

The Development of Perceptual Sensitivity to Polish Sibilants at First Exposure

Ellenor Shoemaker

*Université Sorbonne Nouvelle – Paris 3
Laboratoire de Phonétique et Phonologie (CNRS-UMR 7018)*

1 Introduction

1.1 *The VILLA project* The study of language learning in adults at first exposure is a burgeoning field within the domain of language acquisition and has increasingly received more attention over the last 10 years. The current data were collected as part of the pan-European project VILLA¹ (*Varieties of Initial Learners in Language Acquisition: controlled classroom input and elementary forms of linguistic instruction*), during which participants in five countries (France, Germany, England, the Netherlands, and Italy) participated in a two-week Polish course. Polish was chosen as the target language for this project both because its structural and phonological properties differ greatly from the five first languages (L1s) of the projects' participants, and because Polish is a language that is not frequently taught in the participants' home countries and therefore previous exposure is rare. Learner performance was assessed throughout the course in several different areas in an attempt to investigate what resources learners have access to and what learners are capable of in the very first hours of exposure to a second language (L2). All participants attended nine 90-minute sessions over 10 days (the final day of the course was devoted solely to testing) and received a total of 14 hours of monolingual native speaker input from the same instructor. Empirical investigations explored the development of phonology, the lexicon, and morpho-syntax. Both audio and video recordings were made of the course.

In each participating country, two different participant groups were exposed to different types of input. One group received meaning-based (MB) input, which contained no meta-language or explicit linguistic explanation. The activities in class were communicative in nature and focused on message rather than form. Learners in the MB group received less explicit correction (some correction of learner output was unavoidable in teacher repetitions). The other group was exposed to form-based (FB) input, which included explicit information about structural regularities in the target language (highlighting, tables of grammatical forms). In addition there was extensive corrective feedback from the instructor in the FB group. In neither input group was there any explicit phonological instruction or feedback specifically targeted toward pronunciation.

The data reported here specifically concern the performance of French-speaking participants on the perception of Polish sibilant fricatives at three intervals throughout the 10-day course.

1.2 *Sibilant inventories of Polish and French* Among the world's languages, Polish has a large inventory of fricative sounds and therefore provides an excellent testing ground for the acquisition of non-native phonemic contrasts. According to the UCLA Phonological Segment Inventory Database (UPSID), the most common sibilant contrast across languages is the two-way place distinction between alveolar /s, z/ and palato-alveolar /ʃ, ʒ/. French is an example of a language with this contrast. Polish on the other hand is

* This research was funded by an Open Research Area in Europe for the Social Sciences grant (ANR, DFG, NWO) from France, Germany, and the Netherlands for the period 2011-2014. We would also like to thank Ewa Lenart for help and comments concerning the creation and recording of the stimuli used in this study.

¹ The reader is referred to Dimroth, Rast, Starren, and Watorek (2013) for an in-depth description of the entire VILLA project in all participating countries.

extremely rare in that it contrasts three different places of articulation in sibilants: alveolar² (/s, z/), alveopalatal (/ɕ, ʑ/) and retroflex (/ʂ, ʐ/). It is important to note, however, that the retroflex status of certain Polish sibilants is not globally accepted. Many researchers transcribe the retroflex /ʂ, ʐ/ as palato-alveolar /ʃ, ʒ/. Articulatory data, however, point to a retroflex status for these sibilants. Rochon and Pompino-Marschall (1999) compared productions of /ʂ/ with productions of /s/ and /t/ and showed through eletropalatography (EPG) and articulography (EMA) that /ʂ/ is produced with the tongue tip bent backwards, therefore giving evidence for its retroflex status. Table 1 presents sibilant inventories of Polish and French.

Polish		French	
ALVEOLAR	/s, z/	ALVEOLAR	/s, z/
ALVEOPALATAL	/ɕ, ʑ/	PALATO-ALVEOLAR	/ʃ, ʒ/
RETROFLEX	/ʂ, ʐ/		

Table 1: *Sibilant inventories of Polish and French.*

1.3 Second Language Phonological Acquisition An extensive body of research has shown that adult learners can have difficulty discriminating sounds that are not contrastive in the native language. As noted by Kuhl “no speaker of any language perceives acoustic reality; perception is altered in the service of language” (2000: 11852). In other words, it is a listener’s linguistic experience that determines which phonemic contrasts our perceptual system pays attention to and not the acoustic properties of sounds. While infants are born with acute sensitivity to speech sounds, by the end of the first year of life the ability to perceive segmental contrasts diminishes as the infant perceptual system abandons or prunes away phonetic features that are not contrastive in the native language (Werker & Tees, 1984). As the infant perceptual system becomes attuned to the phonological properties of the L1, infants become in effect ‘adult listeners’ as sensitivity to abandoned contrasts is greatly diminished. This decline in perceptual sensitivity is one factor that has been invoked to account for the striking difference between the facility of L1 acquisition and the difficulties that adult learners experience in acquiring a L2. When first confronted with a novel language, the adult listener may initially have the impression of listening to a stream of noise. Little by little however as the listener is further exposed to the new language, individual sounds and words slowly begin to emerge as discrete and recognizable units from the continuous acoustic stream. The very beginning stages of this process are the focus of the current study.

