Perception of child-produced Polish sibilants: a comparison of native English speakers and Polish Heritage speakers

Perception of child-produced Polish sibilants: a comparison of native English speakers and Polish Heritage speakers

BY DAKYUNG RACHEL LIM, B.A

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the

Requirements for the Degree Master of Science

McMaster University
MASTER OF SCIENCE (2023)
Hamilton, Ontario (Cognitive Science of Language)

TITLE: Perception of child-produced Polish sibilants: a comparison of native English and Polish Heritage speakers.

AUTHOR: Dakyung Rachel Lim, B.A.

SUPERVISOR: Professor M. Stroinska, Professor D. Pape

NUMBER OF PAGES: x, 50

Lay Abstract

Unfamiliar speech sounds are often labelled into sounds of familiar, acoustic categories. The perception of foreign speech sounds is relevant and interesting to study in Canada with its diverse linguistic pool. As fascinating it would be to study many foreign languages, the focus of our study will be on one particular language and, more specifically, one group of speech sounds. This paper explores the findings of a production-perception study on Polish sibilants (“a type of consonant sound where the tip of the tongue is brought near the roof of the mouth for the air to be pushed past to make hiss sounds (ex: s, z, sh, zh in English))”. Encyclopedia Britannica) in typical children and the perception of their productions by others. This study builds on previous work done by Zygis and colleagues (2023) that focused on Polish children's production, self-perception, and native Polish adults' perception of sibilant productions. The child participants were aged 35-106 months and produced words with the sounds /s, ʂ, ɕ/. These symbols are standardized representations of sounds in the International Phonetic Alphabet (IPA) that correspond to speech sounds that may be present in languages. For our study, the audio recordings (from Zygis et al., 2023) were manipulated and played to native English speakers at McMaster University to determine their distinction of the three sibilant contrasts /s, ʂ, ɕ/, which do not exist in English. In English, there is a two-way distinction of the previously mentioned sibilants which is represented by the IPA symbol /s/ (as in “[s]ake”) and /ʃ/ (like in “[sh]ake”). The goal was to observe how the English participants categorized /s, ʂ, ɕ/ into either /s/ and /ʃ/. Another round of the experiment was done with native English participants with varying Polish fluency residing in the Ontario area to capture their perception of the manipulated sibilant stimuli. English participants increasingly categorized all manipulations of /s/ as /s/ and /ɕ/ as the /ʃ/ sibilant, especially with the older children. Their perception for the retroflex /ʂ/ was split, half as /s/ perceptions, across conditions. Phonetic information in the form of vowel information (on top of the isolated sibilant information) did not significantly improve sibilant distinction for the English participants. The Polish

Heritage speakers showed difficulty in correctly identifying /ʂ/ variations, especially in the older children. Phonetic environment and age had varying effects depending on the sibilant. The findings from this study contribute to our understanding of cross-language differences in the acquisition of foreign speech sounds. Note, that acquisition here refers to learning of speech sounds, which does not mean perfection but rather the process of developing the sounds. Our findings contribute to child-produced Polish sibilants and yield further insight into the acoustic characteristics that listeners rely on in making perceptual judgments of foreign sounds.

Abstract

The Polish language has a complex sibilant structure when compared to languages like English. Of particular interest here are the alveolo-palatal and retroflex sibilants. There have been some previous studies on Polish sibilants examining production and perception of children (under 5 years). However, there is a greater need for understanding adult perception of children's productions and the perception of different populations listening to children's productions. Contributing to perception studies would, therefore, allow for a more in-depth analysis of this field of research. This paper builds on the findings of a production-perception study of Polish sibilants in typical children (Zygis et al., 2023) and expands the results by examining English and Heritage Polish population perceptions of Polish children's productions. The Zygis et al. study examined Polish children and their production and perception of the contrasting sibilants. The study looked at the perception of the children for their own production and adults' productions. Their study acquired recordings of 80 Polish children aged 35–106 months producing words with /s, ʂ, ɕ/. One of their tasks involved the child participants hearing their own productions of word-medial sibilants: /kasa/, /kaʂa/ and, /kaɕa/ at random. They then had to choose between three images (corresponding to Polish words, e.g.: kasa for cash register) to indicate the stimuli they heard. Their study found that there were a number of acoustic parameters that children used to identify sibilants. They observed that especially the younger children, "appear [to] pay more attention to formants independent of the sibilant and [that] the cue weighting [for these young children] changes during the acquisition process" (Zygis et al., 2023).

For the present study, we wanted to explore the perception of these word-medial sibilants for different phonetic environments and for non-native listener populations. The three phonetic conditions included: the whole word as in the original study, the isolated sibilant, and the (isolated) sibilant together with the preceding vowel. The audio files (taken from Zygis et al, 2023) were edited and played to both native English and Polish Heritage listeners at McMaster University in Hamilton, to

determine the perception of the three-way Polish sibilant distinction. This distinction is non-existent in English for English listeners or influenced by both Heritage and English phonetics/phonology for Heritage speakers. The sibilant distinction in English lies between /s/ and /ʃ/, therefore the task for the English native participants was to choose between buttons that indicated “kasa | as | s” (for the /s/ sibilant) or “kasha | ash | sh” (for /ʃ/) to indicate which sibilant they perceived. The Heritage speakers of Polish were English participants with varying levels of Polish fluency residing in the Southern Ontario area. They used the same design (three-way sibilant distinction) as the original study. A total of 41 English and 13 Heritage listeners participated in the study. It was hypothesized that the English native listeners would categorize all Polish alveolars as (English) alveolars, but it was not clear how retroflex and alveolo-palatal contrasts from the children’s complex productions would be resolved by the English listeners. It was further assumed that the perception of stimuli with vowel transitions (e.g., /kasa/ and /as/ in contrast to isolated /s/) would significantly differ comparing English listeners and Polish Heritage listeners. In our results, English participants increasingly categorized all manipulations of /s/ as /s/, and /ɕ/ as the /ʃ/ sibilant, especially for the older children’s productions. Their perceptions for the retroflex /ʂ/ was split, half as /s/ perceptions, across conditions. Phonetic information in the form of formants (on top of the spectral noise of the isolated sibilant) did not significantly improve distinction for the English participants. The Polish Heritage speakers showed difficulty in correctly identifying /ʂ/ variations especially in the older children. Phonetic environment and age had varying effects depending on the sibilant. As Polish Heritage participants are familiar with three-way sibilant contrasts, it was interesting to see how these Heritage speakers’ classification differed from that of English participants, especially for stimuli from children who are in the very initial stages of speech development (i.e., decreased articulatory and acoustic accuracy).

Acknowledgements

It is my honor to acknowledge the individuals who have brought me to where I am today in my Master's journey. My parents, Hyunbai Lim and Joohee Lim for their never-ending support in all my efforts and paths I have taken (and will be taking). Thank you, Mom and Dad, for being my longest benefactors despite not understanding what it is you are supporting.

I would like to extend gratitude to my supervisors Dr. Magda Stroinska and Dr. Daniel Pape. Dr. Magda has been such an enthusiastic and caring supervisor despite the busy schedules and numerous changes that have taken place for her during these last years. You have been so willing to offer your help and time in my research. In addition, your excitement to share your knowledge has only fueled my passion for linguistics to a higher level of zeal. Thank you for being such a welcoming and kind supporter to a student who was a little lost at the beginning. Dr. Pape has been an incredible driving force in keeping me motivated and on-track. He has always been more than willing to sit down and offer his guidance while sharing discussions on numerous topics. Working on my thesis with him and being a part of his lab has been such a joy and honor. Thank you for your guidance and consideration in being my supervisor and mentor. I also offer my thanks to Dr. Elżbieta Grodek as the final member in my thesis committee. Thank you for being kind enough to accept being part of my committee and acting as the chair.

To all my friends and colleagues, both in Hamilton and in Edmonton, I am very grateful for your presence and support. Especially those from ARiEAL, some of whom I have worked with in the Phonetics Lab. You are all such kind souls, always willing to help out with my experiments and my questions. And to all my other friends, be that academic encouragement or just offering time out of their day, they have made this academic journey more bearable and worthwhile. This is especially so for someone who has moved across provinces alone and felt isolated, thank you for being such a supportive network for me. I am extremely grateful for the time I have spent and the experiences I have gained in Ontario.

Table of Contents

1. Introduction

Examining foreign phoneme perception can help us to better understand the mechanisms of acoustic perception in general. The present study, building on the work of Zygis et al. (2023), aims to further our understanding of the perception of the three-way contrasting Polish[1] sibilants (dento-alveolar /s/, retroflex /ʂ/, and palatal /ɕ/) by both Heritage Polish speakers and native English speakers. By considering how these sibilants are typically acquired by Polish children and how they are perceived in different phonetic contexts, we can gain a deeper understanding of the underlying acoustic cues that are used in the perception of these sounds. Additionally, by examining how native English speakers perceive these sounds, we can observe the impact of a listener's native language (and by extension different categorical perception) on their ability to perceive non-native phonemes. By contributing to the existing knowledge of Polish sibilant perception, this study may have implications for language acquisition and foreign language perception and comprehension.

1.1 . Sibilant acquisition for typical children

Defining what is considered typical in children's development clarifies the expectations for progress and acquisition at specific age ranges. This includes various aspects such as knowledge, behaviors, and skills that are frequently interdependent on each other (Wadhera et al., 2020). It also provides an objective standard, or norm, in milestones that can be used for assessment comparisons. McLeod et al.'s (2007) metanalysis observed that typical acquisition of sibilants starts at 3 years of age but is accompanied by phonological processes, mistakes, and a sibilant lisp[2]. It is to be expected that these phonemes will be fully acquired later on in life as speech acquisition continually develops throughout one's life. Studies that target child speech (such as Li & Munson, 2016; Hardcastle & Gibbon,

[1] The standard Polish language was the main focus in this study, other Polish dialects were not analyzed.

[2] A sibilant lisp is the inability to correctly pronounce sibilants (for instance /s/ and /z/), where the airstream that should normally be directed in the middle of the oral cavity is spilled to the sides (of the tongue) (Speechlanguage-resources).

2018, and others mentioned by Miodonska et al., 2022) face the issue of inter-speaker variability. Comparative analyses have focused on intra-speaker variation, and even in inter-speaker circumstances, the population has been homogenous. Mcleod et al.'s (2007) metanalysis on child speech acquisition partially resolved this issue by Grouping children into age ranges, observing, and comparing phonological processes that occurred within and between each Group. A shortcoming to this study was the exclusion of other languages, once again highlighting the issue of homogeneity, as the studies focused solely on English-speaking children. Similarly, Zygis et al.'s study (2021) used only Polish children's productions, who then assessed their own productions, but the analysis of variability and cross-language perception was missing. One of the aims of the study presented in this thesis is to analyze the variance among the Polish children's participants as well as the perception of the English participants. There are few studies, or at least reliably translated ones, that observe the sibilant acquisition of typical Polish children. Previous sibilant studies have focused on alveolar /s/ and the post-alveolar / ʃ/ that is present in languages such as English and Japanese. In their study, Li et al., (2011) found that English -speaking children will substitute [s] for the / ʃ/ phoneme while the Japanese children do the opposite. This difference of phoneme implementation, compared to adult-like speech, can be a result of reduced tongue and articulator coordination (Zharkova et al., 2018). Although these phonemes share some qualities, they differ from the Polish dental and retroflex (Miodonska et al., 2022). Putonghua, a Chinese dialect (Li and Munson, 2016), Swedish, and select Indian languages are the closest languages that have the 3-way contrasts similar to Polish (Bukmaier and Harrington, 2016). The phonetic sequence of acquisition for Putonghua was found to be /ɕ/ → /ʂ/ → /s/ with the alveolar phoneme being fully distinct by 5 years of age (Li and Munson, 2016). In contrast, Polish children acquire /ʂ/ relatively later than the other two variants due to its "articulatory instability and weak acoustic information in the vowel for its distinction from [the other variants]" (Bukmaier and Harrington, 2016). Acoustic information in the vowel can range from

the vowel formants, centre of gravity (COG), peaks, and so on. Minczakiewicz (2017) found that, similar to the English children, 30% of Polish children had sigmatism (lisps) during their early school years. They also confirmed that retroflex sibilants were one of the last Group of sounds to be successfully acquired by Polish children. This is interesting to note as the stimuli used for our study consisted of early school age Polish children. The fact of sibilant lisps and difficulty in acquisition will have an influence on the perception results of the children's sibilant productions. A downside to Minczakiewicz's study was that the focus leaned towards speech disorders instead of typical speech acquisition. Therefore, a goal of this study is to capture sibilant perception of typical Polish children's productions.

1.2 . Analyzing the Stimuli of Sibilant Perception Tasks

Previous studies into sibilant perception have focused on sibilants in isolation, CV(Consonant-Vowel) pairs, or sibilants embedded within a word, but rarely inclusive of all three conditions. Zharkova et al. (2018) obtained sibilant productions of CV pair conditions from typical Scottish-English children. In their study, they used a LOC index "which is a ratio of the excursion of the tongue front to that of the tongue back to midsagittal ultrasound imaging data" and fricative centroid (Zharkova et al., 2018). Their study found that for all their age groups of interest, there was more front bunching of the tongue for / ʃ/ than for the alveolar sibilant. There were noticeable differences for /s/ when preceding and following /a/ in comparison to / ʃ/. This was taken to mean that there was higher discriminability for the context of the /a/ pairing and that it increased with the age of the participants. This result showed that in the CV pairings, "increased discriminability in tongue shape [started] from the beginning and end of the frication" (Zharkova et al., 2018). We were interested to see how participants in our study would respond as the VC pairings in the stimuli contain the /a/ phoneme pairings. It is also worth noting that pairing sibilants with the /a/ phoneme is ideal as it "avoid [s the] palatalizing, rounding effects, and phonologically illicit syllables as much as possible" (Zygis and Padgett, 2010).

