Results: A-Level learners: ‘A-Level learners share one phoneme’ – rephrase. They share one tongue shape. We don’t know about the perceptual categories needed for phoneme-hood.

**Response:** Changed.

Final paragraph of this section: is the ‘significant variation in their articulation of /ʃ/’ a typo?

**Response:** Changed to read /ʂ, ɕ/.

B-Level learners: ‘have not acquired the three-way contrast’ – make it clear that this refers to articulation, as we don’t know about their acoustics/perception from this study.

**Response:** Now reads: “This suggests that like the A-level learners, the B-level learners also have not acquired the three-way contrast between /s, ʂ, ɕ/ with respect to their articulation.”

First paragraph on p9: ‘B-level learners continue to use German /ʃ/’. It’s not possible to draw this conclusion from these data as they don’t include analysis of German /ʃ/.

**Response:** Change to read “the same segment.”

/ʃ/ often involves significant lip rounding. This study doesn’t consider lip rounding analysis. This should be taken into account when drawing conclusions about likely German influence.

**Response:** Thank you for this comment.

Highly advanced C-Level learners: It seems there is some inconsistency in labelling these participants between the main text, and the supplementary materials. Consider using 'C04' and 'C05' throughout instead of 'C201' and 'C202'.

**Response:** Changed.

Tongue spline figures: Consider changing the red/orange colour scheme to maximise contrast for readers.

**Response:** Perhaps the only issue with this is that I was required to register the OSF database prior to receiving these comments. As a result, it is no longer possible to change them and there would then be inconsistencies between plot colors in the text and in the supplementary materials.

Discussion: It is suggested here that L2 Sorbian learners experience 'L1 interference in the acoustic-perceptual space'. But at the same time, the text says that Sorbian teachers might not produce this contrast themselves. So, it is possible that learners are doing a really amazing job of repeating exactly what their teachers produce and aren't experiencing L1 influence directly themselves at all. Could you explain how these things can be disentangled, or what is the most likely interpretation of the data?

**Response:** In reality, it is perhaps more likely that the instructors themselves do not reliably produce the three-way sibilant contrasts and this is one of the starkest contrasts between C04/C05 and the rest of the learners. They had explicit instruction from L1 speakers who consistently produced those contrasts, while the contemporary learners likely do not receive that input in any consistent way, if at all. I have added the following text: "One caveat to note is that the current L2 instructors do not have the level of fluency as the L1 instructors that the two advanced speakers (C04 and C05) had access to. As such, it is difficult to interpret how much input for the three-way sibilant contrast (if any) learners received. It is clear from discussions with the learners that pronunciation lessons are not a regular part of the curriculum. It remains very possible the lack of acquisition of the three-way contrast is predominantly due to lack of the three-way contrast in the input for learners."

'contrasts with difficult to perceive differences require specific training to acquire'. Could you give some evidence from the L2 pronunciation literature to back this up? This isn't my exact area of specialism, but my impression is that L2 pronunciation training is extremely difficult and might not be 'successful' in creating L1-like pronunciations.

**Response:** This specific line was removed. However, L1-like pronunciation may be relatively unachievable for a second language learner for a variety of reasons (at least "perfect" L1-pronunciation). But for example, even children who learn multiple languages from birth, do not have the same pronunciation as someone who learned only 1 language as a child. It may perhaps not be the best measure of learning then to say that they do or do not speak

like a monolingual or the same as someone who learned only 1 language as a child. Rather, in some sense it is better to examine to what degree and how contrasts are made and learned by L2 learners.

At the same time, thought needs to be given to what is a realistic target for L2 users and language revitalisation speakers. In applied circles, it is now usually considered more appropriate to aim for intelligible and comprehensible L2 speech, rather than an L1 target for pronunciation.

**Response:** I think this depends on the specific context and goals/aims of the group in question. From personal work, I can say it is very important to many indigenous communities that they maintain pronunciation similar/the same as it use to be spoken. I would not say this is an unachievable goal, but if it is the goal, then certainly specific measures to achieve that goal need to be taken. If it is not the wish of the community to maintain that pronunciation, then it is also not something that should realistically be expected or worked towards. I have added the following text: "If the desire of the community is to maintain specific speech patterns present in older L1 speakers, then from a practical standpoint, it seems that additional resources need to be committed to this achieve this goal."

You conclude that intervention and training is needed here and in the plain language summary. But the advanced C-Level speakers did produce this contrast without having had special training. It could therefore be argued that more language exposure and use is required instead.

**Response:** True, but realistically, that's not possible given the present situation in Lower Sorbian. Few L1 speakers remain and while I agree that significant exposure to L1 speech with such a contrast would/could facilitate learning, that simply isn't the situation in the present context. If the goal of the Sorbian community is to maintain this contrast (and others), then specific resources would need to be mobilized as it is not realistic or even possible that children or adults would receive significant input stimuli from L1 speakers or even L2 speakers who produce the target contrasts reliably.

***Competing Interests:*** No competing interests were disclosed.