Drawing on the fact that the world’s languages compose their sound systems from a universally shared set of possible articulatory gestures, Best (1995; see also Best, McRoberts & Goodell 2001) observes that there is inevitably a great deal of overlap in the production of segments across languages. The *Perceptual Assimilation Model* (PAM) is centered on the notion that the degree of similarity between the L1 and L2 sound systems determines the degree of difficulty that L2 learners will experience in L2 perception. PAM focuses on the perception of non-native *segmental contrasts*, offering predictions as to where successful discrimination of L2 contrasts is expected to be achieved. Essential to this idea is the prediction that learners will have problems detecting differences between native and non-native sounds if non-native segments are perceived as similar to pre-existing native categories. In these cases, when confronted with non-native sounds, the learner will assimilate the new L2 sound into what is perceived as the closest L1 category instead of creating a novel and distinct category.

Best makes a distinction among four types of assimilation of L2 contrasts. First, *Two-Category Assimilation* describes cases in which a non-native contrast is perceived as similar to two separate native categories. This contrast is thus assimilated into two different native categories, with one L2 sound being assimilated to one native category and the other being assimilated to a different native category, neither of which is a good fit. An example of this type of assimilation occurs in speakers of French learning the English dental fricative pair /θ/ and /ð/. The unvoiced member of this pair is perceived as /s/ while the voiced member is perceived as /z/, both of which are existing members of the French phoneme inventory.

² Also seen in the literature as dental (Nowak, 2006).

As a result, neither of the two non-native sounds is correctly classified, but good discrimination is nevertheless achieved.

The second type of assimilation, *Single-Category Assimilation*, refers to cases in which two non-native categories are assimilated into a single existing native category. The native category into which the two non-native sounds are assimilated may constitute a good match for neither of the L2 sounds. Discrimination of the two sounds is assumed to be difficult as both tend to be (poorly) mapped onto one existing category. Japanese learners of English, for example, have well documented difficulty with the English /l-r/ distinction as both tend to be mis-mapped onto the Japanese alveolar flap /ɾ/.

The third type of assimilation described by Best is based on *Category Goodness*, in which two non-native categories are also assimilated into one native category, but unlike *Single-Category* assimilation, one of the non-native sounds represents a good exemplar of a pre-existing native category. For example, English-speaking learners of French often assimilate both French vowels /y/ and /u/ into the pre-existing English category /u:/. In this type of assimilation, discrimination of the two non-native categories is not achieved.

Finally, Best describes non-native speech sounds that may fall outside the phonetic space of the native phonology. These sounds are not assimilated as the non-native contrasts are sufficiently different from native categories to be perceived as such and the learner establishes new categories for each. Discrimination of these sounds is assumed to be strong. This model accurately predicts that Zulu clicks, for example, will be easily perceived by speakers of European languages (Best, McRoberts & Sithole 1988) as there is no close equivalent.

The *Speech Learning Model* (SLM) put forth by Flege (1995) also offers predictions on long-term phonetic development. Like Best, Flege offers multiple classifications of L2 sounds as they relate to those of the L1. He proposes that the learner classifies an L2 sound as *identical to*, *different from* or *similar to* the sounds of the L1. Those sounds that are classified as identical are thought to pose no problem for the learner as they are subsumed into already established phonological categories. L2 sounds that are sufficiently dissimilar from those of the L1 are also thought to be learned with relative ease as novel discrete categories can be established. Sounds that are perceived as similar to L1 sounds, however, pose the biggest obstacle for the learner as these phonemes are perceived as belonging to neither new nor separate phonological categories and thus are incorrectly assimilated into an existing (but different) category.

While both Best and Flege have proposed models that explicate the perceptual classification of native and non-native contrasts during the beginning stages of development, far less focus has been placed on the general time course of adult phonological acquisition or acquisition at first exposure. To date, the use of artificial languages has been the sole avenue for the study of adult learners' sensitivity to novel acoustic input. In a seminal study, Saffran, Newport, and Aslin (1996) showed that English-speaking adults are sensitive to transitional probabilities and prosodic cues in artificial language after just 21 minutes of exposure to novel input and that they can use this information to locate word boundaries. Peña, Bonatti, Nespor, and Mehler (2002) further showed that French speakers are sensitive to transitional probabilities after just 10 minutes of exposure. To our knowledge no first exposure studies to date have dealt with the very first stages of the adult phonological acquisition in a real language. The current study tackles this issue by examining the perception of non-native phonemic contrasts at first exposure to Polish.