Pairings need to be phonetically possible in the language and for Polish, the /a/ phoneme does occur with the three-sibilants of interest (Zygis and Padgett, 2010). A potential issue with Zharkova et al.'s study was their exclusive focus on English and the alveolar vs. post-alveolar voiceless fricatives, as well as the age of their participants, who were well past the age of basic speech acquisition (6 years of age according to McLeod et al., 2007). Li and Munson's (2016) study, on the other hand, found contrasting results. They varied vowel context for the Putonghua sibilants and had the fricatives in word-initial position, creating a CV pairing. Their findings showed that the first acquired / ɕ/ was distinguishable by 3 years of age, but there was no significant effect of the vowels (and therefore formants) in perception accuracy. Interestingly, their experiment involved younger children with varying stimuli, but their phonetic environment focused on CV pairings instead of our intended Vowel-Consonant (VC) order in our study. Zygis et al.'s (2021) study noted that the alveolar-palatal sibilant could be distinguished by formant values (particularly the F2 as well as the F3). However, many studies make use of the CV pair order with the vowel proceeding the focus sibilant phoneme, such as the aforementioned Zygis et al. (2021) study. As well those from Bukmaier & Harrington, (2016) and Li & Munson, (2016). By manipulating the stimuli to have VC pairs instead of the typical CV order, we can observe effects of the preceding transitioning vowel formants on the proceeding sibilant's perception. A study that analyzed Polish sibilants in the VC phonetic environment is from Zygis and Padgett (2010). The 4-way-contrast of [s], [ʂ], [ɕ], and [ʃʲ] was analyzed; the latter categorized as an allophone of [ʂ]. They analyzed reaction times (RT) in discrimination tasks and instructed participants to react as fast as possible to "minimize language-particular differences" (Zygis and Padgett, 2010). This was done to observe processing at the auditory rather than phonetic level. Their Polish participants' reaction times were remarkably similar in terms of the three distinct phonemes [s, ɕ, ʂ] with [s- ɕ] block pair being the longest (but not significantly different). This longer RT is indicative of greater perceptual difficulty implying that they are perceptually close sounds when it comes to

perceptual mapping (Zygis and Padgett, 2010). From their analysis, [s] and [ɕ] may be perceptually similar, leading to the longer RT. Additionally, Padgett and Zygis found a centre of gravity (COG) pattern within the perceptual mappings that has [s] and [ɕ] or [s] and [ʂ] together. As [ɕ] and [ʂ] differ in COG perceptual mapping, then it may be that they are easier to distinguish between the two than it is with the addition of the [s] phoneme. It will be noteworthy to compare the results from our own study to see if there are any significant differences in the vowels for the same VC phonetic conditions. The study presented here is different, from our own study, as the allophone [ʃ] was the focus and their procedures and stimuli varied (ex: unedited stimuli, exclusive focus on VC/CV environment, instructions in Polish, established non-varying stimuli blocks: sa/ɕa, ɕa/sa, sa/sa, ɕa/ɕa; same vs different stimulus order). Therefore, it is valuable to conduct further research to observe conditions that were missed in previous studies. Changing the sibilant positions to word-medial or word-final position resulted in an effect to the fricatives' characteristics via cues in coarticulation (Miodonska et al., 2022). Due to varying results of the influence of phoneme neighborhoods, it will be noteworthy to observe whether there will be similar (lack of) significance in phonetic context for non-native (and native) sibilant distinction in our own experiment. Many of the previous sibilant studies have a degree of exclusivity of phoneme conditions (such as looking solely at words, sibilants, or CV pairs). The aim of this study is to analyze sibilant perception in all manipulated phoneme conditions (inclusive of the three: VC pair, word, sibilant in isolation). Including all three conditions will allow us to examine which conditions lead to improved (more accurate) perception distinction. We can also see whether there is an effect that agrees with or differs from previous studies.

1.3 . Non- Native (English) Speakers' Perception and Categorical decision-making

From birth, humans can discern all possible speech sounds, but this ability narrows according to the parameters of the individual's native language and immediate environment by the first year (Santrock et al., 2021). However, when it comes to

production, according to McLeod et al. (2007), children are not fully intelligible to strangers until at least 6 years of age. This is also when their sibilant lisps (for the English children cohort at least) disappear. Taking these facts into consideration, it is of interest to observe how children's speech sounds are perceived by non-natives. One phenomenon to keep in mind is categorical perception, which states that internal categories of the listener will influence their sensory perception (Goldstone and Hendrickson, 2009). Categorical perception accounts for our human need to categorize what we perceive through perceptual adaptations. These adaptations may at times warp our senses into accentuating or deemphasizing aspects of the actual stimuli. The timing of the adaptations may also vary depending on the task and/or individual (Goldstone and Hendrickson, 2009). It will be difficult to capture and analyze individual adaptations, but task-related adaptations, if any, are more feasible to observe and analyze. Goldstone and Hendrickson claim that in the notion of equivalence classes, "when we place two objects into the same category, we do not treat them as the same thing for all purposes and [they] can be different in different contexts" (2009). Extending this into our area of interest, it may mean that participants treat the perception of the sibilants differently, despite an established category, due to varying phonetic contexts. Another notion to consider is the difference in "fragile" vs "robust" contrasts, in which the former is stated to show "perceptual decline [when] phonetic or phonemic exposure is lacking and are less amendable to training in adulthood" (Polka, 1991), while the latter is easier to modify especially in non-native speakers. Polka found that adult English speakers assimilated the 4-way-contrasting-Hindi-retroflex-stops into existing English categories of /t/, /d/, and /t^h/. This seems to confirm fragile contrasting as the monolingual English speakers lacked any training as part of the study (and we can assume they have limited exposure of the Hindi contrasts) (Polka, 1991). An analysis and comparison of participants' exposure to other language could have shown transfer effects, especially as we know that certain Indian languages (e.g., Sanskrit) share phonetic contrasts with European languages. Conversely, Tamminen

and colleagues (2015) tested the robust contrasts by training participants on non-native speech categories to observe effects on perception. They found that after just three days of training, there was a significant improvement in discrimination and reaction times (Tamminen et al., 2015). However, it is worth mentioning that the training was on fricative voicing contrasts that did not exist in Finnish (the contrast was not on phonetic/phoneme variations). Dehaene-Lambertz and Gilga (2004) add another dimension to non-native perception with issues of repetition (especially in experimental designs). Despite non-native stimuli being a type of new exposure, if the listener perceives multiple stimuli within an extended period (in one session) and categorizes them as the same, this could lead to a decreased response (both neural and behavioral) (Dehaene-Lambertz et al., 2004). Taking all these notions into account, it will be interesting to note the perceptual results of non-native contrasts produced by children. Zygis and Padgett's (2010) study found that their native Polish participants did very well in distinguishing the phonemic contrasts [s, ɕ, ʂ] and that their non-native (English) participants did well for distinguishing [s] from the unfamiliar postalveolars. However, as expected the researchers found that the English participants had difficulty distinguishing within the unfamiliar postalveolars. The reaction times (RT) of both participant groups were also similar (in the CV pair conditions) with the English speakers expectantly having slower RTs. In the contrasting pairs, [sa] vs. [ʂa] had faster reactions times (657 ms) than [ɕa] vs. [ʂa] (812ms) which was anticipated as English does not have the [ɕ] vs. [ʂ] contrast. There was, however, no difference with respect to the VC positions for the English group. Yet the [ɕ] vs. [ʂ] pairing was surprisingly the longest RT contrast for four of the Polish participants (in the VC condition) (Zygis and Padgett, 2010). Zygis and Padgett concluded that sibilant properties and formant transitions were vital for sibilant perception. This will be relevant for our own observations, since both sibilant properties and formants could be judged, independently or combined, by our participants. As previously mentioned, it is important to account for background language experience considerations. As there may be transfer effects from L1 (or

preceding) languages into even the perception of the new language (Flege and Davidian, 1984). Those with exposure to the three-way-sibilant-contrast, such as participants familiar with Polish, Chinese, Swedish, or Indian languages, will be expected to have better perceptual abilities. For this purpose, for our own study, we have recruited English native speakers (with no Polish 3-way-sibilant-contrast-equivalent language fluencies) to perceive and distinguish the Polish sibilants. This would allow us to examine how phoneme contrasts produced by children are distinguished between natives and non-natives.

1.4. Summary and Hypothesis

In summary, the goals of this study are to contribute to the knowledge of sibilant perception in typical Polish children's productions. To achieve this, we have included Polish sibilant productions from native children and manipulated the phonetic environments. This will allow us to assess the Polish Heritage speakers' perceptual accuracy of the sibilants in different phonetic environments. We can also observe the perception of native English speakers of this three-way contrast, which does not exist in the English IPA. Additionally, the interaction (if any) of the two phonologies, English and Polish, can be assessed. Altogether, this study will ultimately contribute to the knowledge of Polish sibilant perception for both Heritage speakers and non-native populations.

Based on previous work and literature, we hypothesize that the English native speakers will categorize the /s/ manipulations as /s/ and the other foreign Polish sibilant phonemes as /ʃ/. These results should change in accordance with the age of the informants as they possess immature articulation (due to incomplete development) and still lack production of typical adult-like articulations. It could have been assumed that responses for younger informants' stimuli perception will therefore vary, leaning towards the /ʃ/ phoneme as sibilant lisps are prevalent for children in the younger age range (Minczakiewicz, 2017). However, as Zharkova et al., (2018) found increased discriminability with the /a/ phoneme in VC pairings

across age Groups, it will be interesting to see the difference in perception according to the varying phonetic environments. As for the participants with Polish fluency (Heritage speakers), we expect better than chance identifications for informants aged 5 years and 6 months to 6 years and 3 months and beyond due to the consideration that children have successfully acquired all phonemes in their native language and the absence of the sibilant lisp (Mcleod, 2007). It remains to be seen if a variation in phonetic environment will affect the perceptual results. Participants with lower fluency in Polish are expected to yield results like those of English native speakers (with no Polish). Both the palatal /ç/ and the retroflex /ʂ/ would not be distinguishable as separate phonemes at the reported stage of the participants' acquisition of and limited exposure to the contrast under investigation in this study.

2. Methods and Observations

2.1. Experimental design

There were two versions of the experiment conducted for this thesis with each version consisting of two parts. The first part in each experiment involved the completion of forms which included a questionnaire for the participants. The questions inquired about: age, gender, known languages, the fluency level of each language, place of acquisition, age of acquisition, handedness, sight/ hearing and/or language disorders, and musical training. The other forms involved COVID information and study background information. The participants' private data was stored in a locked cabinet. The second part of the experiments involved the speech perception portion. Audio for the speech perception was edited using Praat (version 6.2.13, Boersma & Weenink 2022). The original audio files came from Zygis et al's (2021) experiment of 81 child participants' recordings of three Polish words: [kasa], [kaʂa], and [kaɕa]. The original audio was cut so that there were three versions for each child's recording of each Polish word. The original audio containing the whole word was used without any editing. Each version had the whole word, the sibilant on its own, and the preceding vowel with the sibilant (VC pair). Praat was used to cut the audio utilizing the spectrogram plus oscillogram features. As there were 81 children from the original audios, there were three versions for each child's recording. Therefore, after all the audio edits, each word, containing our sibilant of interest, had 243 recordings. Since there were 3 words, the total came to 729 recordings which would all be used for the perception experiment.

The program used to design and run the experiment was Gorilla (Cauldron Science 2016). using all the project and task builder features. Gorilla is a cloud-based experiment builder made for researchers and students to conduct behavior-based experiments. Gorilla was used due to our study taking place during the pandemic and the potential of testing online due to lockdown procedures. The first version of the experiment, which we referred to as the English version, started with a practice run, a countdown, then the actual experiment where participants would hear one

stimulus and be shown a forced-choice two button option, i.e., the participants had to choose on of the two buttons to proceed (and could not choose e.g., “I don’t know”). The buttons were labelled: “kasa |s | as” and “kasha | sh | ash” (as these options were the choices participants could make either the full word, the orthographic representation of the isolation fricative, or the VC sequence). Participants could only continue when they had decided between the two (Figure 1).

Figure 1: Picture illustrating the two-way forced selection buttons and the labels.

There was no time limit, and a pause screen was placed after each task screen with a “continue” button for when the participants were ready to move-on. The stimuli played automatically within each task screen, with no option for replay. Once the participants finished all 729 audio stimuli, they were notified on the screen that the experiment was finished and were guided to the exit. The second version of the experiment, referred to as the Polish version, was very similar except for the set-up of the buttons. As this experiment built up from a previously conducted experiment by Zygis et al (2023), the set-up was accurately replicated to ensure consistency. This version of the experiment had the buttons set up with images of a cash register, a bowl of groats, and a doll. These images were used to maintain consistency from Zygis and colleagues’ (2023) experiment set-up. No labels were placed on the buttons, again to replicate the Zygis et al. (2023) study. Everything else followed the same procedures as the English version: participants completed a questionnaire, got a practice run, a countdown, the actual experiment portion, then a finish screen. Participants were played a recording of the three sibilant contrasts, from the IPA website consonants sounds on a phone audio recorder. In addition, they were given

a labelled diagram of the three-way buttons for them to refer to throughout the experiment (Figure 2).

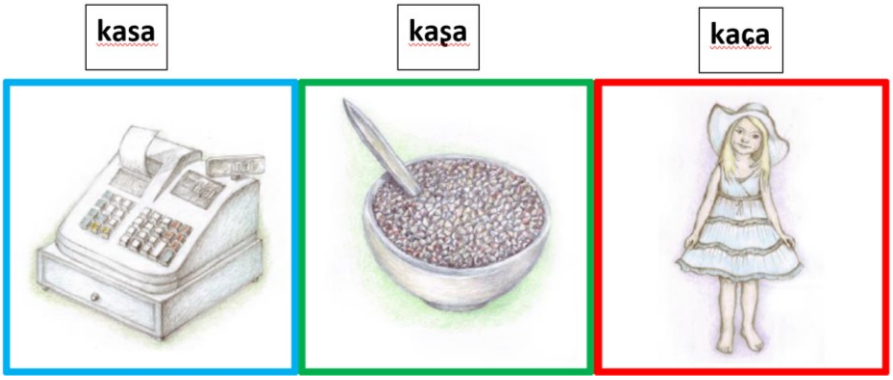

Figure 2: Replica picture, from Zygis et al., (2023), illustrating the minimal triplet: /kasa/ "cash point, /kaʂa/ "groats", and /kaca/ "Cathe, prop.name".

