Citation: Żygis, Marzena & Jaye Padgett (2010). A perceptual study of Polish fricatives, and its implications for historical sound change. Journal of phonetics 38, 207-226.

A Perceptual Study of Polish Fricatives, and its Implications for Historical Sound Change

Marzena Żygis[a], Jaye Padgett[b]*

[a] Zentrum für Allgemeine Sprachwissenschaft, Schützenstr. 18, 10117 Berlin, Germany
marzena@zas.gwz-berlin.de

[b] University of California, Santa Cruz, Santa Cruz, CA 95064, USA
padgett@ucsc.edu

*Corresponding author: Tel.: +1 831 4593157, Fax: +1 831 459 3334, University of California, Santa Cruz, Santa Cruz, CA 95064, USA, padgett@ucsc.edu

Abstract

The present study probes perception of place distinctions among Polish sibilants using an AX discrimination task, and compares results from thirteen Polish- and ten English-speaking subjects. Besides providing information on the relative discriminability of the sibilants, the perceptual study is designed to investigate the claim that a particular kind of diachronic change which has taken place in Polish and other languages, as well as related facts about sibilant inventories, could be perceptually motivated. The results lend support to this claim and to the general view that a principle of dispersion plays a role in explaining sound change tendencies, and therefore in shaping phonological tendencies, for consonants, not only vowels.

Keywords: sibilants, perception, dispersion, perceptual map, sound change.

1. Introduction

Polish has an unusually rich inventory of sibilant sounds, contrasting denti-alveolar [s, z, ts, dz], alveolo-palatal [ɕ, ʑ, tɕ, dʑ], and retroflex [ʂ, ʐ, tʂ, dʐ][1] places of articulation. In addition, a palatalized palatoalveolar sound [ʃʲ] exists as an allophone of /ʂ/ when this phoneme occurs before [i] or [j] as in To[ʃʲ]iba ‘Toshiba’. A good deal is understood about the acoustic and articulatory properties of the three contrastive sibilant places of articulation in Polish (e.g., Dogil, 1990; Halle & Stevens, 1997; Jassem, 1962, 1979; Żygis & Hamann, 2003). Considerably fewer perceptual studies have been conducted on the complex Polish sibilants (Lisker, 2001; Nowak, 2006; Żygis & Hamann, 2003).

This paper presents the results of an AX discrimination task employing Polish sibilants of all four places of articulation mentioned above, in particular the sibilants [s, ɕ, ʃʲ, ʂ]. Results from both English and Polish listeners are presented, based on subject errors and reaction times. This design allows us to observe effects that are language particular, but also effects that are shared by both language groups. In fact, an important goal here is to probe for a language-independent ‘perceptual map’ of the Polish sibilants by place of articulation. This is done by means of a multidimensional scaling analysis using the reaction time data. The results here suggest that both noise spectra and formants contribute significantly to the perception of the Polish contrasts, with dentals and alveolo-palatals grouping against palatoalveolars and retroflexes on the one hand (high noise 'tonality' vs. low), and alveolo-palatals and palatoalveolars grouping against dentals and retroflexes on the other (high F2 vs. low).

The idea that perceptual dispersion might influence vowel inventories is familiar due to pioneering work by Lindblom and later researchers (especially Liljencrants & Lindblom, 1972; Lindblom, 1986, 1990; see also Schwartz, Boë, Vallée, & Abry, 1997 and references therein). The second reason for constructing a perceptual map of Polish sibilants, apart from the goal of learning more about the perception of these sibilants, is to allow us to extend dispersion-theoretic thinking to this class of sounds. In particular, the findings here are related to a category of historical sound change involving sibilants, and to relevant facts about sibilant inventories.

---

[1] Most recent studies on Polish sibilants assume retroflexes instead of palatoalveolars; see Zygis (2006) for a detailed discussion.

Phonemic contrasts among postalveolar sibilants are relatively uncommon, and languages strongly avoid a three-way postalveolar contrast (e.g., alveolo-palatal [ɕ], palatoalveolar [ʃ], and retroflex [ʂ]). There are several independently attested cases in which a language's palatoalveolar series underwent an unconditioned historical change to retroflex, a change called ‘sibilant retroflexion’ in this paper. In Polish, for example, a series of palatalized palatoalveolars depalatalized and became retroflex during roughly the 16th century; examples are given in (1) (Rospond, 1971:91, 110ff; transcriptions are our own).

(1) Polish sibilant retroflexion

[t͡ʃʲi]sto	>	[t͡ʂɨ]sto	‘clean’
[ʃʲi]ja	>	[ʂɨ]ja	‘neck’
[ʒʲi]to	>	[ʐɨ]to	‘rye’

Though depalatalization is not surprising, since it can be viewed as an articulatory simplification, sibilant retroflexion is a puzzling change, since by the usual criteria of comparison retroflexes seem disfavored by languages compared to palatoalveolars. In particular, languages with palatoalveolars far outnumber those with retroflexes, and if a language has only one postalveolar it is likely to be palatoalveolar (Maddieson, 1984).

In the cases of sibilant retroflexion known to us, however, a series of alveolo-palatals already existed in the language. Polish had the alveolo-palatals [ɕ ʑ tɕ dʑ], for example, which emerged from palatalized dentals most likely around the 13th century (Rospond, 1971; Stieber, 1962, 1968). Żygis (2003) argues that sibilant retroflexion was favored because it increased the perceptual distance between the postalveolars, with an alveolo-palatal vs. retroflex contrast being better than an alveolo-palatal vs. palatoalveolar one; see also Padgett & Żygis (2007). The idea is schematically illustrated in Figure 1. (Sound changes are shown by arrows and the relative perceptual distances by horizontal lines.)

Figure 1

The same perceptual reasoning extends to explain an otherwise puzzling fact about sibilant inventories: though palatoalveolars seem to be the ‘unmarked’ postalveolar sibilant place (by diagnostics mentioned above), there are a number of languages having alveolo-palatal and retroflex fricatives as their *only* postalveolar sibilants, as shown in Table I. (See discussion and references in Padgett & Żygis, 2007.)[2] There are also languages with complex sibilant inventories having [ʃ] in contrast with [ʂ], and more rarely, with [ɕ] also shown in Table I.

Sibilant inventory	Language	References
s ɕ ʂ	Mandarin, Komi Permyak, Telugu, Malayalam, Serbian dialects, Naxi	Ladefoged & Maddieson (1996), (Kochetov & Lobanova, 2007; See discussion and references in Padgett & Żygis, 2007.), Maddieson (1984), Mohanan, p.c, Żygis (2006).
s ʂ ʃ	Burushaski, Hmong, Jacaltec, Tulu, Tarascan, Guambiano, Pasto, Camsa, Acoma	UPSID (Maddieson & Precoda, 1992)
s ʃ ɕ	Chuvash, Tehuelche	UPSID (Maddieson & Precoda 1992)

Table I: Phonemic inventories of sibilant systems.

Based on a typological survey of Slavic sibilants, Żygis (2006) concludes that, among languages having a contrast in postalveolar sibilants, one of them is strongly preferred to be retroflex, as in languages with inventories [s ɕ ʂ] and [s ʂ ʃ] (see Table I). There are only two languages out of the 453 found in UPSID in which [ʃ] contrasts with [ɕ] (Chuvash and Tehuelche). Taking into consideration acoustic results, Żygis (2006) hypothesized that the typological preferences in sibilant systems are perceptually grounded: [ɕ] vs. [ʂ] and [ʂ] vs. [ʃ] create well-dispersed contrasts whereas [ɕ] vs. [ʃ] does not.[3]

The experiment presented here seeks to test this explanation of the sound change discussed above. In hypothesizing that perceptual constraints play a role in explaining sound

---

[2] As far as a two-way contrast is concerned, it appears that the most frequent contrast is between [s] and [ʃ], as attested in 187 languages included in UPSID. In comparison [s] vs. [ʂ] is found in only 5 languages: Mazatec, Papago, Tacana, Luiseño, and Dagur.

[3] Our results below do not support the notion that [ʂ ʃ] are well dispersed. This may be because the proxy for [ʃ] in our experiment is Polish [ʃʲ]. See later discussion.

change and/or phonological tendencies, we follow work of Ohala (e.g., 1981, 1983) and Stevens (e.g., 1972; Stevens & Keyser, 1989), and Lindblom (1986, 1990), among others.

The article is organized as follows. Section 2, reviews four previous studies of the perception of Polish fricatives (Lisker 2001, Żygis & Hamann 2003, Nowak 2006 and McGuire 2007, 2008). Section 3 reports on the methodology of the perceptual experiment conducted with English and Polish informants, presenting the acoustic properties of the stimuli. Section 4 presents the results of the perceptual experiment, section 5 is devoted to a discussion of the results, and section 6 concludes.