1.4 The perception of Polish sibilants The discrimination of Polish sibilants by non-native speakers has previously been the focus of a number of studies, all of which have tested native speakers of English. Lisker (2001) tested 12 native speakers of American English. Participants were first trained via feedback to identify the CV non-word stimuli as containing either /s/, /ʂ/, or /ɕ/ and the vowel /a/. Three types of stimuli were used: full CV syllables, only the frication portion of the syllables and only the vowel portion (including formant transitions). Lisker's results showed that speakers of English could not reliably identify /ʂ/, or /ɕ/ and frequently confused the two in a labeling task, while identification of /s/ was consistently at ceiling. This result was predicted as the English phoneme inventory has only a two-way place distinction between /s/ and /ʃ/ and therefore, while /s/ was identified without difficulty, the unfamiliar phonemes /ʂ/ and /ɕ/ were both assimilated to /ʃ/. Interestingly, participants were better able to identify both of the excised portions of the syllables than the full CV sequences. Lisker proposes that in the excised portions participants were perceiving the stimuli in a purely auditory (non-speech) mode focusing primarily on the acoustic properties of the sounds, while the CV syllables, which more closely resemble speech, were perceived in speech mode and therefore were mapped onto existing categories in the listeners' native phonologies.

McGuire (2007) investigated the weighting of different acoustic cues and explored the effects of training in the perception of alveopalatal and retroflex voiceless Polish sibilants by native speakers of English. In this pilot study, McGuire showed that though English speakers initially identify both /ɕ/ and /ʂ/ as English /ʃ/, they can be trained to distinguish the alveopalatal and retroflex sibilants reliably. Furthermore, he found that English speakers attend more to vocalic information than frication, a finding that differs from work examining native speakers. Nowak (2006) showed that native speakers of Polish rely more heavily on frication in the discrimination of sibilants than on formant transitions. Native speakers can reliably distinguish the three sibilant categories based on frication alone.

Employing an AX discrimination task in an investigation as to whether diachronic sound change can be perceptually motivated, Zygis and Padgett (2010) tested 10 native speakers of English and 13 native speakers of Polish on their discrimination of non-word CV and VC syllables incorporating four Polish sibilants /s, ɕ, ɛ, ʂ/ and the vowel /a/. Not surprisingly, the English-speaking subjects performed significantly worse than the Polish participants. Also as expected, the English speakers were better in discriminating /s/ syllables from the non-native sibilants (/ɕ, ɛ, ʂ/), than they did discriminating among the non-native sounds.

The above work has established that the perception of Polish sibilants by non-native speakers of the language is challenging and furthermore that, as predicted by Best, unvoiced alveopalatal and retroflex Polish sibilants are mapped onto existing native phonemic categories, namely palato-alveolar /ʃ/. The work of Lisker (2001) and McGuire (2007) has further suggested that non-native listeners can improve in the perception of these sibilants with minimal but targeted training.

The perception of Polish sibilants by native speakers of French has not previously been investigated, however given that French shares a similar sibilant inventory with English, we predict that, at least initially, learners will have difficulty distinguishing alveopalatal /ɕ, z/ from retroflex /ʂ, ʐ/, and both will be perceived as French palatoalveolar /ʃ, ʒ/, respectively. To put this specifically in terms of Best's Perceptual Assimilation Model, we predict a case of *Single-Category Assimilation* in that the alveopalatal and retroflex contrast will (initially) both be assimilated to the French palato-alveolar sibilants, with the unvoiced /ɕ/ and /ʂ/ both being assimilated to unvoiced /ʃ/ and voiced /z/ and /ʐ/ both being assimilated to voiced /ʒ/. Note as well that the current study test both the voiced and unvoiced sibilants, while previous work has only focused on the voiceless set.

The current study did not incorporate targeted phonetic training as seen in Lisker (2001) and McGuire (2007), however, participants were exposed to native Polish input for a total of 14 hours in a classroom setting. Using artificial languages, Saffran *et al.* (1996) and Peña *et al.* (2002) have shown that the perceptual system can track statistical distributions in novel acoustic input from very early on. We aim to extend this line of inquiry by investigating whether learners' sensitivity to non-native phonemic contrasts can improve without the benefit of training in the first hours of exposure to a second language⁴ with input which includes no explicit phonological instruction or training. Specifically, by investigating the perception of non-native phonemes at first exposure, we seek to understand how and when learners begin to form novel phonemic categories at the very beginning of exposure to a L2.

2 Methods

2.1 Participants Thirty-six native speakers of French took part in a 10-day Polish course in Paris, France and received a total of 14 hours of oral Polish input from a native-speaking instructor. None of the participants had previously been exposed to Polish or another Slavic language. The participants were divided in two input groups: MB input (17 participants; 4 male; aged 19-27 years) and FB input (19 participants; 7 male; aged 19-27 years). A control group of 15 monolingual⁵ native speakers of Polish was also tested in Poland (aged 20-25 years).

2.2 Stimuli The stimuli consisted of pairs of CV non-words including six sibilants from the Polish

³ The sound /ʂ/ is an allophone of /ɕ/ when /ɕ/ occurs before /i/ or /j/.

⁴ The current participants all had other L2s, namely English, and in most cases another Romance language, so strictly speaking Polish represented a L3 or L4, however, we use the term 'second language' here to mean any language learned after the acquisition of the L1 is complete.