The set-up took place in a sound treated room (treated with acoustic absorption foam) in the Phonetics Lab which is located in McMaster's ARiEAL centre (Figure 3). A laptop was connected to a set of Sennheiser HD 599 Open back ivory headphones which in turn was connected to a Focus rite Scarlett Solo 3RD Gen USB audio interface. A Logitech wired mouse was also connected for participants to utilize. Data was automatically stored in Gorilla automatically but also downloaded and stored on an external 1 TB hard drive, as well as on the ARiEAL cloud server.

Figure 3: Picture of the Testing Setup for Participants

The only exception to this standard set-up was that for some of the paid Polish Heritage participants the location/room where the experiment was conducted was at a restaurant (before and during operating hours, to ensure a quiet environment) but the equipment was the same. This was done due to accessibility issues for these Heritage Polish participants.

2.2. Participants

Participants were recruited using the SONA credit system at McMaster University with the approval of the McMaster Research Ethics Board (MREB) and with additional ARiEAL approval for the paid participants. For-credit participants used the SONA system to sign up for experiments and were compensated by bonus SONA credits for one hour of the study (study lasted a total of 1 hour). Paid participants were recruited via online advertisements. The criteria for involvement in the English version required participants to be native speakers of English with no conditions that would prevent them from naturally hearing or seeing the stimuli from the computer set up. Due to this, there were four participants whose data could not be used as they had not qualified (through their own admission). One of the participants did not finish the experiment, their data was incomplete and excluded for analyses. In another instance, a participant had to take a longer pause due to a computer audio malfunction, but their data was intact and included. Therefore, in total there were 37 English speaking participants whose data was included for analysis. For the Polish version, the criteria remained the same but included the requirement of some fluency of Polish. Most of the Polish participants were paid participants and compensated $15 for one hour of the study because they did not qualify to receive SONA credit bonuses. There were 10 Heritage Polish speakers and three Learners of Polish. Thus, the total number of participants in the Polish version was 13. The data collected through the experiments was first transferred and sorted from Microsoft Excel then R Studio Software for analyses (version 2023.03.0+386, RStudio Team 2023). The programs used for analysis involved the ggplot function to plot the

English participants responses to each sibilant category: kasa, kasia, kasza. The plots were further organized into “Audio Type” (stimulus type) which varied by: VC pair, isolated sibilant, or word. The Polish participants’ responses were also plotted using a stacked ggplot to see correctness. Additionally, their data was further subdivided into Heritage speakers vs. Learners. The Heritage speakers were considered the baseline for comparison purposes (to the Zygis et al. study). Similar graphs were plotted: by “fricative” (the fricative presentation) and “Audio Type” (stimulus type) but in terms of percentage correct (to “Response”). Corresponding ages of the children participants were added using the R studio program *fuzzyjoin* program (Viviario, 2023). This was then applied to another ggplot for the English participants for each “Audio Type” (stimulus type) showing their “responses” (/s/ vs /ʃ/) according to each sibilant (fricative representation) “kasa” vs. “kasia” vs “kasza.” Due to the small sample size of our Learners, their data was not included for discussion or analysis. Any interpretation of such small data would have to be taken with much caution as small sample sizes results are not generally applicable to the target population.

3. Results

In this chapter, the results of the English participants will be discussed first, followed by the Polish Heritage participants. Within each participant group, the results are ordered as follows: the perceptual proportion results (accuracy for the Polish) showing the response for each fricative within each stimulus condition (in one array). Then, we present the results considering the children's age, and plot them with fricative and stimulus condition (in multiple arrays).

3.1. English Perceptual Proportion Results

Figure 4 exhibits the perception results from the English participants for the response for each fricative (in the order /kaɕa/, /kasa/, and /kaʂa/) in the x-position. The responses, as stated in the methods section for the English Group, are divided into responses for /s/ in red (at the bottom) or /ʃ/ in blue (stacked on top). The y-axis shows proportions out of the total responses for each fricative. The correlating response counts are displayed inside each response bar. Each plot is labelled with the stimulus condition in this exact order: Sibilants, VC pair, and word. We observe that across all stimulus conditions, as expected, that the Polish /kasa/ fricative had the highest number of English /s/ response proportion. The /s/ responses were all above 75% of the total responses for each stimulus condition. The /kaɕa/ fricative showed the lowest proportion for /s/ perception (less than 25% of the total), followed by the /kaʂa/ fricative with approximately 50% /s/ responses across conditions. The results do not appear to differ significantly between stimulus conditions (i.e., isolated fricative on the left, VC condition in the middle, complete minimal pair word on the right) for the same fricative. Within each fricative, there seems to be a small effect that the sibilant condition seems to render more /s/ responses for /kaɕa/ (Figure 4 far-left plot). The VC pair for /kaɕa/ follows (middle plot) with the word condition being the last (far-right plot). There seems to be a tie for /s/ proportions for the /kasa/ fricative in all conditions Lastly, for /kaʂa/, the sibilants again, like for /kaɕa/, have the

highest /s/ response, followed by the word stimuli, and then the VC pair being the lowest.

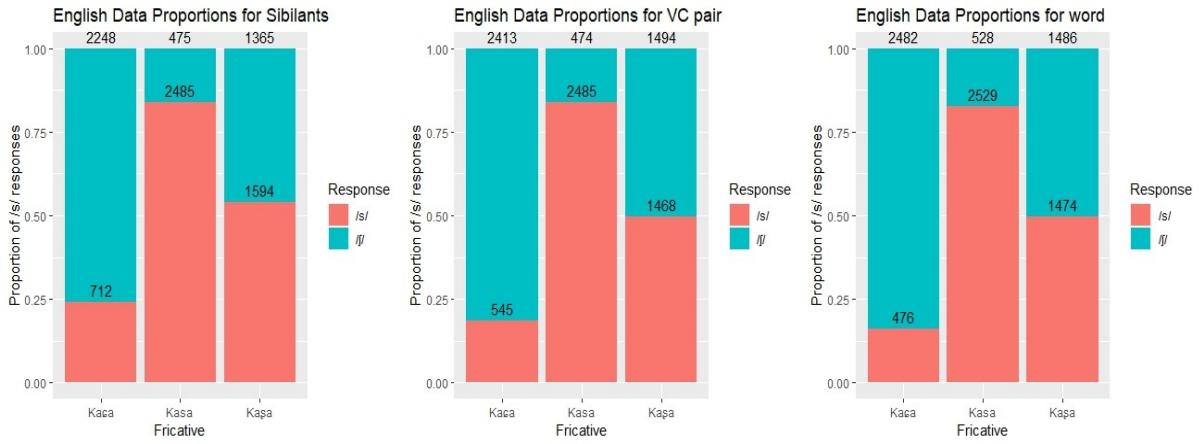

Figure 4: English Perceptual Results. X-axis with the different fricatives (/kaʦa/, /kasa/, and /kaʂa/) and y-axis with the proportion of /s/ responses with the case count labels (i.e., how many items occurred per plotted condition). Response is color coded for each fricative as /s/ or /ʃ/.

3.2. English Results with the Children's Age Factor

Figures 5 to 7 are similar to Figure 4 but have the additional consideration of the children's ages (of the spoken stimuli presented). The children's ages are sorted according to the Groups used in Zygis and colleagues' (2023) paper, which are as follows: Group 1 includes ages 2 years 11 months to 3 years 9 months, Group 2 comprises 3 years 10 months to 4 years 7 months, Group 3 covers 4 years 8 months to 5 years 5 months, Group 4 encompasses 5 years 6 months to 6 years 3 months, and Group 5 includes 6 years 4 months to 7 years 11 months (refer to Table 1 in Appendix). The stimulus conditions had to be separated due to the ages, and thus each condition had its own array of plots for the three fricatives of interest. The order of the stimulus conditions remains consistent with previous plots, where it is sibilant, VC pair, and word order. In Figure 5, the sibilant conditions for the English participants are displayed, and overall, we can see that /kasa/ elicits the most /s/ responses across all ages. There is a progressive trend as the older children's stimuli are increasingly categorized as /s/ (and less as /ʃ/). In the final age group, Group 5, there are only about 3% /ʃ/ perception responses out of the total stimuli for this age group. On the other hand, the majority of responses for /kaʦa/ resulted in /ʃ/

across all age groups. The proportion for /s/ perception of /kaɕa/, excluding the total, appears to have a general downward trend with increasing age. The age group that has the highest percentage /ʃ/ is as follows: Group 4 (86%), Group 5 (82%), Group 3 (81%), Group 2 (70%), then Group 1 (67%). The results of the English participants' perception show that the /kaɕa/ sibilant stimuli generally sounded more like /ʃ/ in an increasing manner for the voices of older children. For the / kaʂa / fricative, when looking exclusively at the /s/ perceptions, increased age leads to a decrease in /s/ proportion. Thus, conversely, we can see that with increasing age, there is an increase in the /ʃ/ proportions. It appears that before Group 3, the participants perceived the / kaʂa / sibilant stimuli as /s/ representing 69% and 67% of the total in Group 1 and 2, respectively. However, past that age, the results flip to favor the /ʃ/ perception (/s/ of the total for Group 3, 4, 5 = 48%, 40%, 23% respectively).

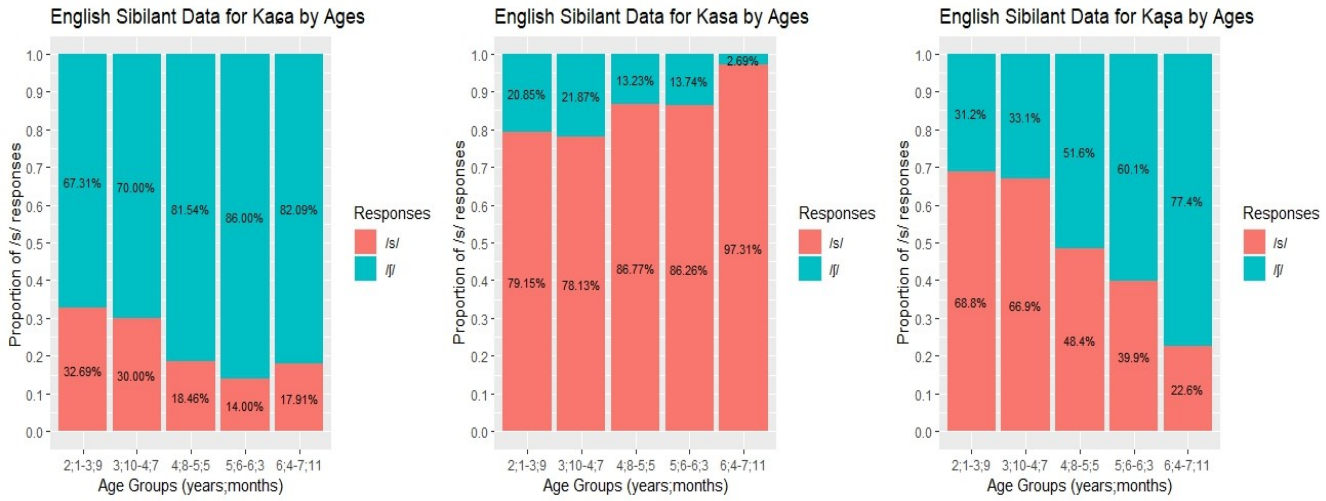

Figure 5: English isolated sibilant Perceptual Responses as a Function of Age Results. X-axis with the different age Groups and y-axis with the response proportions. Response is color coded for each fricative as /s/ or /ʃ/. Each plot shows sibilant perceptions for respective fricatives (/kaɕa/, /kasa/, and /kaʂa/).

For the VC pair stimulus condition in Figure 6, there is almost no difference for the last two (VC pair and word condition) plots compared to the first plot (isolated sibilant). Thus, most importantly, the VC and word conditions do not seem to have any effect but we will nevertheless describe them. The /kasa/ fricative (middle plot) had the highest /s/ proportions across all age groups and showed a downward trend of /ʃ/ selection with increasing ages. For the final group, Group 5, there were only

1.49% of /ʃ/ out of the total. The proportion ranges for /kasa/ between the sibilant (Figure 5) and VC pair (Figure 6) conditions are quite similar for both /s/ and /ʃ/. This similarity is also confirmed in Figure 4, where the /kasa/ proportions are the same for both the sibilant and VC pair conditions. The /kaɕa/ VC pair proportions in Figure 6 have more /ʃ/, again similar to the sibilant condition in Figure 5. This pattern can be observed in Figure 4, where the VC pair /kaɕa/ has more proportion of /ʃ/ (and thus fewer /s/) than the sibilant /kaɕa/ condition. It is apparent that for the /kaɕa/ fricative in Figure 6, Group 4 seems to have the least /s/ (6%) and, in contrast, the highest proportion of /ʃ/ (94%). Other than this one group, the pattern follows that with increased age, the /s/ proportions decrease (Group 1 = 25%, Group 2 = 24%, Group 3 = 16%, Group 5 = 13%). In another interpretation, the /ʃ/ thus increases with age for /kaɕa/ VC pair proportions. This was the same group in Figure 5 that caused a small dip in the downward trend of /s/ (or peak in the /ʃ/ pattern). For the /kaʂa/ plot in Figure 6, Group 2 seems to be the odd peak in the trend with 68% /s/ proportions (or 32% /ʃ/). Excluding Group 2, a decreasing pattern appears for /s/ (or an increasing for /ʃ/). It can also be interpreted, similar to the sibilant condition, that English participants increasingly perceive the /kaʂa/ VC pair as /s/ (with /s/ representing 59% of the total for Group 1 and 68% for Group 2) until Group 3 (/s/ of total of 44%). From Group 3 onwards, the stimuli are increasingly categorized as /ʃ/ (with /s/ representing 37% for Group 1 and 18% for Group 5). Note that in Figure 4, both the sibilant and VC pair /kaʂa/ conditions appear to be around halfway (~50%).