2. Previous studies

In three identification experiments Lisker (2001) investigated the perception of Polish sibilants by English listeners. In each experiment, four tokens of the three Polish nonsense syllables [sɑ, ʂɑ, ɕɑ][4] as pronounced by an adult male speaker of Polish were recorded for presentation to ten English listeners. Three versions of each token were then employed: (i) one exactly as recorded, (ii) one with only the frication noise interval, and (iii) one with the vowel and transitions which remain after removing the frication noise interval.

In the first experiment subjects were presented with all stimulus types. The full CV stimuli were recognised as expected: [sɑ] was clearly distinct from [ʂɑ] and [ɕa], while the latter two could not be well separated. This can be attributed to the phonological background of English listeners, who have a two-way contrast [s] vs. [ʃ]. When the isolated noise intervals were presented, the [ɕ] vs. [ʂ] contrast was better recognized. The presentation of the vocalic portion yielded the best results with respect to the recognition of the [ɕ] vs. [ʂ] contrast. The latter result, as suggested by Lisker, is probably due to the fact that listeners perceived the formant transition of the alveolo-palatals as the palatal glide, which created a better contrast to retroflexes.

The subjects' better performance categorizing the postalveolars that were removed from their vocalic context led Lisker to conclude that subjects might be responding to such stimuli in a purely auditory (non-speech) mode. In Lisker's next experiments, subjects therefore reported auditory judgements of relative loudness and pitch. The stimulus set was

---

[4] The transcription is in accordance with symbols used in the present study. In Lisker (2001) the symbols [s, š, ś] were used.

limited to the one contrast [ɕ] and [ʂ] presented in the three forms described above. Listeners uniformly found [ʂ] to be louder than [ɕ] in the isolated noise conditions; they were less reliable in making this judgement for the full CV syllables. The latter result could be related to the fact that listeners found the isolated vocalic portions of [ʂɑ] to be quieter than that of [ɕɑ].[5] For the task involving pitch, listeners were asked to judge the relative "pitch" of the stimuli in all conditions (though this would mean rather different things for the vocalic vs. fricative noise portions of stimuli). The results indicated that the English speakers judged [ɕɑ] to be "higher-pitched" than [ʂɑ] for all conditions (full CV, isolated frication, isolated vocalic portion).

In Żygis & Hamann's (2003) study the goal was to examine the claim made by Halle & Stevens (1997) and Ladefoged & Maddieson (1996) that the Polish alveolo-palatal fricatives [ɕ ʑ] are palatalized post-alveolars [ʃʲ ʒʲ]. As noted above, the sound [ʃʲ] exists in Polish independently of [ɕ], as an allophone of /ʂ/ before [i] and [j]. On the basis of two perceptual experiments Żygis and Hamann argued that [ɕ ʑ] and [ʃʲ ʒʲ] constitute two separate phonetic categories in Polish.

For the purposes of Żygis and Hamann's experiments the Polish sibilants [s ʂ ɕ ʃʲ] were recorded in pre-, post- and intervocalic position using the three vowels [ɑ i u]:

a) prevocalic context _ (i, u, ɑ)
b) postvocalic context (i, u, ɑ)_
c) intervocalic context i_i, u_u, ɑ_ɑ

In the first experiment, only stimuli in the context [_ɑ] were used. They were presented to twenty Polish native speakers who identified the given sound either as [s], [ʂ], [ɕ] or [ʃʲ]. The results showed a high number of correct identifications (more than 96% for every sound). It also appeared that the listeners had more difficulty in discriminating the contrast between [ɕ] and [ʃʲ] than between other pairs.

---

[5] For those stimuli lacking the fricative noise, subjects were asked to judge the loudness of the vowel's "onset".

In the second experiment, the influence of the vowel context was tested in an AX discrimination task. In addition, the role of native language background was investigated by asking twenty five Polish and twenty five German native speakers to discriminate the fricatives. The results confirmed that the alveolo-palatal [ɕ] and the palatalized palatoalveolar [ʃʲ] are distinct phonetic categories for both Polish and German speakers. As the latter group performed slightly worse in discriminating the sibilants (especially [ʃʲ] from [ɕ]), the conclusion was drawn that native perceptual background influenced the results. In one context, intervocalic [ʂ] and [ʃʲ], German speakers were slightly better in performing the task.

Nowak (2006) examined the role of vowel transitions and frication noise in the perception of Polish sibilants [s, ʂ, ɕ]. In one experiment Nowak studied how native speakers of Polish categorize fricatives occurring without any vocalic cues. Eighteen repetitions of all three sibilants [s, ʂ, ɕ] were presented to five native speakers who had to circle the appropriate fricative on an answer sheet. The results showed that frication alone is a sufficient cue for Polish native speakers to recognise the fricatives. All three categories were excellently recognized (100% correct responses for dentals, 99.6 % for retroflexes and 97.8 for alveolo-palatals).

In another experiment nine nonce words [ese, eʂe, eɕe, asa, aʂa, aɕa, usu, uʂu, uɕu] were divided into vocalic and consonantal parts. One token of each word was chosen to contribute the fricative part, and another to contribute the vocalic part. The fricative and vocalic parts were recombined in various splicing conditions in order to test the relative importance of noise vs. formant cues. The subjects, seven native speakers of Polish, were asked to assign the medial fricative into the dental, retroflex or alveolo-palatal category by circling the corresponding orthographic symbol on an answer sheet. The results showed that the recognition of alveolo-palatals was most affected by splicing conditions; in particular, without the following ([j]-like) transition this sound was easily confused with retroflexes. On the other hand, dental noise led to dental identification virtually regardless of splicing condition.

In summary, Nowak's study showed that the three sibilant categories were reliably identified by Polish native speakers based on frication noise alone. However, when formant cues conflicted with noise cues they could override them in particular cases.

McGuire (2007, 2008) studied the effect of training on the ability of English listeners to discriminate and label stimuli as [ɕa] or [ʂa]. McGuire altered tokens of these syllables, produced by a native speaker of Polish, so as to produce a two-dimensional continuum of stimuli, where the dimensions were fricative noise and vocalic transition. English listeners initially categorized and were better able to discriminate tokens based on the vocalic transition. However, McGuire's results showed that they could be trained to improve at both tasks on both dimensions, whether tokens were varied in only one or both of the dimensions, though there was a good deal of variability in subjects' abilities.

The perceptual studies summarized above lead to a number of conclusions. First, fricative noise provides a strong basis for distinguishing among sibilants of different place in Polish, perhaps a primary basis for Polish listeners. However, formant transitions do play a role in distinguishing Polish sibilants too, at least for the purposes of distinguishing alveolo-palatals from retroflexes (where the transition following the alveolo-palatal is most important). Both of these conclusions – that noise is often sufficient and may be primary, while formant transitions can nevertheless matter – are consistent with a line of research on perception of sibilants in a number of languages (e.g., Mann & Soli, 1991; Nowak, 2006, and references therein; Repp, 1981; Whalen, 1991). Third, English speakers (unsurprisingly) have difficulty distinguishing alveolo-palatals from retroflexes, and they do so primarily based on formant transitions; however, they can be trained to become sensitive to the relevant differences in fricative noise spectra. Finally, the largely allophonic sound [ʃʲ] is distinguished from the three other sibilant series by Polish and German listeners, though the distinction between this sound and the alveolo-palatal [ɕ] causes some difficulty even for Polish listeners.

The experiment reported here differs from those discussed above in a number of respects. First, unlike the studies of Lisker (2001), Nowak (2006), and McGuire (2007, 2008), it includes not only the contrastive sibilant series but the allophonic [ʃʲ]. Second, it compares results for English and Polish listeners. (See Mcguire 2007 for a comparison of English and Mandarin listeners on the [ɕa] - [ʂa] contrast.) Third, a major goal is to draw conclusions about the perceptual space of this four-sibilant inventory and to relate those results to cross-linguistic inventory patterns and sound changes as discussed above. For example, if Polish sibilant retroflexion were perceptually motivated, then one would expect that the output inventory [s ɕ ʂ] is overall better dispersed than the input to the sound change, [s ɕ ʃʲ].

3. Methodology

3.1. Subjects

The English-speaking subjects were 10 undergraduate students from the University of California, Santa Cruz, ages ranging from 19 to 26. All had some training in linguistics, and they were paid for their participation. None were native speakers of any other language. The Polish-speaking subjects were 13 high school students recruited in Szczecin, Poland. They were 16-17 years old and native speakers of Standard Polish. They could speak German and English on a basic or intermediate level. All subjects reported normal hearing.