⁵ Though the Polish control participants all studied second languages to varying degrees, they were all raised in monolingual households and primarily spoke Polish on a daily basis.

phonemic inventory followed by the vowel /a/: /sa/, /za/, /ea/, /za/, /ʒa/, /z̥a/ (see Table 2). Following Zygis and Padgett (2010), the vowel /a/ was chosen because it conforms to Polish phonotactic patterns and can appear after all of the sibilants employed in our study without eliciting allophonic modification (/ʒ/, for example, is impossible before /i/ as it would entail palatization). Six tokens of each non-word were recorded by a female native speaker of Polish in a sound-attenuated booth at a sampling rate of 44.1 kHz. Tokens were normalized for peak intensity using Praat (Boersma & Weenink 2007). Pairs of the CV non-words were subsequently created in all possible combinations and in both possible orders (e.g. /sa/-/za/ as well as /za/-/sa/) resulting in 30 different pairs (15 combinations in both presentation orders) and 6 same pairs. Same pairs always consisted of two different tokens of the same syllable (e.g. /sa₁/ - /sa₂/). Interstimulus interval (ISI) for each pair was 250 msec.

	ALVEOLAR	ALVEOPALATAL	RETROFLEX
unvoiced	/sa/ <i>sa</i>	/ea/ <i>sia</i>	/ʒa/ <i>sza</i>
voiced	/za/ <i>za</i>	/za/ <i>zia</i>	/z̥a/ <i>rza</i>

Table 2: Non-word stimuli in phonetic transcription and corresponding Polish orthography (in italics).

2.3 Procedure The French-speaking participants were tested on an AX discrimination task at three time intervals throughout the Polish language course (T1= 0h of exposure, i.e. prior to the course; T2= 4.5h of exposure; T3= 10.5h of exposure). The control group of 15 native Polish speakers was tested one time only. The experimental protocol was created using E-Prime experimental software (Schneider *et al.*, 2002) and was presented on either laptop or desktop computers. Stimuli were presented binaurally through headphones. Instructions were presented on the computer screen in the first language of the participants. In each experimental trial, participants heard a pair of non-words and were asked to report whether they heard two instances of the same word or two different words by pressing either (1) or (2), respectively, on the computer keyboard. Participants first completed a training portion of nine trials in order to familiarize them with the procedure before beginning the experimental portion. Items included in the training portion were not included in the experimental portion. The experimental portion consisted of 240 trials, of which 180 were test trials and 60 were distractor trials. The 180 test trials consisted of 90 different pairs (30 pairs presented 3 times each) and 90 same pairs (30 pairs presented 3 times each). The 60 distractor trials (30 same pairs/30 different pairs) included the nonce syllables /ga/, /ka/ and /pa/, which all contain phonemes that exist in both Polish and French and are phonotactically licit. Stimuli were presented in randomized order. No feedback as to the accuracy of responses was given in either the training or the experimental portion of the test. No response limit was set; participants were instructed to respond quickly, though not so quickly as to sacrifice accuracy. Each testing session lasted approximately 12 minutes.

2.4 Results We present analyses here for only the 30 different pairs. We have also collapsed both presentation orders for analysis, i.e. responses for /sa/-/za/ and /za/-/sa/ are analyzed together. For this reason, all analyses that follow include 15 different pairs.

The Polish control group was tested on the AX discrimination task one time only. One control participant was excluded from analysis because s/he had a mean accuracy rate across all pairs of 52%, suggesting that responses were given at random. Results for the control group thus include 14 participants. Mean accuracy (i.e., correct ‘different’ responses) on all 15 different pairs was 97.1% (SD 0.10). A one-way Analysis of Variance (ANOVA) showed a significant difference in accuracy rates among the 15 pairs: $F(14, 195) = 7.12, p < .0001$. Post-hoc analyses (Tukey) showed that the alveopalatal-retroflex pairs (P5, the voiced pair, and P10, the unvoiced pair) were discriminated significantly worse (78.6% and 87.0%, respectively) than the other 13 different pairs. (See Table 3 for a complete list of pairs.)

We also calculated d' (d-prime) for the control group. The sensitivity index d' is used in signal detection theory (see for example Macmillan & Creelman, 1991) and is calculated by transforming participant hit rates (H; correct ‘different’ responses) and false alarm rates (F; ‘different’ response to ‘same’ pair) to z scores and then calculating the difference: $d' = z(H) - z(F)$. Because d' takes into account not only accuracy (as measured by hits), but also any possible response bias (as measured by false alarms), it is

considered to be a more sensitive measure of discrimination than accuracy alone. Good discrimination is generally thought to be reflected by a d' score of 3 or above, while a score of below 1 is taken to be indicative of very low (or a lack of) discrimination. We should also note that in calculating d' neither H nor F can be 0 or 1. This is a particular problem in the current data because many of the different pairs were discriminated at 100% accuracy (a hit rate of 1). Hit rates of 1 were therefore converted to 0.999 and false alarm rates of 0 were converted to 0.001.

A significant difference among the 15 different pairs was observed for mean d' scores for the control group: $F(14, 195) = 4.07, p < .0001$. Post-hoc analyses (Tukey) showed that the alveopalatal-retroflex pairs (P5 and P10) were discriminated significantly worse than the other 13 different pairs. Mean accuracy and mean d' scores along with standard deviations for the Polish control group are presented in Table 3.