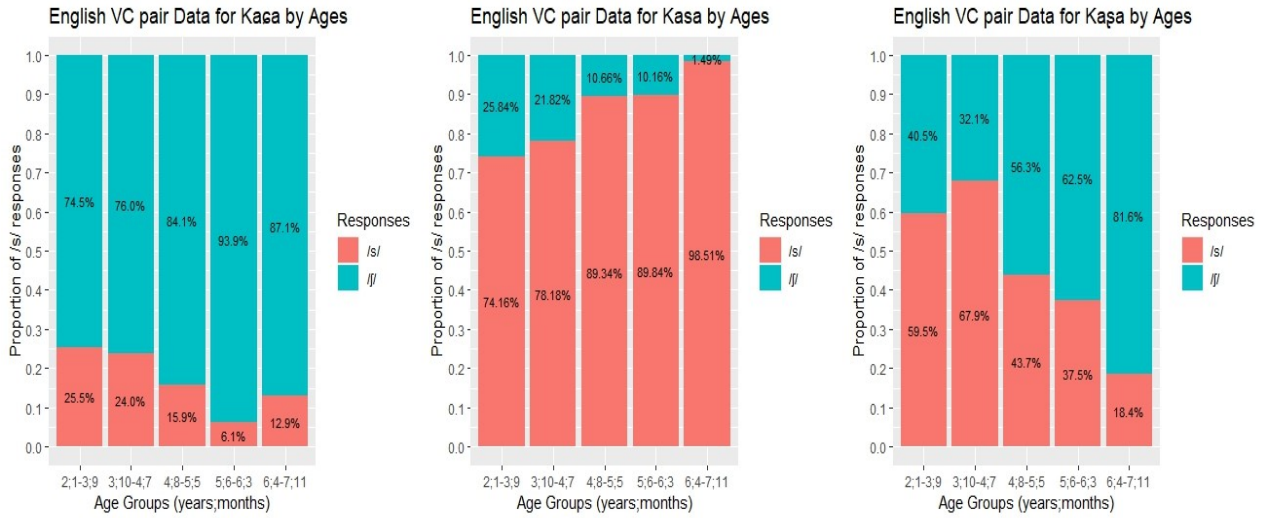

Figure 6: English VC Pair Perceptual Proportions as a Function of Age Results. X-axis with the different age Groups and y-axis with the response proportions. Response is color coded for each fricative as /s/ or /ʃ/. Each plot shows VC pair perceptions for respective fricatives (/kaɕa/, /kasa/, and /kaʂa/)

In our final word stimuli condition, Figure 7 exhibits the exact same patterns as the previous conditions. The fricative /kasa/ has the highest /s/ proportions across all age groups, while /kaɕa/ has the lowest /s/ in all age groups. Generally, there is an increase in the /s/ perceptions for /kasa/ as the children get older (with the highest /s/ proportions of the total for Group 3 and 4, both at 90%), except for the final Group 5 (where the /s/ has a total of 85%). This exception deviates from the previous patterns of a general increase. Interestingly, in Figure 4, the /kasa/ word condition had the highest /s/ proportions, but this is not true when comparing Figure 5 to 7. We can observe that the VC pair and sibilants have the highest /kasa/ /s/ proportions in subsequent figures. In Figure 7, there is a downward trend for /kaɕa/ /s/ perceptions, starting with 22% of the total for Group 1 and decreasing to 8% for Group 5. Therefore, with increased age, the perception for /ʃ/ increase (reaching a peak at Group 4 with 96%), with the exception of Group 5, which shows a slight decrease (/ʃ/ at 92%). Group 4 also leads to exceptions in the trend for the other stimuli conditions. When comparing /kaɕa/ across all stimulus conditions, we observe that the word condition has the lowest /s/ proportion. The sibilant /kaɕa/ appears to have the highest proportion of /s/. These observations can be confirmed in both Figure 4

and when comparing Figures 5 to 7. Like the sibilant and VC pair conditions, the word condition for /kaʂa/ can be observed to either have a general decrease in /s/ proportions or an increase until Group 3. In the overall decrease interpretation, we see that Group 1 starts with 58% of /s/ and drops to 17% in Group 5. Group 2 is the exception, with a sudden peak of /s/ at 70%. In the increase to decrease interpretation, Groups 1 and 2 display an increase in /s/ categorization with 58% and 70% respectively. We then see a dive in Group 3, with /s/ representing 46% of the total, and a shift in perception from this age onwards to the /ʃ/ sound. Overall, trends seem to stay consistent within each fricative across stimulus conditions. Patterns remain the same, down to the exact group that shows the sudden peak (Group 2) or dive (Group 3) in /kaʂa/. In relation to the other /kaʂa/ stimulus conditions, the VC pair condition has the lowest peak (68 % of /s/), which aligns with Figure 4. The sibilant condition has the highest /kaʂa/ proportion of /s/ in Figure 4, but its peak at 69% for /s/ responses in Figure 5 is lower than the word condition's 70% in Figure 7. Therefore, there is less consistency for /kaʂa/ when comparing the figures with each other.

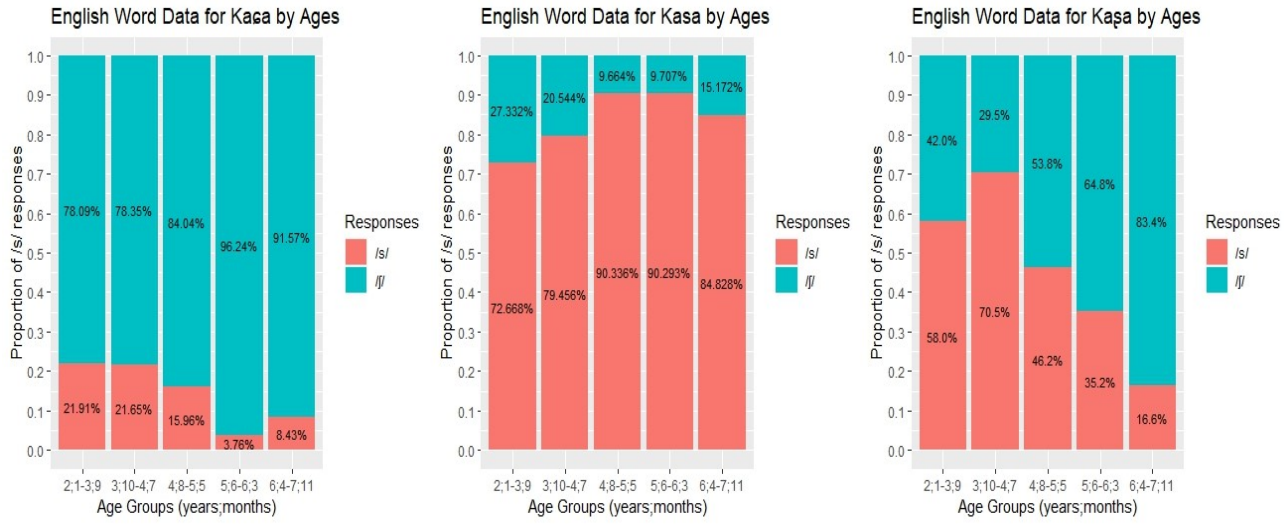

Figure 7: English Word Perceptual Proportions as a Function of Age Results. X-axis with the different age Groups and y-axis with the response proportions. Response is color coded for each fricative as /s/ or /ʃ/. Each plot shows word perceptions for respective fricatives (/kaɕa/, /kasa/, and /kaʂa/)

3.3. Polish Heritage Speakers Accuracy Results

In this section, we will discuss the results of the Heritage Polish speakers. The plots for the Polish participants depict the proportion of correct responses (accuracy) instead of proportion of responses, unlike in the English plots. The correct responses are on the bottom (teal-colored bar) and incorrect responses are stacked on the top (fuchsia-colored bar).

Figure 8 shows the accuracy of the Polish Heritage speakers for the three fricatives in each stimulus condition. Across all three plots, we can see that the /kasa/ fricative has the highest accuracy, with all values above 60% correct. Within the /kasa/ fricative, the sibilant condition yields the lowest accuracy. The /kasa/ word condition has the highest accuracy, with 75% accuracy. The /kaʂa/ fricative, across all stimuli conditions, shows the highest rate of incorrect responses (below 20% correct responses), indicating the lowest accuracy. The remaining fricative /kaɕa/ exhibits the opposite effect: the sibilant condition had the highest accuracy, while the word condition had the lowest. Therefore, for the fricatives /kasa/ and /kaʂa/ (to a much lesser extent), the increased phonetic information available does seem to lead to changes in accuracy. However, for /kaɕa/, the presence of additional information actually, surprisingly, results in less accurate responses.

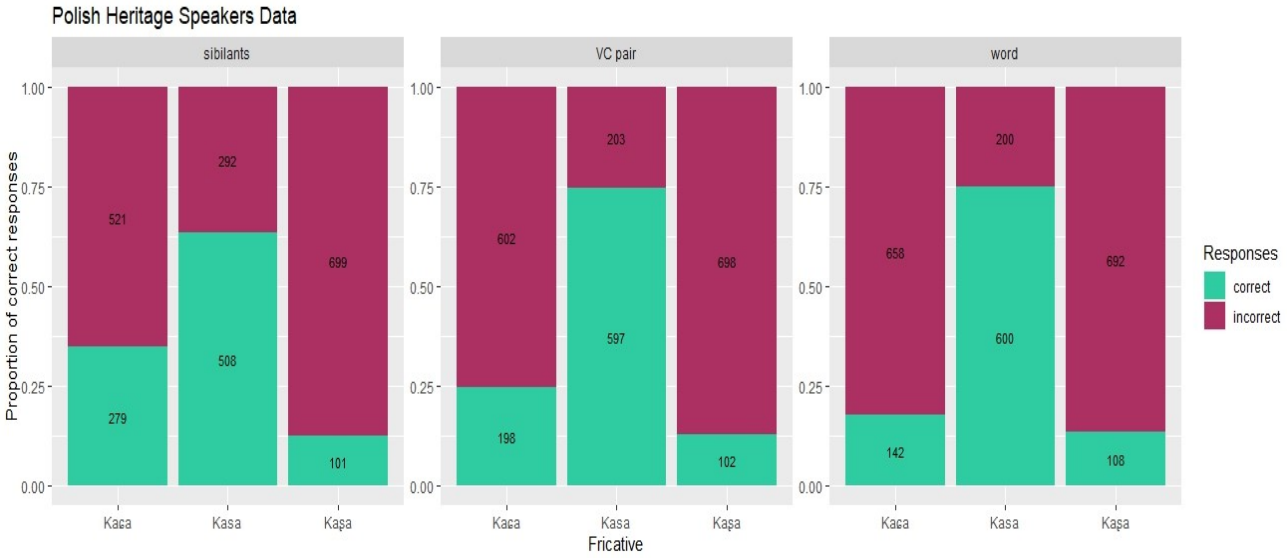

Figure 8: Polish Heritage Speakers Accuracy Results. X-axis with the different fricatives (/kaɕa/, /kasa/, and /kaʂa/) and y-axis with the accuracy in percentage scales. Response is color coded for each fricative as correct (bottom teal bar) or incorrect (top fuchsia bar).

3.4. Polish Speakers Results with Children's Age Factor

Adding children's ages into consideration, Figures 9 to 11 illustrate the accuracy according to the children's ages for the Polish Heritage speakers. Figure 9 specifically focuses on sibilants for the three fricatives of interest. In the sibilant condition, we observe that the /s/ fricative yields the highest number of correct responses across all age groups, while /kaʂa/ has the lowest accuracy results. When considering proportions, the accuracy appears to rise with age for the /kasa/ sibilants (middle plot). The accuracy starts at 59% for Group 1 and reaches 76% in Group 5. In contrast, the /kaʂa/ sibilants (right plot) show a reverse pattern where accuracy decreases with increasing age. The correct percentage for the /kaʂa/ sibilant starts at 14% in Group 1 and falls to 10% in the final group. It seems that for this fricative, the older the group the lower the categorization accuracy. There does not appear to be a general pattern for the /kaɕa/ sibilant as the accuracy rates range from 32% to 38% in a seemingly random order. It is worth noting that Group 3 had the most accurate results (38%), while Group 2 and 5 had the lowest (31.67%). However, the difference in accuracy ratings does not appear significant enough to conclusively set apart specific groups. It is noteworthy that this pattern was observed in Figure 8 and 9, where the /kasa/ sibilant had the highest accuracy among the fricatives. Additionally, the fact that the /kaɕa/ fricative in the sibilant condition had average results and the /kaʂa/ fricative had the lowest accuracy remains consistent in all the figures (8 and 9) so far.