3.2. Stimuli

The stimuli used in the experiment included both CV and VC pairs, where V is always [a] and C stands for [s], [ʂ], [ɕ], or [ʃ^{j}]. The diagram in Figure 2 schematically indicates the phonological status of each of the sibilants tested. While [s ɕ ʂ] are distinct phonemes of Polish ([s]ali ‘hall’ gen.sg., [ɕ]ali ‘they sowed’ [ʂ]ali ‘scale’ gen.sg.), [ʃ^{j}] is regarded as an allophone of /ʂ/ which occurs when /ʂ/ is followed by [i] or [j], e.g. To[ʃ^{j}]iba. Within words this sequence is generally found only in borrowings. But this allophone of /ʂ/ also occurs widely in phrases where a word ending in /ʂ/ is followed by a word beginning with [i] or [j], e.g. ko/ʂ/ /j/abłek is pronounced as ko[ʃ^{j} j]abłek ‘a basketful of apples’ (Rubach, 1984; Wierzchowska, 1971).

Figure 2

In spite of its allophonic status, [ʃ^{j}] is included in the experiment because of its relevance to the questions about sound change and inventories addressed here. Since sibilant retroflexion, and the existence of postalveolar sibilant inventories having [ɕ] and [ʂ] but no palatoalveolar, might be explained at least in part by the perceptual ‘location’ of palatoalveolars with respect to these other postalveolars, it is important to include a palatoalveolar in the study. Since sibilant retroflexion in Polish *eliminated* the class of (palatalized)[6] palatoalveolars, it should

---

[6] All palatoalveolars at the relevant stage of Polish were palatalized – there were no plain correspondents – because they had been derived through a palatalizing mutation of velars, e.g. [k^{j}] > [tʃ^{j}].

have been impossible to do this employing only Polish sounds. But as noted above [ʃʲ] has re-entered the language. In addition, Żygis and Hamann (2003) have shown Polish speakers do rather well distinguishing this sound from other postalevolars.[7]

The choice of [a] for vocalic context was chosen in order to avoid palatalizing and rounding effects, and to avoid phonologically illicit syllables as much as possible. All four sibilants tested do occur before this vowel in Polish, though [ʃʲ] only marginally (see note 7). In comparison, [ʂ] is impossible before [i], and [ʃʲ] is impossible before any remaining Polish vowel. As for the context [a_], [ʃʲ] is found only if the next word begins with [i] (or the glide [j]); this point may be relevant to the results.

Seven tokens of each stimulus were recorded by a native speaker of Polish (the author MZ) and digitized at 22 kHz. The data were recorded on a DAT-Recorder TASCAM DA 20 MK 2 in a sound-proof laboratory at the Zentrum für Allgemeine Sprachwissenschaft in Berlin. The stimuli were all scaled to have the same peak amplitude using Praat (Boersma & Weenink, 2007).[8] Stimulus pairs were constructed based on these tokens. 'Same' pairs were always constructed of different tokens of the same stimulus type.

3.3. Experimental procedure

The experiment employed was an AX discrimination task. The stimulus pairs were grouped into 12 blocks based on the six types of different CV pairings and similarly six VC pairings, and each block contained trials involving only the same two fricatives; for example, one block contained only the pairs sa/ɕa, ɕa/sa, sa/sa, ɕa/ɕa. Since there were four trial types (same vs. different x stimulus order) and seven instances of each trial (constructed from the original seven tokens of each stimulus), there were 28 trials per block. This gave a total of 12×28 = 336 trials per subject. Trials within blocks and also blocks themselves were randomized between listeners. The stimuli were natural, unedited, and without added noise, and the interstimulus interval was 30 ms.

---

[7] Furthermore, [ʃʲ] occurs in some proper names even before the vowel [a]. We would have liked to include [ʃ] as a possible stimulus, but our stimuli come from a Polish speaker (see below) and [ʃ] does not occur in Polish. Having a speaker of another language produce it would have introduced a confound between distinct sounds and distinct speakers. However, Padgett and Żygis (2007) conducted an acoustic study of Bulgarian, Polish, and Russian sibilants. Based on that study's results Bulgarian [ʃ] and Polish [ʃʲ] were grouped into one sibilant set due to their similarity in center of gravity.

[8] The Praat command "Scale" was used to adjust each sound file's amplitude such that peak amplitude was 0.8 (out of 1).

The choice to use a fixed discrimination design and stimuli without noise was made with the goal of making the task as easy as possible for the English speakers relative to the Polish speakers. The point of this, and the use of reaction times with encouragement that subjects respond as quickly as possible, was to obtain results that might minimize language-particular differences (though the latter are interesting and did emerge especially in errors made). There is reason to assume a distinction between short-term auditory and phonetic category memories (Pisoni, 1973). Building on this idea, Babel and Johnson (to appear) and Johnson and Babel (to appear) argue that sufficiently speeded RTs in a discrimination task can access processing at the auditory ('psycho-acoustic') level rather than at the phonetic ('subjective perceptual processing') level.

The experiment was carried out in a quiet room, at UC Santa Cruz (English speakers) and a Szczecin high school (Polish speakers), using headphones and an iMac G3 (English) or Powerbook G4 (Polish) running Superlab Pro version 1.75. Instructions were presented on the computer screen. Subjects were advised to respond to trials as quickly and accurately as possible. (All instructions and feedback for Polish speakers were in Polish.) No time limit was employed, and all responses were captured. The experiment began with a practice block of 28 trials comparing tokens of [as] and [aɕ]. Subjects pressed a color-coded key on the computer keyboard to indicate whether they found a pair of stimuli to be the "same" or "different". After each trial they were given visual feedback about the correctness of their response and reminded to respond as quickly and accurately as possible. Both correctness and reaction times were recorded.[9] Reaction times were measured from the end of the second stimulus. There was an optional rest period between blocks.

3.4. Acoustic properties of the stimuli

This section presents the acoustic properties of the stimuli. In doing so, it focuses especially on two measures: the center of gravity of the fricative noise portion (COG) and the rate of change of the second formant transition.

It is well known that second formant transitions provide an important cue to place of articulation in fricatives as with other consonants. The research cited in section 2 makes clear that this is the case with Polish sibilants. Another candidate measure is F3, especially since one class of Polish fricatives is described here as retroflex. However, Polish retroflexion

---

[9] Superlab was set to use "Time Manager 1" and the "Generic Keyboard" as input. According to the Superlab manual, this allows for a reaction time resolution of ±16.67 ms given the machines we used.

differs articulatorily from the more extreme retroflexion found in some languages (see Żygis 2006 for discussion and references), and Żygis (2006) found a great deal of variation in the F3 values of vowels preceding and following Polish sibilants, compared to F2 values, especially in the case of retroflexes. F3 was therefore not measured here.

There is less consensus in the literature about what the most important acoustic measures are in the case of sibilant noise itself. In the case of Polish specifically, Dogil (1990) and Kudela (1968) measured the spectral peaks of fricatives. The two studies agree that the alveolo-palatals, retroflexes and palatalized palatoalveolars are similar with respect to the presence of prominent spectral peaks in the frequency range contigous with the F3-F4 formants. What distinguishes retroflexes and palatalized palatoalveolars from alveolo-palatals is the presence of a major amplitude peak in the F2 frequency region in the former sounds (Dogil 1990). Table II shows the peak frequencies found in Dogil's study.

	F2	F3	F4
ɕ	not found	2670Hz	3210Hz
ʃʲ	1680 Hz	2740 Hz	not found
ʂ	1280 Hz	2440Hz	2970Hz

Table II: Peak frequencies of Polish sibilants (Dogil 1990).

Spectral peaks raise challenges as a summary measure of sibilant place. There are typically several peaks in a sibilant spectrum; whether a particular peak is highest, and even whether it appears, can vary significantly from token to token. In the hopes of finding a more integrated and reliable summary measure of sibilant spectra, the center of gravity (COG) measure was chosen here instead. COG is the average of frequencies over some frequency domain weighted by the amplitude, and it has been found useful in distinguishing among fricatives (see also Forrest, Weismer, Milenkovic, & Dougall, 1988; Gordon, Barthmaier, & Sands, 2002; Jassem, 1979; Jongman, Wayland, & Wong, 2000; Nittrouer, Studdert-Kennedy, & McGowan, 1989). Żygis and Hamann (2003) found that the Polish sibilants can be reliably ordered by COG as follows (from highest to lowest): [s > ɕ > ʃʲ > ʂ].

Before turning to the formant rate and COG results, Figure 3 presents the average duration of the fricative portion of the stimuli in CV and VC syllables. This third parameter was investigated because duration has been found in general to influence perception of fricative place (Jongman, 1989).

Figure 3

The fricatives in CV syllables are of similar length (the differences are not significant). In VC sequences the palatalized palataloveolar [ʃʲ] is significantly shorter than the other fricatives (for [ʃʲ] vs. [ʂ] and [ʃʲ] vs. [s] $p < .001$ and for [ʃʲ] vs. [ɕ] $p < .01$, based on a post-hoc Scheffè test with *duration* as dependent variable and *sibilant type* as independent variable, see Table 1a in Appendix I). Note that [aʃʲ] is the one stimulus that is not a phonologically licit phrase-final sequence in Polish (though it could occur in a phrase if [i] or [j] followed). However, it is unclear why that might lead to its different duration.