Pair	Accuracy		D'	
	Mean	<i>SD</i>	Mean	<i>SD</i>
P1: /zɑ/-/sa/	100	0	5.25	1.18
P2: /zɑ/-/ea/	100	0	5.25	1.18
P3: /zɑ/-/ʂɑ/	98.8	4.4	5.10	1.39
P4: /zɑ/-/za/	98.8	4.4	5.10	1.20
P5: /zɑ/-/za/	78.6	26.5	3.69	1.95
P6: /sa/-/ea/	98.8	4.4	5.63	0.90
P7: /sa/-/ʂɑ/	98.8	4.4	5.63	0.90
P8: /sa/-/za/	100	0	5.78	0.79
P9: /sa/-/za/	100	0	5.78	0.79
P10: /ea/-/ʂɑ/	87.0	14.9	3.51	1.79
P11: /ea/-/za/	97.6	6.0	5.19	1.37
P12: /ea/-/za/	98.8	4.4	4.89	1.60
P13: /ʂɑ/-/za/	100	0	5.33	1.03
P14: /ʂɑ/-/za/	98.8	4.4	5.19	1.60
P15: /za/-/za/	100	0	5.49	0.83
All pairs	97.1	17	5.12	1.37

Table 3: Mean accuracy and d' scores with standard deviations for Polish control participants for each different pair. The alveopalatal-retroflex pairs (P5 and P10) are in bold.

For the experimental groups, we first examined whether the participants in the MB group ($n=17$) and in the FB group ($n=19$) performed differently on the AX task in a repeated-measures ANOVA. No significant difference was found between the two input groups: $F(1,16) = .341, p = .5673$. We therefore collapsed both input groups for all following analyses.

We then examined participants' accuracy in the discrimination of the 15 different pairs as a function of time in order to ascertain whether discrimination of the non-native phonemes changed over the three testing sessions (T1, T2, T3). Mean accuracy results for the experimental group showed a main effect of Test in a repeated-measures ANOVA: $F(2,35) = 5.541, p = .0058$. Post-hoc pairwise comparisons (Holm-corrected) showed that there was a significant difference between T1 and T3 ($p = .0011$), but that the difference just missed significance between T1 and T2 ($p = 0.0624$) as well as between T2 and T3 ($p = .0532$). Results also showed a significant effect of Pair: $F(14, 35) = 160, p < .0001$, indicating that perceptual sensitivity was different among the 15 different pairs. Post-hoc tests showed significant differences between performance on P5 and P10 and the other 13 different pairs. Furthermore, there was a significant interaction among the factors Test and Pair: $F(28, 35) = 4.591, p < .001$, indicating that the effect of Test was not equivalent for all the Pairs. This can be attributed to the fact that only pairs P5 and P10 showed significant improvement.

We next examined d' scores for the experimental group, which showed no significant effect of Test (T1, T2, T3) in a repeated-measures ANOVA: $F(2,35) = 0.962, p = .387$. Post-hoc tests showed however that the difference between T2 and T3 just missed significance ($p = .059$). Results further showed a significant effect of Pair: $F(14, 35) = 124.5, p < .0001$, indicating that perceptual sensitivity was different among the 15 different pairs. Furthermore, there was a significant interaction among the factors Test and Pair: $F(28, 35) = 4.591, p < .001$, indicating that the effect of Pair was not equivalent at each of the three

test times. Mean accuracy and mean d' along with standard deviations for the experimental group at all three test times are presented in Table 4.

Pair	Accuracy						D'					
	T1		T2		T3		T1		T2		T3	
	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>
P1: /zɑ/-/sɑ/	99.5	2.8	98.6	4.6	98.6	4.6	5.40	1.10	5.04	1.55	4.72	1.46
P2: /zɑ/-/cɑ/	97.3	7.4	99.1	3.8	99.1	3.8	5.15	1.18	5.11	1.43	4.79	1.47
P3: /zɑ/-/ʂɑ/	98.6	4.6	98.6	4.6	98.2	6.6	5.29	1.13	5.05	1.36	4.73	1.44
P4: /zɑ/-/zɑ/	97.7	5.8	97.7	5.8	97.7	5.8	5.17	1.33	4.93	1.50	4.61	1.36
P5: /zɑ/-/zɑ/	24.9	31.0	30.1	35.9	41.2	40.4	0.78	1.18	0.83	1.92	1.33	2.21
P6: /sɑ/-/cɑ/	97.7	5.8	100	0	97.7	5.8	5.27	1.33	5.71	0.83	5.53	1.05
P7: /sɑ/-/ʂɑ/	97.2	7.4	99.1	3.8	99.1	3.8	5.26	1.33	5.59	0.90	5.70	0.85
P8: /sɑ/-/zɑ/	99.5	2.8	99.1	3.8	99.1	3.8	5.51	0.92	5.53	1.04	5.70	0.85
P9: /sɑ/-/zɑ/	99.1	3.8	98.6	4.6	99.1	3.8	5.45	1.06	5.53	1.03	5.70	0.96
P10: /cɑ/-/ʂɑ/	24.5	28.6	28.2	31.9	37.5	36.0	0.94	1.57	1.07	1.73	1.15	1.55
P11: /cɑ/-/zɑ/	99.1	3.8	100	0	98.2	5.3	5.29	1.21	5.38	0.94	4.77	1.42
P12: /cɑ/-/zɑ/	97.3	6.2	98.2	6.6	98.2	8.7	5.06	1.32	5.19	1.00	4.87	1.27
P13: /ʂɑ/-/zɑ/	98.2	6.6	98.2	5.3	99.5	2.8	5.24	1.10	5.41	1.05	5.20	0.97
P14: /ʂɑ/-/zɑ/	99.5	2.6	98.6	4.6	97.7	5.8	5.38	0.94	5.47	1.13	4.96	1.16
P15: /zɑ/-/zɑ/	100	0	99.5	2.8	98.6	4.6	5.73	0.80	5.57	0.89	5.48	1.14
All pairs	88.7	27.7	89.6	27.0	90.6	24.9	4.73	1.92	4.76	1.96	4.62	1.90