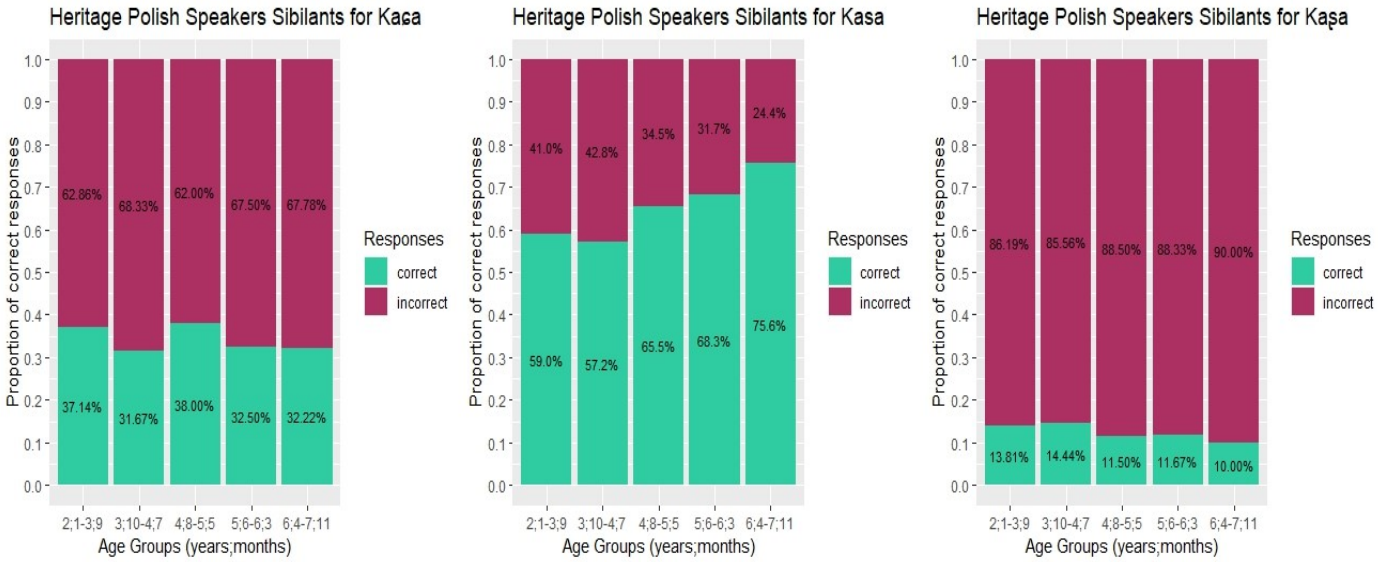

Figure 9: Polish Heritage isolated sibilant Accuracy as a Function of Age Results. X-axis with the different age groups and y-axis with the accuracy response percentages. Response is color coded for each fricative as correct (bottom teal bar) or incorrect (top fuchsia bar). Each plot shows sibilant perceptions for respective fricatives (/kaɕa/, /kasa/, and /kaʂa/)

In Figure 10, the VC condition, the overall pattern remains, with /kasa/ exhibiting the highest accuracy and /kaʂa/ the lowest. There is a deviation in the /kasa/ accuracy for Group 3 (84%), but the accuracy generally increases with Group 5 peaking at 94% accuracy. Compared to the sibilants, the VC pair for /kasa/ shows much higher accuracy rates (63 to 94%). For /kaʂa/, Group 3 (3.50%) again deviates from the pattern, and without this group, there is a decreasing trend for /kaʂa/ with increasing ages. Group 1 starts with 22% accuracy, and it drops to 4% in the final group. The range of accuracy for /kaʂa/ is more varied in the VC pair condition (4 to 22%) compared to the sibilant-only condition (10 to 14%). Lastly, the /kaɕa/ VC pair does not exhibit a distinctive trend, similar to sibilants, with a small difference in the accuracy range (21 to 29%). Referring back to Figures 8 and 9's VC pair plots, we observe that the order of accuracy remains the same, with the highest accuracy for /kasa/, followed by /kaɕa/, and finally /kaʂa/ with the lowest accuracy.

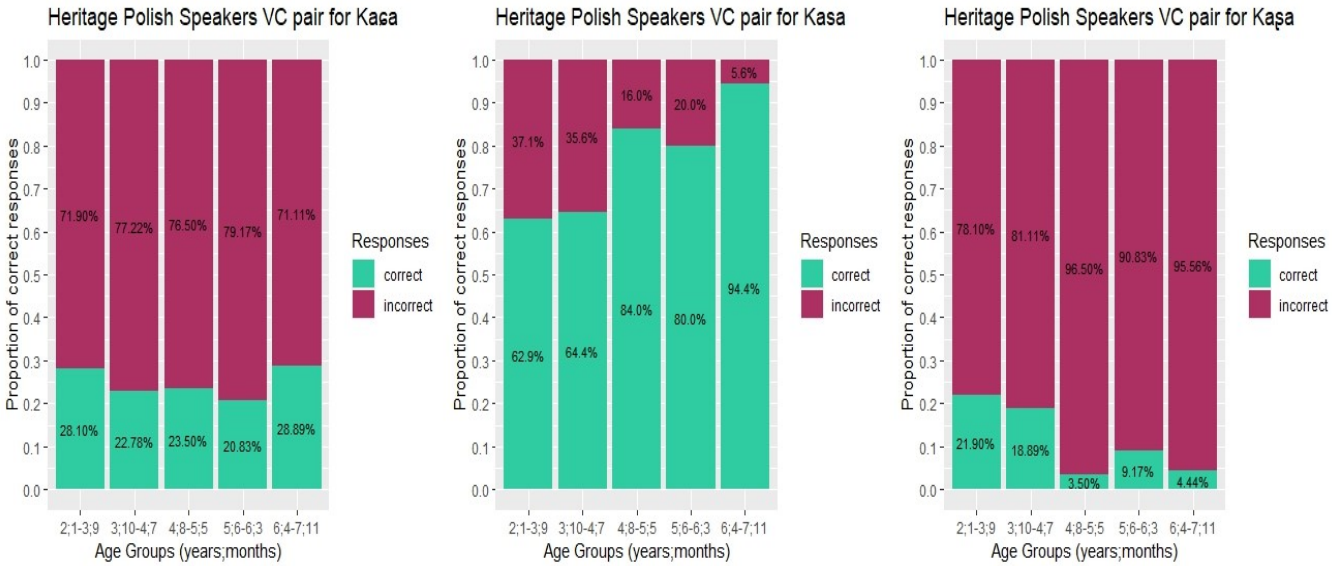

Figure 10: Polish Heritage VC Pair Accuracy as a Function of Age Results. X-axis with the different age groups and y-axis with the accuracy response percentages. Response is color coded for each fricative as correct (bottom teal bar) or incorrect (top fuchsia bar). Each plot shows VC pair perceptions for respective fricatives (/kaɕa/, /kasa/, and /kaʂa/)

In Figure 11, we can observe the results for the word condition in the Heritage Polish speakers. Similar to the other stimuli conditions, the /kasa/ fricative has the highest accuracy among all the age groups in the word condition. The accuracy increases from Group 1 (62%) to an impressive near-complete accuracy (99%) in the final group. There is a slight dip in accuracy from Group 3 (85%) to Group 4 (83%), which was also observed in the VC pair condition for the same group. Another similarity we can observe is that the word condition is the highest accuracy both in Figure 11 and in Figures 8. The lowest accuracy in Figure 11 is for the /ʂ/ fricative. Previously, the pattern for /kaʂa/ showed a clear decrease with increased age groups. However, Figure 11 seems to be an exception as there is a decline from Group 1 (26%) to Group 3 (5%), followed by a small uptake from Group 4 (6%) to Group 5 (7%). The other stimuli conditions did not show a clear upward trend, with perhaps one group as an exception, but nothing that indicated an increase as seen in Figure 11. Nevertheless, as the proportion of the increase is small, it may be dismissed as a significant change. The /ʂ/ fricative for the word condition, although the lowest between the fricatives, yielded the highest results among the stimulus conditions.

This is also seen in Figure 8, where the word condition (far right) yields the highest accuracy within the /kaʂa/ fricative comparisons. Lastly, the /ɕ/ fricative appears to finally have a noticeable trend. There was no clear pattern in other figures, but for Figure 11's word condition, a discernable trend emerges. With the exception of Group 4 (13% accuracy), the increase in age groups resulted in a decrease in accuracy for the /kaɕa/ word condition. Group 1 starts at 20% accuracy and drops to 14% in Group 5. Among the /kaɕa/ fricative, the word condition consistently shows the lowest accuracy results in both Figure 8 and when comparing Figures 9 to 11.

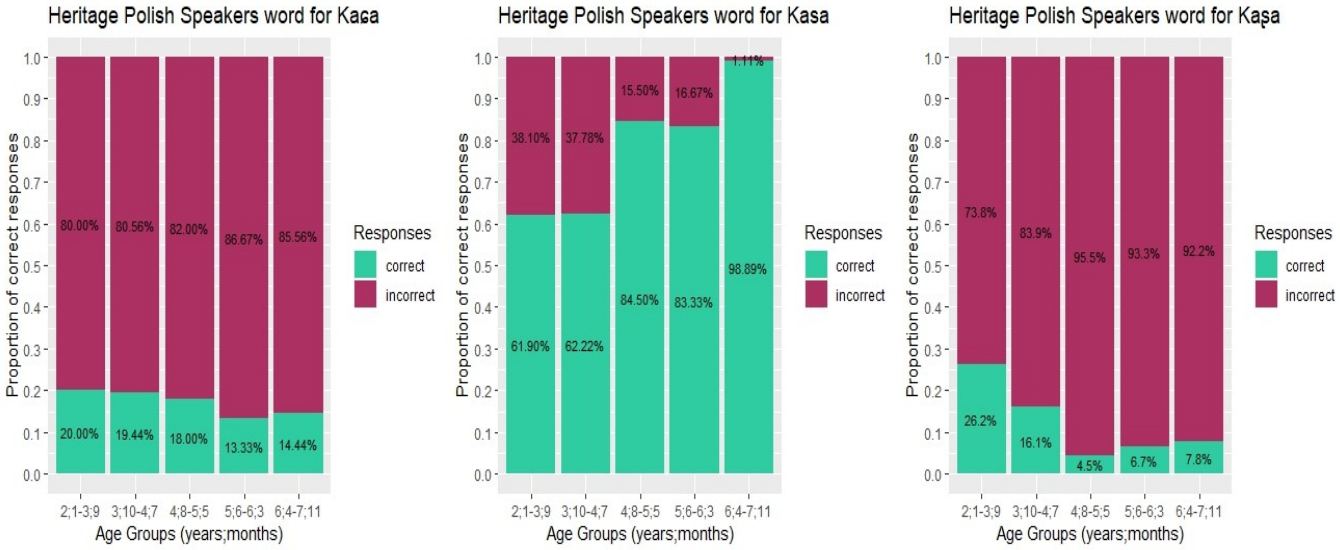

Figure 11: Polish Heritage Word Accuracy as a Function of Age Results. X-axis with the different age groups and y-axis with the accuracy response percentages. Response is color coded for each fricative as correct (bottom teal bar) or incorrect (top fuchsia bar). Each plot shows word perceptions for respective fricatives (/kaɕa/, /kasa/, and /kaʂa/)

4. Discussion

As outlined in the introduction, the goals of this study were to contribute to the existing knowledge of typical sibilant perceptions for both Heritage speakers of Polish and non-native populations by assessing the perceptual accuracy of Polish Heritage speakers for sibilants in varying phonetic environments. We also aimed to observe how native English speakers would perceive the foreign sibilant contrast. The discussion begins with an analysis of Polish Heritage speakers and the effect of

children's ages. Then we conclude the discussion with comments on the perceptual performance of the native English participants.

4.1. Polish Heritage Speakers' Perceptions and Children's Ages

In section 1.1, we cite Bukmaier and Harrington (2016) who found that the /ʂ/ sibilant was the last to be acquired among the three-way-contrasts. Zygis and colleagues (2023) confirmed this acquisition order, with /ɕ/ being acquired first, followed by /s/, and finally /ʂ/. We can observe this pattern in the results of our study (section 3.3, Figure 8). Across all stimulus conditions (isolated sibilant, VC pair, word), /kaʂa/ exhibited the lowest accuracy with less than 25% correct responses. This aligns with the child stimulus results from Zygis et al. (2023), where /kaʂa/ also had the lowest overall accurateness (Figure 12 from Zygis et al., 2023) with percentages ranging between 50-75%. However, our results showed improvement with the addition of phonetic information. From the isolated sibilant to the whole word (see Figure 8), there was a very small increase in accuracy for /kaʂa/. The sibilant variations of /kasa/ in our study showed an increase in accuracy, while /kaɕa/ showed a decrease with increased phonetic information. Therefore, the claim by Bukmaier and Harrington (2016) regarding the difficulty in /ʂ/ identification between contrasting sibilants being connected to the weak articulatory and acoustic information is not evident from our results. Furthermore, contrary to previous studies (cf. section 1.2) that concluded the addition of phonetic information, such as preceding vowel formants, did not lead to improved perception, our results demonstrated an improvement (with the exception of /ɕ/). It is interesting to note that the /ɕ/ sibilant, which was acquired first, did not have the highest accuracy in our study (Figure 8 accuracy ranged from 18-35%). Our results (Figure 8) showed that /kasa/ had the highest accuracy, with all responses well above 33.3% chance accuracy. In contrast, Zygis et al. (2023) found that the children's stimulus for /kaɕa/ had the highest accuracy (Figure 12 from Zygis et al., 2023), while /kasa/ had average correctness. This difference in accuracy may be a result of the interaction between English and

Polish phonology for our Polish Heritage participants. Since the English language lacks the other two sibilants (/ʂ/ and /ɕ/), the increased exposure to the common /s/ phoneme in both languages results in greater familiarity and confidence for all the /s/ conditions. This supports the notion that the lack of phonetic exposure leads to a decline in perception, thus confirming Polka's (1991) "fragile" contrast (cf. section 1.3). The Polish children in Zygis et al.'s study (2023) had less exposure to /s/ compared to the /ɕ/ sibilant and to those with English fluency. This is attributed to their social and linguistic environments in Poland, where obviously English is not as common as it is in Ontario, Canada. Native Polish child participants heavily relied on spectral properties, particularly COG, for distinguishing /s/ and /ʂ/ (Zygis et al., 2023). However, it was for the /ɕ/ sibilant that the formants, specifically the "F2 of preceding and following vowels," provided the most crucial cues for perception (Zygis et al., 2023). In our own study, we observed the opposite effect, where there was a greater reliance on formants for /s/ and /ʂ/, and an increased confidence in spectral cues for /ɕ/. In other words, the native Polish children in Zygis et al. (2023) performed well with spectral properties for /s/ and /ʂ/, while relying more on formants for /ɕ/. On the other hand, the Polish (adult) Heritage speakers in our study exhibited lower accuracy with spectral properties for /s/ and /ʂ/ and much less confidence in utilizing formants for /ɕ/ perception.