Figure 4 presents COG values of each fricative in CV and VC syllables, averaged across subjects. For measurements of COG each fricative was divided into three equal intervals. (COG 1-3 correspond to the initial, middle, and final intervals respectively.) All fricatives have dynamic spectral characteristics, but these intervals were distinguished especially with palatalized [ʃʲ] in mind, in order to make clear any effects on COG of formant transitions. In the measurements here, the frequency range was 0-11,025 Hz, and weighting was done by using the power spectrum ($p = 2$) in PRAAT without preemphasis of the signal.

Figure 4

In both CV and VC syllables the stimuli can be ordered from highest to lowest COG value: [s] > [ɕ] > [ʃʲ] > [ʂ] after averaging the three intervals. In both CV and CV syllables all of these differences were highly significant: $p < .05$ for [ʃʲ] > [ʂ] in VC syllables and $p < .001$ for all other comparisons within CV and VC syllables. (See Tables 2f and 2l in Appendix I.) Within fricatives there were few significant differences by interval. In the case of [ɕ] and [ʃʲ] in CV syllables the first interval is significantly lower than the second ($p < .05$), suggesting that the COG values rise in the second and third phase of palatal sounds (see Table 2b). In VC syllables two differences are statistically significant: the COG value of the second interval of [ɕ] is higher than that of its third interval (*p* < .05) and the first interval of [ʃʲ] is lower than its third interval (*p* < .05). (See Table 2h in Appendix 1.)

Figure 5 plots the rate of change (slope) of the second formant transition in CV and VC position. This was defined as the difference in F2 between the vowel at consonant-vowel transition versus at its midpoint, divided by the number of milliseconds between these two points. Formants of the vowel segments were measured semi-automatically by means of Linear Predictive Coding (LPC). Prior to formant analysis the sounds were down-sampled to 11 kHz. The LPC was then calculated by using the following parameters: pre-emphasis frequency 50 Hz, analysis window duration 0.0256 s, time step 0.001 s and a prediction order of 13. LPC spectra were calculated at four time instants: (i) the midpoint of the preceding vowel, (ii) the end of the preceding vowel transition, (iii) the beginning of the following vowel, and (iv) the midpoint of the following vowel.

Figure 5

What stands out in these data are the relatively steep transitions of [ɕ] in both positions and of [ʃʲ] in CV position. Those F2 slopes are significantly higher than all other F2 slopes, except for the comparison between [s] and [ɕ] in VC position. No other differences are significant. (See Table 3a,b Appendix 1). A similar pattern has been obtained in other studies such as Lisker (2001) and Nowak (2006). The rising F2 in [aɕ] and the falling F2 in [ɕa] is due to the fronted tongue blade/body in the production of this fricative. As far as [ʃʲ] is concerned, the F2 is falling in CV syllables and only slightly rising in VC syllables. As the transcription suggests, this sound has a palatal offglide but no corresponding onglide.

The relevance of these results for perception is discussed in section 5.

4. Results

This section presents first the error results and then the reaction time (RT) results of the experiment. As noted earlier, the experiment was designed with the expecation that the two kinds of results might differ in an interesting way. Given the use of unaltered stimuli with no noise added (and given the non-roving experiment design), one might expect that subjects would make relatively few errors in ‘same’ vs. ‘different’ judgements. One might further expect that Polish speakers would do very well on all distinctions, because they are all phonemic in Polish, or in the case of [ʃʲ], because other studies have found Polish listeners to perceive this sound well. English listeners, by comparison, would be expected to make more errors in the case of the postalveolar contrasts. From the point of view of the present study, the RT results are the most telling. If the present methodology succeeds to any extent in tapping a psychoacoustic level of response, one might expect the results to be more similar across the languages.

4.1. Error data

4.1.1. Polish subjects

Figure 6 presents the mean d-prime scores (Macmillan & Creelman, 1991) for all Polish speakers combined, broken down by position (CV versus VC) and fricative pair.[10] It also presents data by subject to give a sense of the variation. Figure 7 shows the mean bias scores for the combined Polish speakers. (A negative bias indicates a predisposition by subjects to respond ‘different’, a positive one ‘same’.) These d-prime and bias data are summarized in Tables 1a-b in Appendix II.[11] The d-prime data were subjected to a two-factor repeated measures analysis of variance with the two factors Position (VC vs. CV) and Fricative Pair. The main effects of Position and Fricative Pair were both significant ($F(1, 12) = 12.06$; $p < .01$ and $F(5,60) = 15.52$; $p < .001$ respectively). The interaction of Position and Fricative Pair was also significant ($F(5,60) = 12.47$; $p < .001$). All of these effects seem attributable to the pair [ʃʲ] vs. [ʂ] in the VC environment.

These Polish speakers would be expected to do very well distinguishing the phonemic contrasts [s ɕ ʂ], especially given the absence of noise or editing of the stimuli, and this is indeed the case. These speakers also do well distinguishing [ʃʲ] from the other sibilants, just as Żygis and Hamann (2003) found. The one exception is for the pair [ʂ] and [ʃʲ] in the VC environment, where the average d-prime score across speakers was much lower. This is also the only condition where bias was not negative overall (indicating less bias to respond

---

[10] In figures and in the text we employ the convention of presenting fricative pairs in the order suggested by their COG ordering (s > ɕ > ʃʲ > ʂ). Therefore we write ‘[ɕ] vs. [ʃʲ]’ but ‘[ʃʲ] vs. [ʂ]’.

[11] A problem arises in computing d-prime scores for a subject who gets all answers correct, or none correct, for any pair of fricatives. This is because d-prime is calculated based on the z transformation of the proportion correct, and it is undefined when the input is either 1 or 0. A typical approach in such cases is to adjust 1 or 0 slightly down or up (resp.) before computing d-prime. We substituted 0.999 and 0.001 (resp.).

‘different’.) As noted earlier, [ʃʲ] is not phonologically licit in phrase-final position as in these [aʃʲ] stimuli, but the question is why this stimulus is confused with [aʂ] and not [as] or [aɕ] in this context. The fricatives [ʂ] and [ʃʲ] have very similar onset F2 transitions and center of gravity values. (See previous section.) No other pair of fricatives are so similar in both regards. (For example, [ɕ] and [ʃʲ] are also similar in COG but have very different onset transitions.) In fact, the substantial acoustic feature distinguishing these fricatives is the falling offset F2 transition associated with [ʃʲ] but not [ʂ]. But this cue is endangered by the lack of any offset vowel in VC stimuli. Putting [aʂ] vs. [aʃʲ] aside, the Polish speakers overall had very high d-prime scores on all comparisons. The individual speaker behavior reflects these overall results very consistently: for all subjects the [aʂ] vs. [aʃʲ] comparison fared the worst. (See Table 1a in Appendix II.)

Figure 6

Figure 7

4.1.2. English subjects

The d-prime scores for English speakers are given in Figure 8, and the related bias scores in Figure 9. Data by subject are summarized in Tables 1c-d in Appendix II. The main effect of Position was not significant for this group ($F(1,9) = 3.57$; $p = .091$). The main effect of Fricative Pair and the interaction of Position and Fricative Pair were both significant ($F(5, 45) = 19.64$; $p < .001$ and $F(2.02, 18.17) = 6.39$; $p < .01$).[12]

The expectation that these subjects would do relatively well distinguishing [s] from the postalveolars, and less well distinguishing among postalveolars, is largely supported by these results. In particular, a post-hoc comparison of all Fricative Pair levels averaged across positions indicates that scores for each fricative pair containing [s] (the first three points in each graph) were significantly better than scores for each fricative pair involving only

---

[12] For the interaction effect, the figures reported reflect the Greenhouse-Geisser adjustment. This adjustment is necessary when the means being compared do not pass the test of ‘sphericity’, that is, when the differences between means are not equal in variability.

postalveolars (the last three points in each graph). Within each of these groups differences were mostly insignificant. (For details see Table 2 in Appendix II.)

When viewed by Position, however, subjects do not appear to do equally poorly on all postalveolar comparisons. For VC stimuli performance seems worst on [aʂ] vs. [aʃʲ] just as with the Polish speakers, though only six of the ten subjects show this pattern (see Appendix II, 1c). For CV stimuli performance is worst in the case of [ɕa] vs. [ʃʲa] (nine out of ten subjects). The fricatives [ɕ] and [ʃʲ] have similar offset formant transitions and center of gravity profiles. (The pair [ʂa] vs. [ʃʲa] are also close in COG values but have very different formant transitions.) The difficulty subjects had with these pairs – [aʂ] vs. [aʃʲ] for VC and [ɕa] vs. [ʃʲa] for CV – might also be reflected in the bias scores for them: the general bias toward 'different' responses is least evident for these pairs.