Table 4: Mean accuracy and d' scores with standard deviations for experimental participants for each different pair at each test time. The alveopalatal-retroflex pairs (P5 and P10) are in bold.

3 Discussion

Before any discussion of our experimental results, it is first worth noting that the Polish control group performed significantly worse on the discrimination of the voiced and unvoiced alveopalatal-retroflex pairs (P5 and P10) than on the discrimination of the other 13 different pairs according to both accuracy and d' scores. This result is not in line with previous work (Zygis & Padgett, 2010), in which native Polish speakers did not show any perceptual differences in an AX discrimination task that employed pairs of CV and VC syllables containing the unvoiced Polish fricatives /s, ʂ, ɛ, ʃ/ (though participants did perform worse on the discrimination of /ʂ/ and /ʃ/ in VC syllables, which the authors attribute to phonotactic constraints). We can only speculate as to why the native speaker group in the current study showed less sensitivity to pairs P5 and P10. One possibility lies in methodological differences between the current study and previous work. Firstly, the current study tested both the voiced and unvoiced pairs of Polish sibilants possibly rendering the task more difficult. In addition, Zygis and Padgett (2010) used an ISI of 30 msec, while the current study used an ISI of 250 msec. A relatively shorter ISI is thought to allow listeners to compare stimuli while they remain in auditory sensory memory and thus give the listener an advantage in discriminating acoustic variation, therefore the longer ISI in the current study could also have rendered discrimination more challenging. One further possibility is that the stimuli used in the current study differ along some unidentified parameter from the stimuli used in previous work. However, even though the native speakers performed worse on the alveopalatal-retroflex pairs, they did perform well above chance on accuracy (78.6% correct on P5 and 87.0% correct on P10) and showed strong discrimination in d' scores (3.69 on P5 and 3.51 on P10), attesting to the fact that this distinction is not a difficult one for native speakers.

The experimental group showed greatly diminished discrimination of the pairs P5 and P10 relative to the other pairs. Participants performed basically at ceiling according to both accuracy and d' scores for all pairs except P5 and P10. This result is in line with previous work that tested native speakers of English (again only on the unvoiced alveopalatal-retroflex pair, P10, in the current study) and found that listeners have difficulty distinguishing the Polish contrast (Zygis & Padgett, 2010). Our participants did not take part in an identification task in which they were asked to label the stimuli as has been seen in previous work

(Lisker, 2001, McGuire, 2007), however based on previous work and following the predictions of Best's (1995) Perceptual Assimilation Model, we believe that, initially at least, the alveopalatal /ç, z/ and retroflex /ʂ, ʐ/ sibilants were both assimilated to the existing French palato-alveolar categories /ʃ/ and /ʒ/, respectively, and therefore made it difficult for French listeners to discriminate the two.

Improvement in the discrimination of this contrast has also been observed in native English speakers following phonetic training and feedback (Lisker, 2001; McGuire, 2007). The current study did not incorporate any type of phonetic training or targeted pronunciation instruction, yet the current results suggest a significant increase in the ability of participants to discriminate the non-native sounds after just hours of exposure to the target language. An analysis of mean accuracy shows significant improvement in perceptual sensitivity between T1 (0 hours of exposure) and T3 (10.5 hours of exposure). Differences between T1 and T2 and between T2 and T3 also approached significance. An analysis of d' scores did not reveal a significant effect of Test, however, the difference between perceptual sensitivity at T2 (4.5 hours of exposure) and at T3 just missed significance ($p = .059$). Given that d' is thought to be a more sensitive measure of perceptual sensitivity than accuracy, we are hesitant to claim that participants significantly improved on the discrimination of these non-native sounds. However, the ensemble of both accuracy and d' results suggests that after 10.5 hours of input in the target language, participants were better able to discriminate the unfamiliar phonemes. At T2, after 4.5 hours of input, however, the amount of input was not sufficient to improve discrimination.