McLeod (2007) found that speech acquisition in typical children begins around the age of 3. During this stage, phonological processes and a prevalent sibilant lisp are common. By the age of 6, children's intelligibility improves as they outgrow the sibilant lisp (McLeod, 2007). The increase in intelligibility corresponds with our findings for /kasa/ (c.f section 3.4), where we observed that the youngest children's productions in Group 1 and 2 (under 4 years 8 months) were the most difficult to distinguish, resulting in the lowest accuracy across stimulus conditions for /kasa/. Group 3 (4 years 8 months to 5 years 5 months) and older children's productions exhibited significant improvements in precision especially with more phonetic information. However, we observed the opposite trend for /kaʂa/, where older

children's productions resulted in less accurate distinctions (refer to figures in section 3.4). This can be explained by the limited exposure our Polish Heritage speakers had to /ʂ/ and even less exposure to the specific word "kaʂa" (meaning "groats", which is not common in the Canadian English vocabulary)[3]. Based on Polka's (1991) concept of fragile contrast, since our Heritage speakers spent more time using English due to their sociolinguistic environment, their contrasts are "less amendable to training in adulthood". Additionally, younger children tend to speak more slowly, often drawing out each phoneme, due to their underdeveloped articulators compared to the older children (McLeod, 2007). This is especially true for phonemes that pose difficulties, such as the last acquired /ʂ/ sibilant contrast. The slight difference in stimulus exposure (for the participants) can explain the higher accuracy observed in the younger children's productions. This relates to Polka's (1991) concept of exposure in fragile contrasts, although the level of exposure in our study is relatively small compared to its intended definition. We can further narrow down the exposure time within each stimulus perception. Therefore, even if it is a matter of seconds between stimuli, the additional time and exposure may have contributed to an improvement in perception. No clear age pattern emerges for /kaɕa/ (see figures in section 3.4), as there is a slight decrease in accuracy with the addition of phonetic information (from an average sibilant accuracy of 34% to 17% at the word level). Once again, this aligns with how spectral properties provide the most reliable cues for Heritage speakers, similar to how English native speakers perceive the contrast between the two English sibilants, which is purely based on spectral information and not at all based on formant transitions. The lack of age patterns within each /kaɕa/ stimuli condition can be attributed to the lack of exposure in fragile contrasts, as described by Polka (1991,) or the process of child speech acquisition, as discussed by McLeod (2007). Therefore, considering that /ɕ/ is the first sibilant contrast to be acquired in Polish, native Polish children would not be struggling to articulate this

[3] Terms like "kaʂa" for groats is most likely used by Polish speakers that practice Polish cuisine as many dishes can be translated into "kaʂa" (ex: buckwheat, cornmeal, etc.). They may be familiar with this word if this is the case (Stroinska, personal communication, 2023).

phoneme compared to /ʂ/, due to the resulting reduced exposure. However, since our Polish Heritage speakers would have acquired /ɕ/ first, having more familiarity and exposure to it, their perceptions resulted in average accuracies for /kaɕa/ in our study. We can confirm that our hypothesis (refer to section 1.4) holds true for the /kasa/ manipulations. All accuracies for /kasa/ beyond 5 years and 6 months surpassed 33.3% chance. Unfortunately, the other two sibilants did not support our hypothesis, as their accuracies remained below 33.3% chance despite increasing ages.

4.2. English Speakers and a Comparison between all Groups

The analysis of the English participants is based solely on choice proportions and not accuracy, like in the Polish groups. Across stimuli conditions (refer to Figure 4), it appears that the addition of phonetic information (e.g., formants added to the spectral noise) did not influence the categorization of the sibilants. Closer inspection may show that for /kaɕa/ variations increased phonetic information led to a higher proportion of /ʃ/ perception. However, the effects are not strong enough to warrant that the addition of formant to spectral information influenced the English participants' perception. This confirms part of our hypothesis that /s/ variations would be categorized as /s/ and the other two Polish sibilants as /ʃ/. It is interesting to note that /kaɕa/ word variations had the highest proportion of /ʃ/ perceptions, despite the retroflex sibilant /kaʂa/ being closer in place of articulation to the post-alveolar English fricative / ʃ/ (see Figure 4). Perhaps part of this can be attributed to the fact that the alveolo-palatal fricative /ɕ/ is further back in articulation, making /ɕ/ perceptions more favored towards /ʃ/ rather than /s/. Contrary to our hypothesis, the retroflex /kaʂa/ produced mixed results, with approximately equal proportions of /s/ and /ʃ/ perceptions, hovering around 50%. Further addition of (following or preceding vowel) formants to spectral noise did not significantly affect the distinctions, as the average /s/ proportions remained steady at 45% for the /kaʂa/ *word* level (see section 3.2 figures). The proximity of the place of articulation to the two phoneme

options (/s/ and /ʃ/) to the retroflex fricative could explain why the results were split around 50%. Once again, just as how the alveolo-palatal fricative being further back than the alveolar /s/ led to increased /ʃ/ categorization, it could be that the retroflex, in comparison, was perceived to be centered around the alveolar-postalveolar region of the mouth. The English participants' categorical perception appears to be established for /kasa/ and /kaɕa/, but variable or uncertain for /kaʂa/. It may be that despite having the established /s/ vs /ʃ/ categories, English participants recognize that /kaʂa/ needs another category (or at least that it cannot be placed into just one of these already established ones). Padgett et al. (2010), cited in section 1.3, found that their native (English) participants did well in distinguishing /s/ from the unfamiliar postalveolars but had difficulty distinguishing *within* the unfamiliar postalveolars (i.e., /ɕ/ and /ʂ/). Our results show that the English participants did well in distinguishing that the alveolo-palatal /ɕ/ was different from /s/. However, they struggled with the retroflex /ʂ/, which may indicate a need for another category in the experimental design (which, of course, is methodologically difficult). Spectral noise and formant transitions, which were stated to be vital for Padgett and Zygis (2010), were decisive factors for the categorizations of /kasa/ and /kaɕa/ in our study, but not as much for /kaʂa/. Additional phonetic information, beyond the preceding formant, evidently influences mainly the /kaɕa/ perceptions. Altogether, the preceding vowel formant in addition to the spectral noise is evidently not a vital cue for the English participants' perception. Additional phonetic information aids perception to a certain extent but again the effect is very minimal.

Factoring in the ages of the children speakers showed increasing confidence in /s/ categorizations for /kasa/. The highest /s/ categorization is at the VC pair level, which again indicates more cues from preceding vowel formants rather than the following vowel formants and spectral noise. However, at the word level, /kasa/ saw a small drop in its average, where /s/ accounted for 86% of the total at the VC pair level and 83% at the whole word (see section 3.2). This aligns with Zharkova et al.'s (2018) study, which found increased discriminability with the /a/ phoneme in VC pairings.

The lowest /s/ categorization (thus highest /ʃ/ selection) is observed in /kaɕa/ and decreases with the older age groups. The same pattern is seen in /kaʂa/ with the older age groups. As the younger children exhibit a sibilant lisp along with lengthened articulations, the sibilants may sound more like /s/ to some extent. Nonetheless, it is evident that in Group 3 (beyond 5 years), with the disappearance of the lisp and more adult-like articulations, perceptions overwhelmingly lean toward /ʃ/. In section 3.2 (Figures 5 to 7), we can see a high /s/ distinction (more than 50% chance) for groups 1 and 2 for /kaʂa/. However, there is a clear shift in Group 3, as /s/ selections drop below chance (50%) across /kaʂa/ stimuli conditions. Thus, the more matured articulatory development of the children appears to increase sibilant distinctions in non-native English speakers.

4.3. Limitations of the study and future work

Due to the timing of our experimental study, which took place during the COVID-19 pandemic, social distancing measures and procedures limited our pool sample size for the Polish Heritage participant group. Students who could have participated were unable to do so because they were not present on campus (due to distance learning), and the experimental setup required in-person participation. The goal of the in-person participation was to limit discrepancies between different headphones and acoustic environments. As a result, our sample size was smaller than desired and too small to be considered conclusive for the Polish Learners. Additionally, the age range of the participants was narrow due to the experiment being conducted at McMaster University. It would have been interesting to include older Heritage speakers (above 25 years of age) and from outside of the Hamilton area in order to obtain a broader perspective. The experimental design itself was also a limitation, as it was created with the potential for conducting online experiments for the participants to complete the experiment in their own homes. This meant the tasks were simple but highly repetitive, which was reported by multiple participants. Future work could be designed to incorporate more focus tasks in-between to provide

breaks in the repetition and reduce the risk of participants making ingenuine selections. The restricted time frame also prevented us from analyzing the influence of participants' background languages and musical training. Although all the participants completed forms that included questions about their language background, exposure, and musical training, we were unable to examine the transfer effects on sibilant-contrast distinctions for both native and non-native speakers. This would have been an interesting avenue to explore in future research.

5. Conclusion

Our data showed that more phonetic information (formants of preceding and following vowels in addition to spectral noise) did not significantly influence English participants' categorical perception. The English participants had an established perception of /kasa/ and /kaɕa/ but were uncertain for /kaʂa/. However, increasing the age of the children in the stimuli, especially around the 5-year mark with the sibilant lisp disappearance, greatly influenced the perception to shift in favor of one sibilant (/s/ or /ʃ/). A change in stimulus conditions and increasing ages seem to have varied effects on certain sibilants for Polish participants. The addition of formants (to existing spectral noise) increased accuracy for the /s/ sibilant but decreased for /ɕ/ (minimal effects for /ʂ/). Increasing the age for the Polish Heritage speakers showed significant improvement in accuracy (for /s/) or its decline (for /ʂ/, no trend for /ɕ/). Putting it altogether, our perceptual results suggest that more formant information, in addition to the spectral noise of the sibilant, do not aid Polish-sibilant distinction in English listeners however increasing the child speaker's age does lead to improvements. Adult Heritage Polish listeners rely on different phonetic environments dependent on the sibilant, with age also shifting their perception. These results may indicate a broader phonetic process in foreign phoneme perception.

References

Boersma, Paul & Weenink, David (2022). Praat: doing phonetics by computer [Computer program]. Version 6.2.13, retrieved 05 June 2022 from http://www.praat.org/

Bukmaier, V., & Harrington, J. (2016). The articulatory and acoustic characteristics of Polish sibilants and their consequences for diachronic change. Journal of the International Phonetic Association, 46(3), 311–329. https://doi.org/10.1017/S0025100316000062

Cauldron Science. (2016). Gorilla: Easily create online behavioural experiments [website]. Gorilla Experiment Builder. https://gorilla.sc/

Dehaene-Lambertz, G., & Gliga, T. (2004). Common neural basis for phoneme processing in infants and adults. Journal of Cognitive Neuroscience, 16(8), 1375–1387. https://doi.org/10.1162/0898929042304714

Flege, J. E., & Davidian, R. D. (1984). Transfer and developmental processes in adult foreign language speech production. Applied Psycholinguistics, 5(4), 323–347. https://doi.org/10.1017/S014271640000521X

Goldstone, R. L., & Hendrickson, A. T. (2009). Categorical Perception. WIREs Cognitive Science.

Li, F., Munson, B., Edwards, J., Yoneyama, K., & Hall, K. (2011). Language specificity in the perception of voiceless sibilant fricatives in Japanese and English: Implications for cross-language differences in speech-sound development. The Journal of the Acoustical Society of America, 129(2), 999–1011. https://doi.org/10.1121/1.3518716

Li, F., & Munson, B. (2016). The Development of Voiceless Sibilant Fricatives in Putonghua-Speaking Children. Journal of Speech, Language, and Hearing Research, 59(4), 699–712. https://doi.org/10.1044/2016_JSLHR-S-14-0142

McLeod, S. (2007). Children's Speech Acquisition. In The International Guide to Speech Acquisition (pp. 175–221). essay, Thomson Delmar Learning.

Minczakiewicz, E., (2017). Dyslalia against other speech defects and disorders in preschool and school-age children, (PL) Dyslalia na tle innych wad i zaburzeń mowy u dzieci w wieku przedszkolnym i szkolnym. Konteksty Pedagogiczne 1, 149–169. 10.19265/KP.2017.018149.

Miodonska, Z., Badura, P., & Mocko, N. (2022). Noise-based acoustic features of Polish retroflex fricatives in children with normal pronunciation and speech disorder. Journal of Phonetics, 92, 101149. https://doi.org/10.1016/j.wocn.2022.101149

Polka, L. (1991). Cross-Language Speech Perception in Adults: Phonemic, Phonetic, and Acoustic Contributions. The Journal of the Acoustical Society of America, 89(6), 2961–2977. https://doi.org/10.1121/1.400734

Rstudio Team (2023). Rstudio: Integrated Development for R [computer program]. Boston, MA. Version 2023.03.0+386: RStudio Inc.

Santrock, J.W. (2021). Child development, 15th international student Ed.