These differences were tested with paired-samples *t*-tests in the following way. For the VC condition, results for the (lowest scoring) fricative pair [aʂ] vs. [aʃʲ] were compared with those for [aɕ] vs. [aʃʲ] and [aɕ] vs. [aʂ]. Neither difference was significant according to a Bonferroni-adjusted significance level of $p < .025$ ($t(9) = 2.11$, $p = .064$ and $t(9) = 2.65$, $p = .026$ respectively).[13] For the CV condition, the (lowest-scoring) fricative pair [ɕa] vs. [ʃʲa] was compared with [ɕa] vs. [ʂa] and [ʂa] vs. [ʃʲa]. Both of these differences were significant at $p \leq .025$, ($t(9) = -3.71$, $p < .01$ and $t(9) = -3.32$, $p < .01$ respectively).

Figure 8

Figure 9

4.2. Reaction time data

4.2.1 Polish subjects

Figure 10 presents reaction time data for correct 'different' responses for the Polish subjects, broken down again by position and fricative pair. Mean RT data (overall and by speaker) are summarized in Table 3a in Appendix II. As stated in the Methods section, there was no time

---

[13] For this significance level we divided .05 by 2 for the two comparisons made within a Position.

out for subject responses and all data were captured. The boxplots here are computed using all of the RT data, but they are drawn without showing outliers, in order to make differences on the y-axis easier to read. For more information on the overall distribution of the scores, Figure 11 plots all individual RT scores (pooled). Polish RTs overall ranged from 4 to 4142 ms. (English RTs, discussed below, ranged from 56 to 7472 ms.) For purposes of analysis no low values (e.g., 4 ms.) were excluded; though RT was measured from the end of the second stimulus, cues to the place of the second fricative are encountered more than 100 ms. before this point, so that very low RTs were not considered errors. Nor were high outliers excluded. Instead, given the high positive skewness of the RT scores, the data were log transformed for the purposes of statistical analysis (Baayen, 2008:71,97; Johnson, 2008:231). For these data both the Position and Fricative Pair main effects were significant ($F(1,12) = 64.72$; $p < .001$ and $F(5,60) = 8.60$; $p < .001$ respectively), while the interaction was not significant ($F(5,60) = 2.29$; $p = .056$).

Figure 10

Figure 11

Subjects were faster overall in responding to CV comparisons than to VC comparisons (490 vs. 430 ms). This fact might be entirely due to the fact that subjects heard the consonants earlier in the CV stimuli than in the VC stimuli. More interesting are the differences by Fricative Pair. The pattern of responses in VC and CV position seem very similar to each other (and rather different from the error results already seen): comparisons involving the sibilant [ʃʲ] seem to have the longest RTs. (Note the second, fourth, and sixth box in each graph of Figure 10.) This observation was tested with paired *t*-tests comparing each of the pairs involving [ʃʲ] with all other pairs (twelve comparisons in all for each Position). The results are given in Tables III a-b.

a. VC

Fricative pair	s ɕ	s ʂ	ɕ ʂ	ʂ ʃʲ	ʃʲ ɕ
s ʃʲ	.099 (.208) *t*(12) = 1.71 *p* = .113	.143 (.280) *t*(12) = 1.85 *p* = .089	.133 (.205) *t*(12) = 2.34 *p* = .037	-.315 (.399) *t*(12) = -2.85 *p* = .015	-.163 (.238) *t*(12) = -2.47 *p* = .030
ɕ ʃʲ	.262 (.239) *t*(12) = 3.94 *p* = .002	.306 (.258) *t*(12) = 4.28 *p* = .001	.296 (.245) *t*(12) = 4.36 *p* = .001	-.153 (.424) *t*(12) = -1.30 *p* = .219	
ʃʲ ʂ	.414 (.352) *t*(12) = 4.24 *p* < .001	.459 (.453) *t*(12) = 3.65 *p* = .003	.449 (.341) *t*(12) = 4.75 *p* < .001		

b. CV

Fricative pair	s ɕ	s ʂ	ɕ ʂ	ʃʲ ʂ	ɕ ʃʲ
s ʃʲ	.145 (261) *t*(12) = 2.00 *p* = .069	.028 (.468) *t*(12) = 1.60 *p* = .136	.170 (.342) *t*(12) = 1.79 *p* = .098	.020 (.286) *t*(12) = .25 *p* = .805	-.043 (.313) *t*(12) = -.50 *p* = .627
ɕ ʃʲ	.188 (.336) *t*(12) = 2.02 *p* = .067	.251 (.287) *t*(12) = 3.15 *p* = .008	.214 (.235) *t*(12) = 3.28 *p* = .007	.063 (.298) *t*(12) = .76 *p* = .459	
ʃʲ ʂ	.125 (.266) *t*(12) = 1.70 *p* = .116	.188 (.360) *t*(12) = 1.88 *p* = .084	.150 (.267) *t*(12) = 2.03 *p* = .065		

Table III: Results of selected *t*-test comparisons. Top line in each cell represents the mean and standard deviation for differences between log RTs of the fricative pairs in rows and those in columns. Shaded cells indicate differences that are significant or marginally so (assuming a Bonferroni-adjusted significance threshold of .05/12 = .004). VC position is shown in (a) and CV in (b).

In the case of VC position specifically: i) the RT for [as] vs. [aʃʲ] did not differ significantly from that for any other fricative pair; ii) [aɕ] vs. [aʃʲ] and [aʃʲ] vs. [aʂ] (the two highest bars for VC position in Figure 10) each had a significantly greater RT than any remaining fricative pair except for [as] vs. [aʃʲ] (549 and 651 ms respectively vs. 451, 441, 438 ms). The high RT for [aʃʲ] vs. [aʂ] might be expected given the especially poor error rate seen for this fricative pair for the Polish speakers (see Figure 6). However, the higher RT for [aɕ] vs. [aʃʲ] has no analogue in the error data.

In the case of CV position, only two of the comparisons approached significance: the longer RT for [ɕa] vs. [ʃʲa] (474 ms) compared to [sa] vs. [ʂa] (394 ms) and [ɕa] vs. [ʂa] (404 ms).

These comparisons between Fricative Pair RTs may be easier to appreciate by means of the diagram in Figure 12. The general conclusion to be drawn is that the fricative pairs [ɕ] vs. [ʃʲ] and [ʃʲ] vs. [ʂ] caused particular difficulty for subjects overall. In VC position the highest average RT was induced by one of these two fricative pairs for ten of the thirteen subjects (all but 6, 9 and 10, see Table 3a in Appendix II). For two subjects (9 and 10) [as] vs. [aʃʲ] had the longest RT. In CV position there is more variability, but one or both of the pairs [ɕa] vs. [ʃʲa] and [ʃʲa] vs. [ʂa] induced the longest average RTs for four subjects (1,2,6,7) and the second longest for five more (4,5,8,12,13). The pair [sa] vs. [ʃʲa] caused the longest or second-longest RT for six subjects (1,3,4,5,9,10).

Figure 12

4.2.2 English subjects

Figure 13 presents reaction time data for the English subjects. Mean RT (overall and by speaker) are summarized in Table 3b in Appendix II. The main effects of Position and Fricative Pair were both significant for these speakers ($F(1,9) = 12.96$; $p < .01$ and $F(2.80, 25.23) = 13.00$; $p < .001$ respectively, the latter Greenhouse-Geisser adjusted). Once again the interaction was not significant ($F(5,45) = 1.12$; $p = .364$).