We see several factors that could account for the discrepancy between the accuracy and d' data. Figure 1 shows accuracy and d' scores at the three test times for only pairs P5 and P10. Accuracy data show not only a steady increase in discrimination, but also a clear elbow in the slope of the increase at T2 indicating that learning sped up between T2 and T3 (a difference which just missed significance at $p = .053$). Furthermore, mean accuracy for P5 and P10 show similar developmental trajectories. The d' data are more complex and suggest that performance on the two test pairs did not follow the same trajectory. While P10 showed a steady increase across the three test times, P5 started with lower discrimination and then showed a marked increase between T2 and T3, creating an interaction between the two pairs and the three test times.

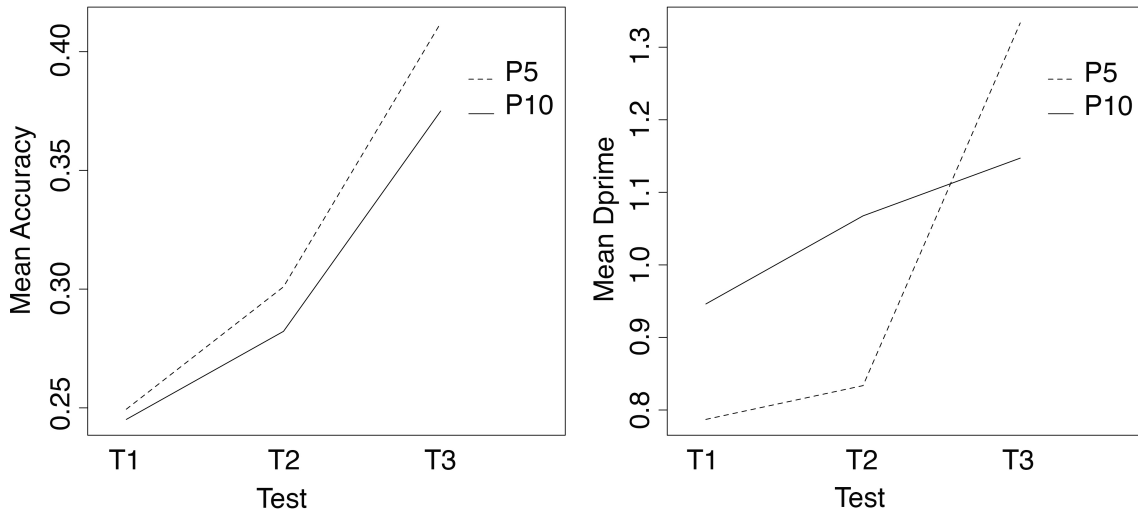


Figure 1: Mean accuracy (left) and mean d' scores (right) at each test time for alveopalatal-retroflex pairs P5 (/ʒa/ - /ʂa/) and P10 (/ea/ - /ʂa/).

A significant increase in accuracy rates but not in d' scores would imply an increase in false alarms (i.e., a 'different' response to a same pair) across the three test times. Given that participants were tested three times on the same task using the same stimuli, it is possible that participants developed a testing strategy that involved responding 'different' more frequently. The objective of the test was not discussed at any time with the participants, however it is feasible that they could have deduced that the test involved distinguishing sounds that are difficult to discriminate thus giving them a slight bias toward expecting more pairs to be 'different'. A bias toward 'different' responses would not be reflected in accuracy rates.

Furthermore, closer analysis of both accuracy and d' data reveals that the two pairs in question, P5 and P10, had much higher standard deviations than the other 13 pairs, indicating a high level of individual variability among participants. In the case of d' prime scores, higher variability among participants could further have diminished the effect of Test even though overall discrimination improved.

Participants' discrimination of non-native phonemes improved without any explicit attention or instruction given to either individual segments or general pronunciation during the course. We interpret this improvement to be evidence of implicit or statistical language learning (i.e. learning without intention or awareness). While our study was not specifically designed to test a statistical learning model of phonological acquisition, the fact that discrimination improved without targeted instruction or attention to pronunciation lends support to a theory that proposes that listeners track distributions of acoustic information in the input. As the input accumulated over time, our participants became better at discriminating non-native contrasts to which they had not been previously exposed. Furthermore, no significant difference was observed between the meaning-based and form-based input groups, suggesting that phonological learning was independent of the type of exposure, further supporting an implicit learning model.

Crucially, however, though performance improved on the non-native contrasts after just 10.5 hours of exposure to novel input, mean accuracy on pairs P5 and P10 remained below chance (41.2 % and 37.5% correct, respectively) and d' scores remained below 1.5 (1.33 and 1.15, respectively) at T3, indicating very low perceptual sensitivity. This raises the issue as to whether or not novel categories were actually formed even though perceptual sensitivity improved. Our results do not definitively show that learners can form novel phonemic categories after just 10.5 hours of exposure, however, they do suggest that the perceptual system begins very early on to extract regularities in the acoustic signal and that perceptual sensitivity to these distributions can improve with minimal input.