Speechlanguage-resources. (2013). Lateral lisp therapy. Lateral Lisp Therapy. http://www.speechlanguage-resources.com/lateral-lisp-therapy.html

Stroinska, M. (2023). [Email correspondance]

Tamburelli, M., Sanoudaki, E., Jones, G., & Sowinska, M. (2015). Acceleration in the bilingual acquisition of phonological structure: Evidence from Polish–English bilingual children *. Bilingualism: Language and Cognition, 18(4), 713–725. https://doi.org/10.1017/S1366728914000716

Tamminen, H., Peltola, M. S., Kujala, T., & Näätänen, R. (2015). Phonetic training and non-native speech perception — New memory traces evolve in just three days as indexed by the mismatch negativity (MMN) and behavioural measures. International Journal of Psychophysiology, 97(1), 23–29. https://doi.org/10.1016/j.ijpsycho.2015.04.020

Wadhera, T. & Kakkar, D. (Eds.). (2020). Interdisciplinary Approaches to Altering Neurodevelopmental Disorders. IGI Global. https://doi.org/10.4018/978-1-7998-3069-6

Viriaro. (2023, April 12). *Adding a new column based on information in existing column* [Online forum post]. Reddit. https://www.reddit.com/r/RStudio/comments/12jz5y9/comment/jg8yevm/

Zharkova, N., Hardcastle, W. J., & Gibbon, F. E. (2018). The dynamics of voiceless sibilant fricative production in children between 7 and 13 years old: An ultrasound and acoustic study. The Journal of the Acoustical Society of America, 144(3), 1454–1466. https://doi.org/10.1121/1.5053585

Żygis, M., & Padgett, J. (2010). A perceptual study of Polish fricatives, and its implications for historical sound change. Journal of Phonetics, 38(2), 207–226. https://doi.org/10.1016/j.wocn.2009.10.003

Żygis, M., Pape, D., Jaskuła, M, & Koenig, L. (2023) "Do children better understand adults or themselves? An acoustic and perceptual study of the complex sibilant system of Polish." Journal of Phonetics.

Appendix

Table 1
Children's age groupings used in the graphs.

	Age (year; month)	Number of children	Number of observations in perception blocks
Group 1	2;11–3;9	21 (9 female, 12 male)	903
Group 2	3;10–4;7	17 (9 female, 8 male)	785
Group 3	4;8–5;5	19 (8 female, 11 male)	884
Group 4	5;6–6;3	12 (7 female, 5 male)	534
Group 5	6;4–7;11	7 (4 female, 3 male)	318

Table 1. From Zygis et al (2023) of Children's Age Groupings respective to their experiment

Figure 1: Picture illustrating the two-way forced selection buttons and the labels for the English only participants.

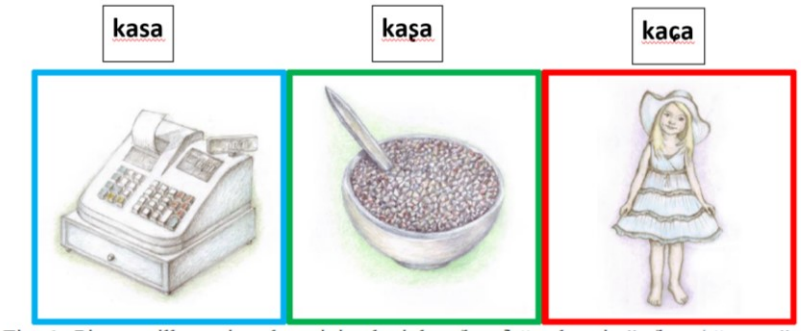

Figure 2: Replica picture, from Zygis et al., (2023), illustrating the minimal triplet: /kasa/ "cash point, /kaʂa/ "groats", and /kaca/ "Cathe, prop.name".

Figure 3: Picture of the Testing Setup for all participants

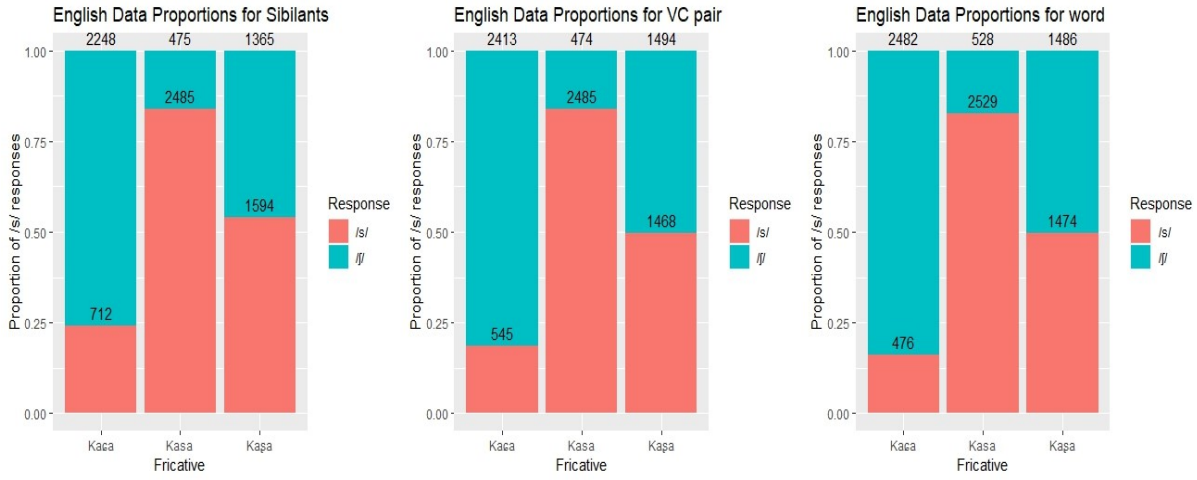

Figure 4: English Perceptual Results. X-axis with the different fricatives (/kaca/, /kasa/, and /kaşa/) and y-axis with the proportion of /s/ responses with the case count labels (i.e., how many items occurred per plotted condition). Response is color coded for each fricative as /s/ or /ʃ/.

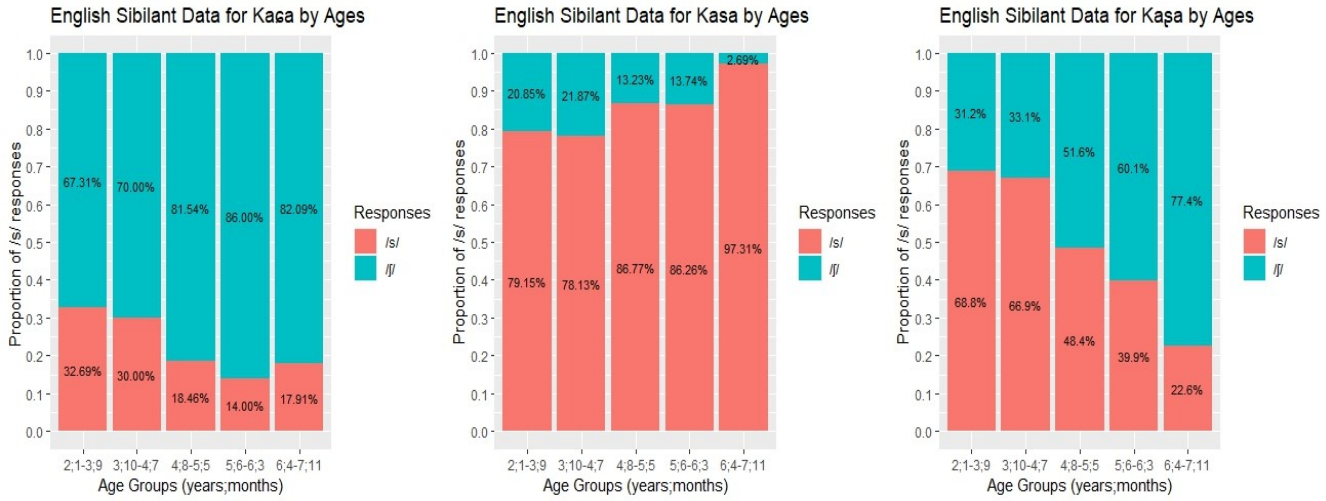

Figure 5: English isolated sibilant Perceptual Responses as a Function of Age Results. X-axis with the different age Groups and y-axis with the response proportions. Response is color coded for each fricative as /s/ or /ʃ/. Each plot shows sibilant perceptions for respective fricatives (/kaɕa/, /kasa/, and /kaʂa/).

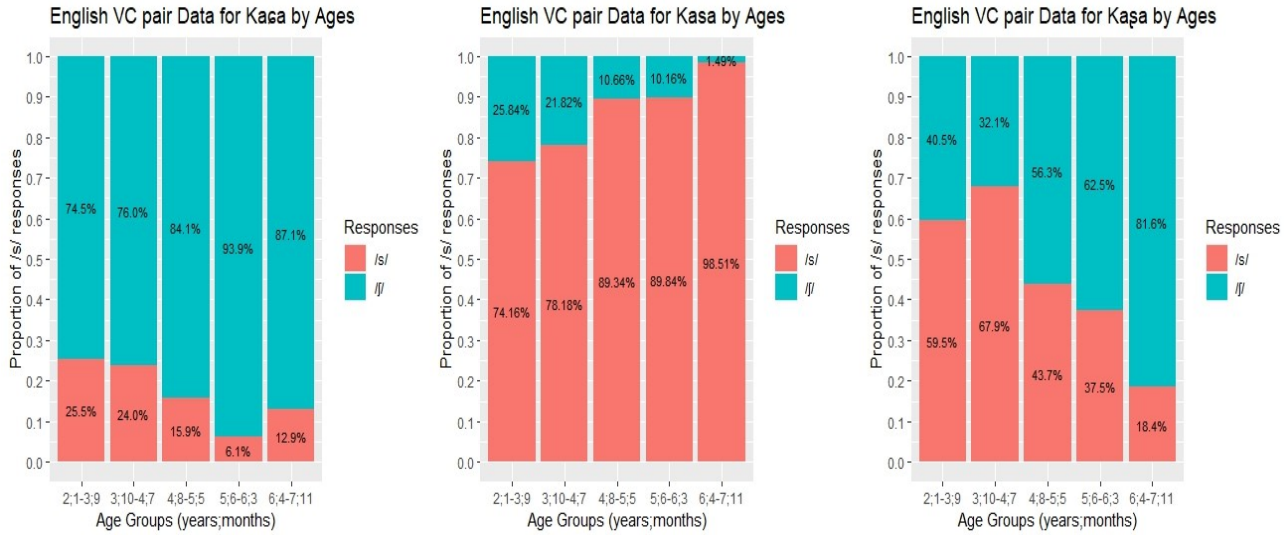

Figure 6: English VC pair Perceptual Proportions as a Function of Age Results. X-axis with the different age Groups and y-axis with the response proportions. Response is color coded for each fricative as /s/ or /ʃ/. Each plot shows VC pair perceptions for respective fricatives (/kaɕa/, /kasa/, and /kaʂa/)

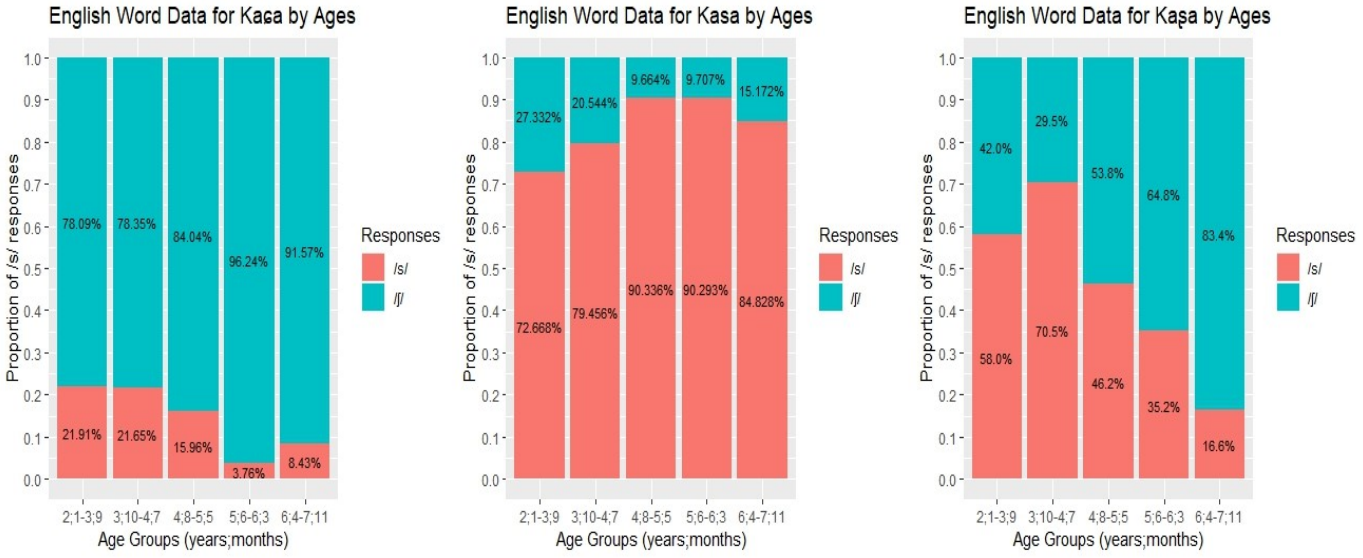

Figure 7: English Word Perceptual Proportions as a Function of Age Results. X-axis with the different age Groups and y-axis with the response proportions. Response is color coded for each fricative as /s/ or /ʃ/. Each plot shows word perceptions for respective fricatives (/kaɕa/, /kasa/, and /kaʂa/)