Figure 13

Subjects were faster responding to CV pairs (815 vs. 924 ms), just as the Polish subjects were, though as already noted this can be viewed as an artifact of the experiment design. Once again the differences by Fricative Pair are of most interest. Though English speakers had notably slower RTs than Polish speakers overall, there is some similarity between these RT results and those for the Polish speakers, since comparisons involving [ʃʲ] give the worst RTs in most cases for both VC and CV position. This observation was tested in a manner similar to that seen above for Polish subjects, with detailed results given in Table IV.

a. VC

Fricative pair	s ɕ	s s̢	ɕ s̢	ʃʲ s̢	ɕ ʃʲ
s ʃʲ	.112 (.229) *t*(9) = 1.55 *p* = .156	.158 (.301) *t*(9) = 1.66 *p* = .131	.016 (.205) *t*(9) = .24 *p* = .815	-.068 (.232) *t*(9) = -.92 *p* = .380	-.072 (.328) *t*(9) = -.69 *p* = .506
ɕ ʃʲ	.184 (.243) *t*(9) = 2.39 *p* = .040	.230 (.157) *t*(9) = 4.62 *p* = .001	.088 (.297) *t*(9) = .93 *p* = .376	.004 (.178) *t*(9) = .08 *p* = .942	
ʃʲ s̢	.180 (.223) *t*(9) = 2.55 *p* = .031	.226 (.166) *t*(9) = 4.31 *p* = .002	.083 (.231) *t*(9) = 1.14 *p* = .284		

▯

b. CV

Fricative pair	s ɕ	s ʃʲ	s s̢	ɕ s̢	ʃʲ s̢
ɕ ʃʲ	.260 (.311) *t*(9) = 2.65 *p* = .027	.322 (.231) *t*(9) = 4.40 *p* = .002	.498 (.306) *t*(9) = 5.15 *p* = .001	.245 (.199) *t*(9) = 3.89 *p* = .004	.137 (.197) *t*(9) = 2.21 *p* = .054
ʃʲ s̢	.123 (.334) *t*(9) = 1.16 *p* = .274	.184 (.273) *t*(9) = 2.13 *p* = .062	.361 (.393) *t*(9) = 2.90 *p* = .018	.107 (.280) *t*(9) = 1.21 *p* = .256	

Table IV: Results of selected *t*-test comparisons. Top line in each cell represents the mean and standard deviation for differences between log RT for each pair of fricative pairs. Shaded cells indicate differences that are significant (assuming a Bonferroni-adjusted significance threshold of .05/12 = .004 for VC and .05/9 = .006 for CV). VC position is shown in (a) and CV in (b).

For VC position, RTs for [aɕ] vs. [aʃʲ] and [aʃʲ] vs. [as̢] were both significantly greater than that for [as] vs. [as̢] (981 and 987 ms respectively vs. 813 ms). No other differences (among those tested) were significant. For CV position only the pairs [ɕa] vs. [ʃʲa] and [s̢a] vs. [ʃʲa] were tested, against all other pairs, since the pair [sa] vs. [ʃʲa] did not show a long RT overall. The pair [ɕa] vs. [ʃʲa] had a significantly greater RT than did [sa] vs. [ʃʲa], [sa] vs. [s̺a], or [ɕa] vs. [s̺a] (1017 ms vs. 801, 657, 812 ms respectively). These relationships are summarized in Figure 14.

Figure 14

As the Polish RT results did, these results single out the pairs [ɕ] vs. [ʃʲ] and [ʃʲ] vs. [s̺] as particularly difficult for subjects overall. For VC position, one or both of these pairs induced the longest RTs for five out of ten subjects (subjects 1,2,5,8,10). Still, for four subjects (3,4,6,7) [aɕ] vs. [as̺] caused the longest RT, and for subjects 3 and 9 the RT for [aɕ] vs. [aʃʲ] seems particularly low. For CV position the results are more consistent: the pair [ɕa] vs. [ʃʲa] had the longest RT for five out of the ten subjects (1,2,5,7,9), and it was among the top three highest RTs for every subject. The lack of significant effects involving [ʃʲa] vs. [s̺a] (in spite of appearances in Figure 13) is reflected in the individual subjects, who treated this pair very inconsistently. The same is true of the pair [sa] vs. [ʃʲa] (for which the overall results did not suggest particular difficulty).

Summing up so far: first, results from the error data differ by language in a way that would be expected, given the well known finding that language experience affects perception. Polish speakers perform near the ceiling on all pairs except for the VC pair [as̺] vs. [aʃʲ] (with [aʃʲ] an illicit phrase-final sequence). In contrast, English speakers have difficulty discriminating among the postalveolars. In addition there was evidence from the English error data that subjects have particular difficulty discriminating CV [ɕa] vs. [ʃʲa]. Second, the RT results for the Polish- and English-speaking subjects differ from the error results in being rather similar to each other. Specifically, they provide evidence that the fricative pair [ɕ] vs. [ʃʲ] causes relatively more perceptual difficulty than other fricative pairs in CV position; in VC position this same pair and [ʃʲ] vs. [s̺] cause relatively more difficulty.

5. Multidimensional scaling

In order to further explore these results and their relation to the hypothesis about perceptual distance and sound change, multidimensional scaling (MDS) analyses were carried out. MDS is a set of techniques which, given an input of (dis)similarities (distances) between pairs of points, derives a ‘map’ of those points in space (of some dimensionality). The algorithm seeks an arrangement of the points that reproduces the input distances to the best degree possible. One common measure of this goodness-of-fit, ‘stress’, depends on the sum of the squared deviations of the input distances from the respective derived distances. The user attempts to interpret the output ‘map’ by associating dimensions of that map to known or hypothesized underlying perceptual dimensions. In speech research MDS can be used in the following way: measures of perceived similarity, such as d-prime scores or reaction times, can serve as input ‘distances’; the user can then attempt to interpret the output by means of acoustic or perceptual dimensions (e.g., F2, COG). What MDS contributes beyond the d-prime and RT results already seen is a way of *integrating* all of the information into one perceptual ‘map’. Works in phonetics employing MDS include Pols, van der Kamp, and Plomp (1969), Gandour and Harshman (1978), Jassem and Lobacz (1995), Iverson and Kuhl (1995), Lambacher, Martens, Nelson, and Berman (2001), and Huang (2001). (Cf. the phonological notion 'P-map' of Steriade, 2001.) An overview of the technique in phonetics can be found in Johnson (2003).

For input distances the reciprocals of the mean (log transformed) RTs were employed (Shepard, Kilpatric, & Cunningham, 1975). The use of RTs rests on the idea that, when listeners take longer to react to a pair of different stimuli such as [sa] vs. [ɕa] – even when they *correctly* find them to be different – the longer RT indicates more perceptual difficulty (Takane & Sergent, 1983). In terms of the spacial metaphor, greater difficulty implies closer sounds on the perceptual map (hence the use of the reciprocal of the RT for distances.)

Separate analyses were done for each language group and position (VC vs. CV), and the results are given in Figure 15. These analyses were done using SPSS 15 for Windows.[14] (For the meaning of the added arrows in the diagrams see the next section.)

Figure 15

---

[14] The analyses assumed square symmetric matrices, matrix conditionality, a ratio level of measurement, and a euclidean distance scaling model.

It should be emphasized that the orientation of each output map is arbitrary. If one had begun with distances between known cities (a commonly used example), the output would show the relative positions of the cities, but with no regard for one's desire that north be 'up', or even that north correspond to either of the labeled dimensions '1' or '2' (which are therefore ignored). Thus plot (b) seems to be flipped along a diagonal compared to plot (a), and plot (c) seems rotated 45 degrees counterclockwise compared to (d). The point is that one can nevertheless infer the same groupings for each plot. Specifically, in each plot we find a grouping of [s] and [ɕ] versus [ʃʲ] and [ʂ] along one dimension, and a grouping of [s] and [ʂ] versus [ɕ] and [ʃʲ] along the other. These dimensions can be plausibly attributed to the acoustic parameters of COG (center of gravity) and F2 respectively, given the acoustic properties of these sounds described in section 3.4.[15] That is, [s] and [ɕ] group together as having relatively high COG values while those of [ʂ] and [ʃʲ] are comparatively low. At the same time, [ɕ] and [ʃʲ] group together as having relatively high F2 transitions while those of [s] and [ʂ] are comparatively low. (Since [ʃʲ] has a high F2 only at offset, one might expect to find [ʃʲ] closer to [ʂ] when in VC position. This seems to be the case, though the differences are not as large as we would expect.)

Figure 15 also shows two measures of goodness of fit for these results, Stress and RSQ; these are averaged across individual subjects for each group. These values indicate the the fit is poor for all four maps, with the input distance values accounting for only 14-36% of the scaled results according to RSQ. This is probably due to the considerable variation between subjects in mean RTs for the same fricative pairs. The individual fits are poor. However, first, the results have plausible interpretations in terms of the known acoustic properties of these sibilants, as discussed above. Second, the derived groupings of sibilants hold independently for each of the four analyses, and they are not trivial: if the positions of [s] and [ɕ] were switched in any of the plots, for example, there would be no obvious interpretation of the dimension understood here as F2 based on acoustic criteria. These results are therefore of interest.

---

[15] For plot (c) we find these groupings by drawing the COG and F2 axes at a 45-degree tilt.

5. Discussion

5.1. COG, F2, and phonetic similarity

The error data results revealed some unsurprising differences between the Polish- and English-speaking subjects. Polish subjects performed near the ceiling on all fricative pairs except for [ʃʲ] vs. [ʂ] in VC position. (As noted earlier, [ʃʲ] does not occur phrase-finally in Polish, and in this position [ʃʲ] and [ʂ] are very similar in both COG and formant transitions.) English-speaking subjects did well discriminating [s] from postalveolar fricatives but more poorly in distinguishing among the latter, as would be expected given their language experience.