The current results are not only in line with seminal work demonstrating that humans are sensitive to distributional regularities in novel acoustic input after very limited exposure (Saffran *et al.*, 1996), but our results also support research specifically examining the learning capacities of learners at first exposure, that is during the very first hours of exposure to a L2. Much of this research has focused on word learning at first exposure (for a review, see Carroll, 2013). Gullberg, Roberts, and Dimroth (2012) found that after just 7 minutes of exposure to Mandarin Chinese in the form of a video presenting a weather report, Dutch-speaking participants with no prior exposure to Mandarin Chinese were able recognize words that had appeared in the video. Shoemaker and Rast (2013) found that French-speaking participants improved on their ability to segment and extract lexical forms from Polish sentences after just 6.5 hours of exposure to the target language. These authors attributed the improvement to increased sensitivity to the Polish phonological system. Specifically, the authors posit that French participants gained sensitivity to the regular distribution of stress in Polish (stress consistently falls on the penultimate syllable of Polish words) and that they used this information to locate word boundaries and extract discrete lexical items from running speech. The current results, which examine perceptual sensitivity at the segmental level, thus contribute to an ever-expanding field that seeks to examine learning capacities and strategies of learners at the very beginning stages of exposure to a new language.

4 Conclusion

The current study examined the development of non-native phoneme discrimination in the very first hours of exposure to a second language. To our knowledge this is the first study to investigate the development of the perception of non-native phonemes in real language input at first exposure. Our results suggest a rapid increase in the ability of participants to discriminate non-native sounds after just 10.5 hours of input in the target language, shedding light on the developmental course of adult phonological acquisition. While our participants did show evidence of increased discrimination of the non-native phonemes, discrimination was still low relative to syllables containing familiar phonemes. Further research is thus needed to determine more precisely at what point in phonological development learners can create and maintain stable L2 categories.

References

- Best, Catherine. (1995). A direct realist view of cross-language speech perception: New Directions in Research and Theory. In Winifred Strange (ed.) *Speech perception and linguistic experience: Theoretical and methodological issues*. (pp. 171–204). Baltimore: York Press.

- Best, Catherine, McRoberts, G. & Sithole, N. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu Click Discrimination by English-Speaking Adults and Infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3), 345-360.
- Best, Catherine, McRoberts, G. & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, 109, 775-794.
- Boersma, Paul & David Weenink (2007). Praat: doing phonetics by computer. [Computer program].
- Carroll, Susanne. (2013). Introduction to the special issue: Aspects of word learning on first exposure to a second language. *Second language research*, 29(2), 131-144.
- Dimroth, Christine, Rebekah Rast, Marianne Starren & Marzena Watorek. (2013). Methods for studying a new language under controlled input conditions: The VILLA project. *Eurosla Yearbook*, 13, 109-138.
- Flege, James. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 233-273). Timonium, MD: York Press.
- Gullberg, Marianne, Leah Roberts & Christine Dimroth. (2012). What word-level knowledge can adult learners acquire after minimal exposure to a new language. *The International Review of Applied Linguistics in Language Teaching*, 50, 239-276.
- Kuhl, Patricia. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews: Neuroscience*, 5, 831-843.
- Lisker, Leigh. (2001). Hearing the Polish sibilants [s š ś]: phonetic and auditory judgements. In N. Grønnum, & J. Rischel (Eds.), *Travaux du Cercle Linguistique de Copenhague XXXI. To honour Eli Fischer-Jørgensen* (pp. 226-238). Copenhagen: C.A. Reitzel.
- McGuire, Grant. (2007). English listeners' perception of Polish alveopalatal and retroflex voiceless sibilants: A pilot study. UC Berkeley Phonology Lab Annual Report: 391-415.
- Macmillan, Niel & C. Douglas Creelman. (1991). *Detection Theory: A User's Guide*. New York: Cambridge University Press.
- Nowak, Pawel. (2006). The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of Phonetics*, 34 (2), 139-152.
- Peña, Marcela, Luca Bonatti, Marina Nespors & Jacques Mehler. (2002). Signal-driven computations in speech processing. *Science*, 298, 604-607.
- Rochon, Marzena & Bernd Pompino-Marschall. (1999.) The articulation of secondarily palatalized coronals in Polish. Proc. 14th ICPHS, San Francisco, 1897-1900.
- Saffran Jenny, Elyssa Newport & Richard Aslin. (1996). Word segmentation: The role of distributional cues. *Journal of Memory & Language*, 35, 606-621.
- Schneider, Walter, Amy Eschman & Anthony Zuccolotto. (2002). *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools, Inc.
- Shoemaker, Ellenor & Rebekah Rast. (2013). Extracting words from the L2 speech stream: a first exposure study. *Second Language Research*, 29 (2), 35-53.
- Werker, Janet & Richard Tees. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49-63.
- Zygis, Marzena. (2003). Phonetic and Phonological Aspects of Slavic Sibilant Fricatives. In Hall, T. A. and S. Hamann (eds). *ZAS Papers in Linguistics*, 32, 175-213.
- Zygis, Marzena & Jade Padgett. (2010). A perceptual study of Polish sibilants, and its implications for historical sound change. *Journal of Phonetics*, 38(2), 207-226.