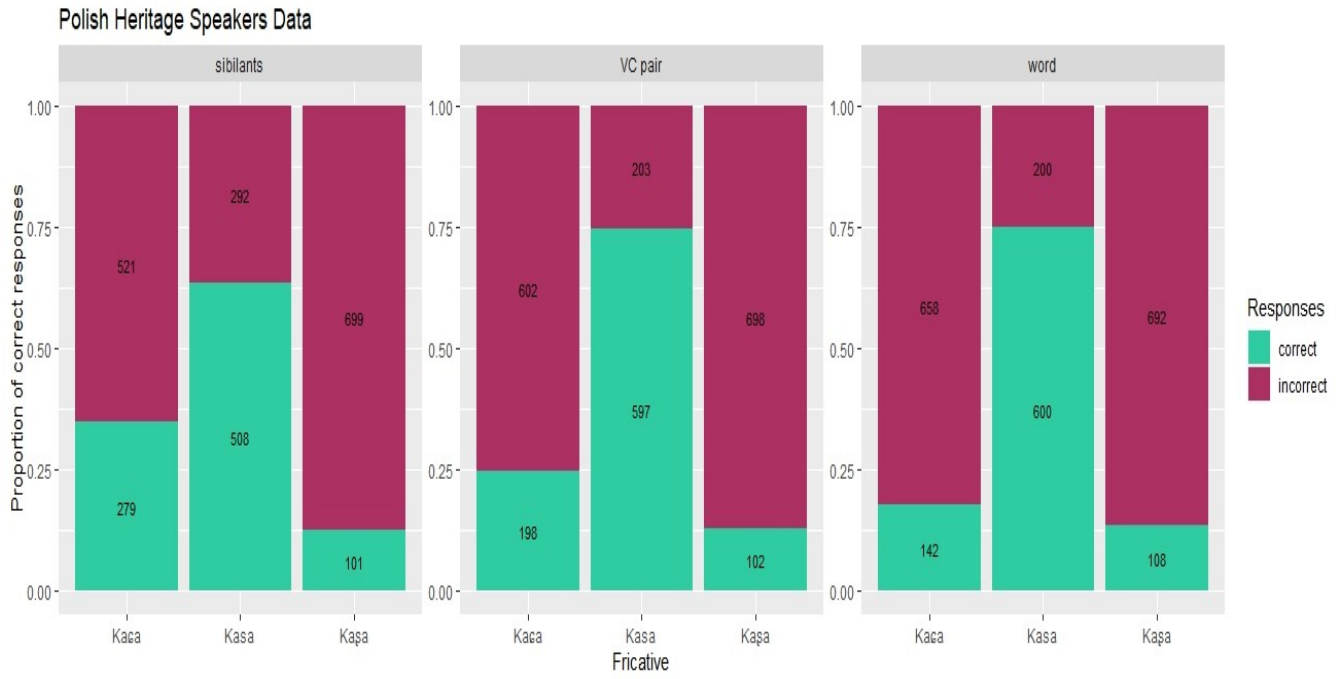

Figure 8: Polish Heritage Speakers Accuracy Results. X-axis with the different fricatives (/kaca/, /kasa/, and /kaşa/) and y-axis with the accuracy in percentage scales. Response is color coded for each fricative as correct (bottom teal bar) or incorrect (top

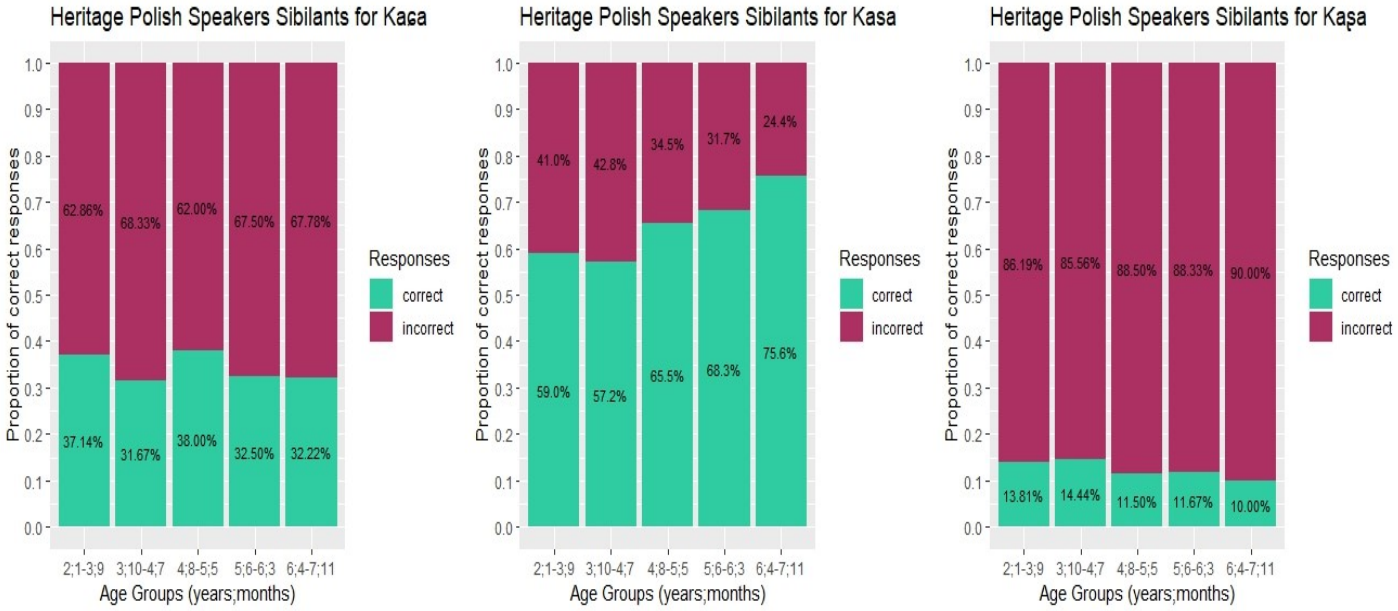

Figure 9: Polish Heritage Sibilant Accuracy as a Function of Age Results. X-axis with the different age groups and y-axis with the accuracy response percentages. Response is color coded for each fricative as correct (bottom teal bar) or incorrect (top fuchsia bar). Each plot shows sibilant perceptions for respective fricatives (/kaca/, /kasa/, and /kaşa/)

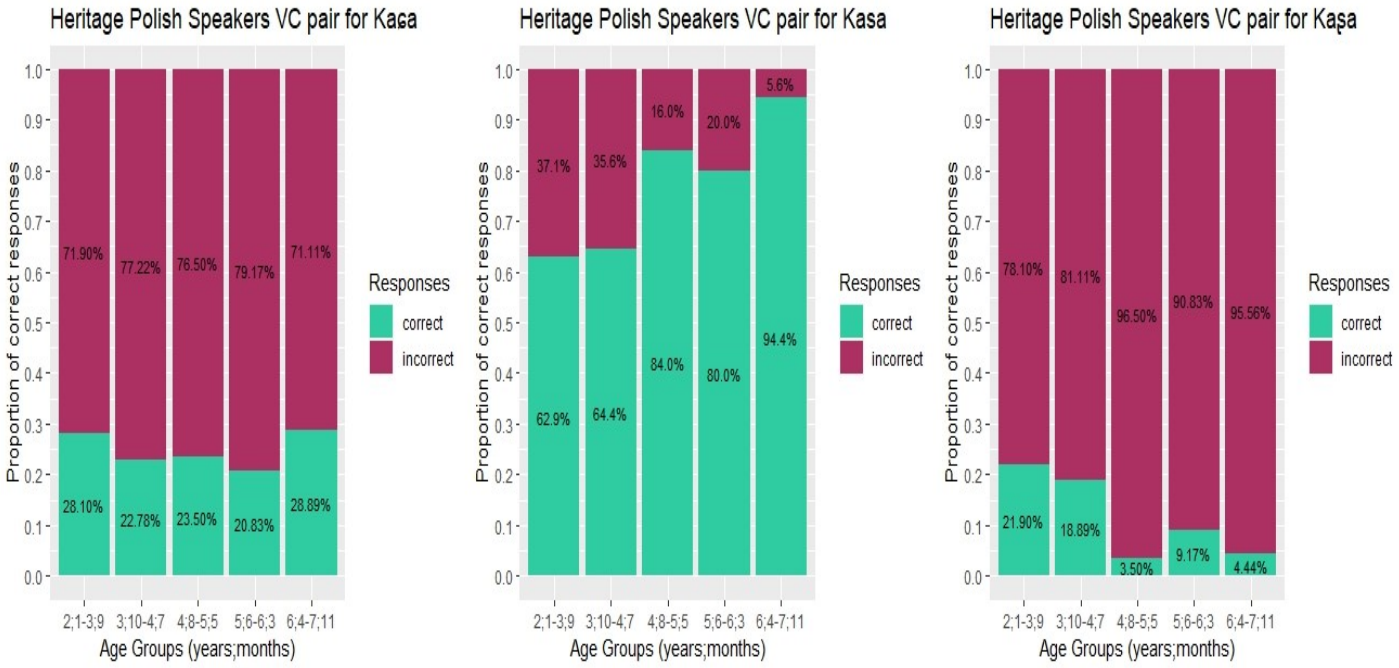

Figure 10: Polish Heritage VC Pair Accuracy as a Function of Age Results. X-axis with the different age groups and y-axis with the accuracy response percentages. Response is color coded for each fricative as correct (bottom teal bar) or incorrect (top fuchsia bar). Each plot shows VC pair perceptions for respective fricatives (/kaɕa/, /kasa/, and /kaṣa/)

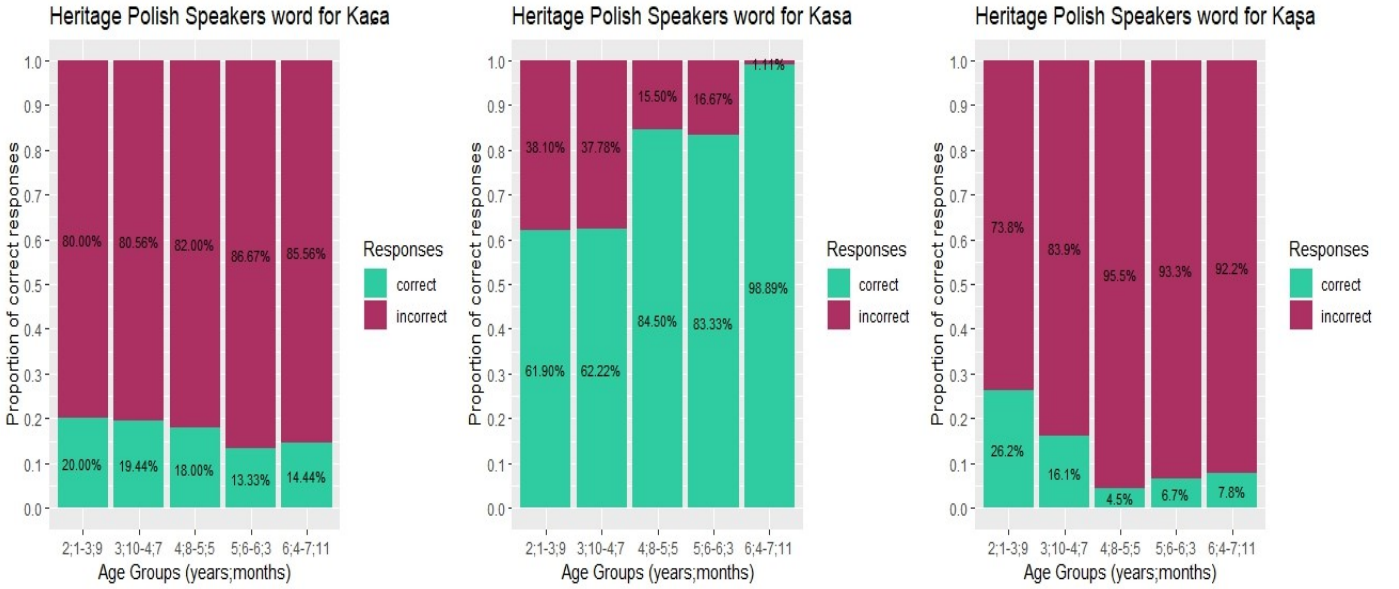

Figure 11: Polish Heritage Word Accuracy as a Function of Age Results. X-axis with the different age groups and y-axis with the accuracy response percentages. Response is color coded for each fricative as correct (bottom teal bar) or incorrect (top fuchsia bar). Each plot shows word perceptions for respective fricatives (/kaɕa/, /kasa/, and /kaṣa/)

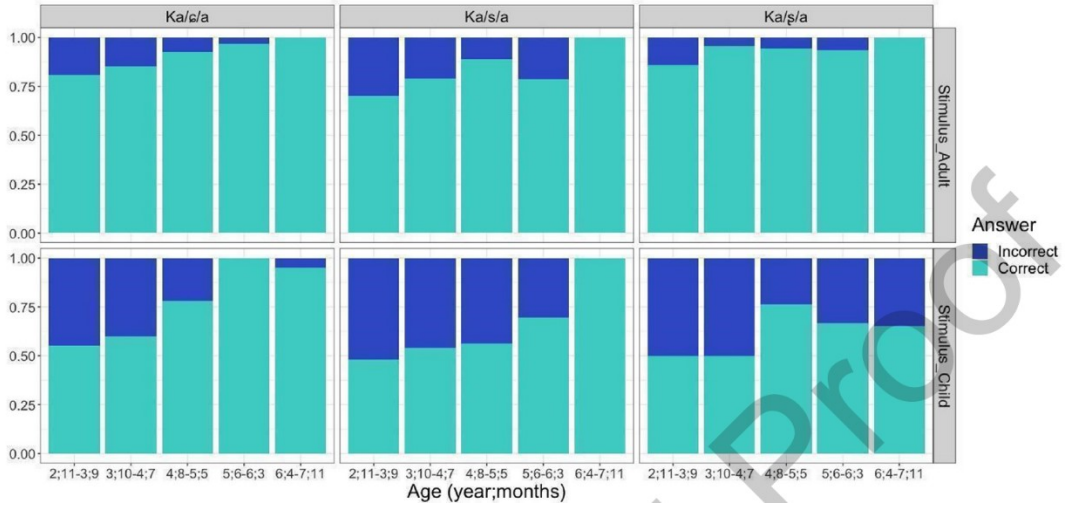

Figure 12: (From Zygis et al., 2023) Proportion of correct/incorrect answers as a function of the STIMULUS TYPE (Adult, child) and AGE for the word-medial contrast