However, the interest here lies more with results that might be argued to be independent of language experience. This is because the focus is not on how phonology influences perception but rather on the possibility of the opposite relation: does perception influence the shape of phonological systems? To answer this question requires, first, that one understand the perceptual relations of sounds, abstracting away from language experience as much as possible. The reaction time results are encouraging from this point of view. In contrast to what was found with the error results, the RT results reveal a rather striking similarity across the two languages. Furthermore the patterning of the RT results is qualitatively distinct from the patterning of the error results of *either* Polish or English. Though we do not claim that the RT results are unaffected by experience with the respective languages, we take these results to support a tentative conclusion that the RT results do reveal *some* perceptual properties of these sibilants that are independent of the particular languages, as suggested by Johnson and Babel (to appear) and Babel and Johnson (to appear). These works argue that the degree of independence from language may increase with faster RTs, particularly with RTs in the neighborhood of 500 ms, and perhaps to a lesser extent with RTs up to 800 ms. Our Polish subjects, who have the category contrasts we are trying to abstract away from, had RTs under 800 ms. These results may therefore legitimately bear on the question of how perception could influence phonology.

Figure 16 presents a crude picture of the acoustic properties of these sibilants relative to each other, based on the acoustic analysis of the stimuli in section 3.4. They can be ordered along the y-axis as shown according to their COG values. Along the dimension of F2 [s] and [ʂ] are similar in having low F2 values, while [ɕ] has high F2 transitions. In Polish, the sound [ʃʲ] has a rather low *onset* F2 value and a high *offset* value. This dynamic character is represented by means of the arrow.

Figure 16

This diagram bears some obvious similarity to the MDS results. Based on this diagram, and assuming that these acoustic properties have important perceptual correlates, one might predict confusion particularly between [ɕ] and [ʃʲ] on the one hand and between [ʃʲ] and [ʂ] on the other: only these pairs involve sibilants that both lie adjacent in COG and share at least one F2 transition type (high or low). These are indeed the pairs that were found to present the most difficulty according to the RTs of both Polish- and English-speaking subjects. More specifically, one might expect the greatest difficulty between [ɕ] and [ʃʲ] in CV position, since these sounds have similar onset F2 values, and the greatest difficulty between [ʃʲ] vs.[ʂ] and in VC position, since this pair have similar offset F2 values. The RT results reviewed above are consistent with this expectation (as are indeed the more particular error results for each language group). However, for both subject groups evidence was also found that [ɕ] and [ʃʲ] are highly confusable even in VC position. This is less expected given their onset F2 profiles, and it might suggest that these sounds are more similar in COG (or other relevant spectral properties) than this crude diagram allows.

These results thus support the conclusion that both fricative noise and F2 transitions are important for the perception of these sibilants, a conclusion that Nowak (2006) is also led to. The multidimensional scaling results similarly support this conclusion. Indeed, the MDS plots suggest that COG and F2 are roughly equally important to distinguishing among these sounds for both Polish- and English-speaking subjects.

5.2 Perception, sound change, and phonology

Turning now to the phonological issues raised at the outset, recall first the phoneme inventory puzzle: though [ʃ] seems to be the prefered postalveolar in general, there are languages that have [ɕ] and [ʂ] while lacking [ʃ]. We hypothesized (following Żygis 2006) that a language with a contrast between postalveolars might prefer [ɕ] and [ʂ] over [ɕ] and [ʃ] for reasons of perceptual distinctiveness. The results here support this hypothesis.[16] The MDS ‘perceptual maps’ in particular highlight the perceptual distinctiveness of [ɕ] and [ʂ] compared to either other postalveolar contrast. Table V presents distances between fricative pairs broken down by language group and position, based on the MDS results. The figures are euclidean distances based on the coordinates provided by the MDS algorithm. As the plots themselves suggest, for each of the four analyses the pairs [s] vs. [ʃʲ] and [ɕ] vs. [ʂ] (‘opposite corner fricatives’) are perceptually most distinct. Also for each of the analyses, the pairs [ɕ] vs. [ʃʲ] and [ʃʲ] vs. [ʂ] are the most confusable.

The results here also bear on the more specific question whether the Polish historical sound change [ʃʲ] > [ʂ] might have been motivated by perceptual distinctiveness. It is obviously a leap to assume that the sounds of Modern Polish analyzed here are the same sounds involved in the 13th century sound change discussed earlier. On this assumption, however, the arrows in the MDS plots indicate the effect of this sound change. Again each of the four analyses (two languages × two positions) supports the hypothesis: the summed perceptual distances among the three sibilants improves in every case assuming this change. (See (c) in Table V.)

---

[16] Our experiment included Polish [ʃʲ] instead of [ʃ] for reasons discussed in section 3.2. As noted in footnote 7, these sounds are very similar in COG. Polish [ʃʲ] seems to have lower onset F2 transitions and higher offset ones than a typical [ʃ].

a. Polish

	s	ɕ	ʃʲ	ʂ
s		2.20	2.78	2.18
ɕ	2.12		1.85	2.85
ʃʲ	2.79	1.81		1.77
ʂ	2.15	2.86	1.92	

b. English

	s	ɕ	ʃʲ	ʂ
s		2.13	2.82	2.13
ɕ	2.07		1.85	2.82
ʃʲ	2.83	1.72		1.97
ʂ	2.24	2.80	1.98	

c. Change in *overall* perceptual distance in Polish before and after sound change based on these figures: Polish VC 6.83 → 7.23; Polish CV 6.72 → 7.13; English VC 6.80 → 7.08; English CV 6.62 → 7.11.

Table V: Euclidean distances between fricative pairs derived from the coordinates provided by the MDS analyses: Polish (a) and English (b). For each language, the upper-right portion represents VC position and the lower-left CV position.

What do these results mean for phonological theory? Consider first Hall's (1997) account of sibilant retroflexion based on distinctive feature theory. According to Hall, the postalveolar sibilants [ɕ] and [ʃ] have identical feature specifications: [Coronal, −anterior, +distributed]. (The specification for [ʂ] is [Coronal, −anterior, −distributed].) This assumption immediately entails that [ɕ] and [ʃ] cannot occur in the same inventory; in phonological terms they are versions of the same sound. It also provides Hall with a means to explain retroflexion. As this paper does, Hall links this sound change to the earlier appearance of an alveolopalatal series: since the pre-existing palatoalveolars and the new alveolo-palatals cannot contrast and both be specified [−anterior, +distributed], the palatoalveolars shifted to [−distributed].

One problem for this approach is the fact that [ɕ] and [ʃ] did contrast phonemically in Old Polish. Though retroflexion of palatoalveolars did follow creation of the alveolo-palatals (from palatalized dentals), it did not follow immediately; in fact, the alveolo-palatals and palatoalveolars coexisted for as long as 200 years. (See the references cited in section 1.) As seen earlier, the two categories (with [ʃʲ] now palatalized) are furthermore distinct allophones (and marginally contrastive), in Modern Polish, and Polish speakers are able to distinguish them.

Though the account for retroflexion here is similar to Hall's, it rests on phonetic *similarity* rather than phonological (featural) *identity*. Palatoalveolars and alveolopalatals can coexist, even as independent phonemes; but such a contrast is *dispreferred*. Phoneme inventories having [ɕ, ʂ] but no [ʃ⁽ʲ⁾] (as in Modern Polish or Mandarin) exist – even though palatoalveolars are the preferred postalveolar by standard criteria – as a result of the functional pressure of dispersion. Retroflexion of palatoalveolars, as in Polish (and e.g., Sanskrit), occurs due to the same pressure, which is why it occurs (to our knowledge) only when a language already includes an alveolo-palatal series. In approaching the facts this way this paper is appealing to well known ideas of Lindblom (Liljencrants and Lindblom 1972; Lindblom 1986, 1990).[17] As for phonological features, this paper is influenced by recent work even in phonology minimizing the role of features in explaining cross-linguistic patternings of contrast, arguing that those patterns are better explained by principles of dispersion (e.g., Flemming, 2004; Padgett, 2001). Padgett and Żygis (2007) present an account of Polish retroflexion within Optimality Theory in such a framework.

Though this application of dispersion principles to the sibilant facts seems promising, it raises important questions that are difficult to answer. If a contrast between [ɕ] and [ʃʲ] is perceptually disfavored, why was it created in the first place? Why did palatalized dentals shift to alvelo-palatals when palatalized palatoalveolars were already present in the system? Why did the resulting system last for generations before retroflexion occurred? Why was the resulting system avoided by retroflexion, a shift of palatoalveolars to retroflexes, rather than to (say) a shift to lateral or velar fricatives? We think that answers to such questions are likely to depend on acknowledging that local perceptual pressures of the sort suggested here are constantly competing with other perceptual pressures, and with pressures having sources in articulation, lexical representation, learning, and social factors, among other things.

6. Conclusion

The results here support the general view that perceptual constraints play a role in explaining sound change tendencies, and therefore in shaping phonological tendencies. Since an

---

[17] To acknowledge the importance of functional pressures in phonology or phonetics is not necessarily to imply that speakers act in a goal-oriented way (though we don't believe this should be ruled out). See Wedel (2004) and Boersma and Hamann (2008) for accounts of dispersion through self-organization.

important hypothesis here was about how perception might shape phonology and not vice versa, the perceptual study had a design intended to reduce or ‘skirt’ the effect of native language on a subject’s performance: the task was relatively easy in providing non-degraded stimuli and keeping the comparisons to just two fricatives within any block. Though the results look different by language based on the incorrect response data, with English speakers showing the expected difficulties due to having fewer categories, the results are in fact more similar based on reaction time data.

Based on reaction time data, the results here suggest that both noise spectra and formants contribute significantly to the perception of the Polish contrasts, with dentals and alveolo-palatals grouping against palatoalveolars and retroflexes on the one hand (high noise 'tonality' vs. low), and alveolo-palatals and palatoalveolars grouping against dentals and retroflexes on the other (high F2 vs. low). The results also provide support for the account of the sound change discussed above. The contrast [ɕ] vs. [ʃʲ] was particularly difficult to discriminate for both Polish and English speakers, in both CV and VC syllables.

Acknowledgments

We are grateful to Gerry Docherty and three anonymous reviewers for extensive comments on this article. We would also like to thank Tracy A. Hall, Junko Ito, Keith Johnson, Shigeto Kawahara, Grant McGuire, Armin Mester, and Nathan Sanders for help and feedback. This research has been supported by the German Research Foundation (GWZ-4/11-1-P2), the Federal Ministry of Education and Research (01UG0711), and faculty research funds granted by the University of California, Santa Cruz.

References

Baayen, H. R. (2008). *Practical data analysis for the language sciences with R*.

Babel, M., & Johnson, K. (to appear). Accessing psycho-acoustic perception and language-specific perception with speech sounds. *Laboratory phonology*.

Boersma, P., & Hamann, S. (2008). *The evolution of auditory dispersion in bidirectional constraint grammars*: Ms., University of Amsterdam and Düsseldorf University.

Boersma, P., & Weenink, D. (2007). *Praat: doing phonetics by computer*: Computer program, http://www.praat.org/.

Dogil, G. (1990). *Hissing and hushing fricatives: a comment on non-anterior spirants in Polish*: Ms., Stuttgart University.

Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner & D. Steriade (Eds.), *Phonetically-based phonology* (pp. 232-276). Cambridge: Cambridge University Press.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. *Journal of Acoustical Society of America, 84.1*, 115-123.

Gandour, J., & Harshman, R. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech, 21*, 1-33.

Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association, 32*, 141-174.

Hall, T. A. (1997). The historical development of retroflex consonants in Indo-Aryan. *Lingua, 102*, 203-221.

Halle, M., & Stevens, K. N. (1997). The postalveolar fricatives of Polish. In S. Kiritani, H. Hirose & H. Fujisaki (Eds.), *Speech production and language* (pp. 177-193). Berlin: Mouton de Gruyter.

Huang, T. (2001). The interplay of perception and phonology in tone 3 sandhi in Chinese Putonghua. In *OSU working papers in linguistics 55* (pp. 23-42).

Iverson, P., & Kuhl, P. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of Acoustical Society of America, 97.1*.

Jassem, W. (1962). Noise spectra of Swedish, English, and Polish fricatives. In *Proceedings of the speech communication seminar 1-4*. Stockholm: Royal Institute of Technology Speech Transmission Laboratory.

Jassem, W. (1979). Classification of fricative spectra using statistical discriminant functions. In B. Lindblom & S. Öhman (Eds.), *Frontiers of speech communication research* (pp. 77-91). New York: Academic Press.

Jassem, W., & Lobacz, P. (1995). Multidimensional scaling and its applications in a perceptual analysis of Polish consonants. *Journal of quantitative linguistics, 2.2*, 105-124.

Johnson, K. (2003). *Acoustic and auditory phonetics*. Cambridge, MA: Blackwell.

Johnson, K. (2008). *Quantitative methods in linguistics*. Oxford: Blackwell.

Johnson, K., & Babel, M. (to appear). Perception of fricatives by Dutch and English speakers. *Journal of phonetics*.

Jongman, A. (1989). Duration of frication noise required for identification of English fricatives. *Journal of the Acoustical Society of America, 85*, 1718-1725.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of Acoustical Society of America, 108.3*, 1252-1263.

Kochetov, A., & Lobanova, A. (2007). Komi-Permyak coronal obstruents: acoustic contrasts and positional variation. *Journal of the International Phonetic Association, 37*, 51-82.

Kudela, K. (1968). Spectral analysis of Polish fricative consonants. In W. Jassem (Ed.), *Speech analysis and synthesis* (pp. 93-188). Warsaw.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Cambridge, MA: Blackwell.

Lambacher, S., Martens, W., Nelson, B., & Berman, J. (2001). Identification of English voiceless fricatives by Japanese listeners: the influence of vowel context on sensitivity and response bias. *Acoustical science and technology, 22.5*, 334-343.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language, 48.4*, 839-862.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology* (pp. 13-44). Orlando: Academic Press.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403-439). Dordrecht: Kluwer.

Lisker, L. (2001). Hearing the Polish sibilants [s ś s]: phonetic and auditory judments. In N. Grønnum & J. Rischel (Eds.), *To honour Eli Fischer-Jørgensen. Travaux du cercle linguistique de Copenhague XXXI* (pp. 226-238). Copenhagen: Reitzel.

Macmillan, N. A., & Creelman, C. D. (1991). *Signal detection: a user's guide*. Cambridge: Cambridge University Press.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

Maddieson, I., & Precoda, K. (1992). *UPSID and phoneme. UCLA phonological segment inventory database version 1.1*.

Mann, V., & Soli, S. D. (1991). Perceptual order and the effect of vocalic context on fricative perception. *Perception and psychophysics, 49.5*, 399-411.

McGuire, G. (2007). *Phonetic category learning*. Berkeley, CA: Ph.D. dissertation, UC Berkeley.

McGuire, G. (2008). *Selective attention and English listeners' perceptual learning of the Polish post-alveolar sibilant contrast*. Santa Cruz, CA: Ms., UC Santa Cruz.

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of speech and hearing research, 32*, 120-132.

Nowak, P. M. (2006). The role of vowel transition and frication noise in the perception of Polish sibilants. *Journal of phonetics, 34*, 139-152.

Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick & M. F. Miller (Eds.), *Chicago Linguistic Society: papers from the parasession on language and behavior* (pp. 178-203). Chicago, IL: Chicago Linguistic Society.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. MacNeilage (Ed.), *The Production of Speech* (pp. 189-216). New York: Springer-Verlag.

Padgett, J. (2001). Contrast dispersion and Russian palatalization. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 187-218). San Diego, CA: Academic Press.

Padgett, J., & Żygis, M. (2007). The evolution of sibilants in Polish and Russian. *Journal of Slavic linguistics, 15.2*, 291-324.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and psychophysics, 13*, 253-260.

Pols, L. C. W., van der Kamp, L. J. T., & Plomp, R. (1969). Perceptual and physical space of vowel sounds. *Journal of the Acoustical Society of America, 46:2B*, 458-467.

Repp, B. H. (1981). Two strategies in fricative discrimination. *Perception and psychophysics, 30.3*, 217-227.

Rospond, S. (1971). *Gramatyka historyczna jezyka polskiego*. Warszawa: Panstwowe Wydawnictwo Naukowe.

Rubach, J. (1984). *Cyclic and Lexical Phonology: The Structure of Polish*. Dordrecht: Foris.

Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997). The dispersion-focalization theory of vowel systems. *Journal of phonetics, 25*, 255-286.

Shepard, R. N., Kilpatric, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive psychology, 7*, 82-138.

Steriade, D. (2001). Directional asymmetries in place assimilation: a perceptual account. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 219-250). New York: Academic Press.

Stevens, K. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), *Human Communication: A Unified View* (pp. 51-66). New York: McGraw Hill.

Stevens, K., & Keyser, S. J. (1989). Primary features and their enhancements in consonants.

Language, 65, 81-106.

Stieber, Z. (1962). *Rozwoj fonologiczny jezyka polskiego*. Warszawa: Panstwowe Wydawnictwo Naukowe.

Stieber, Z. (1968). *The phonological development of Polish*. Ann Arbor: University of Michigan, Department of Slavic Languages and Literatures.

Takane, Y., & Sergent, J. (1983). Multidimensional scaling models for reaction times and same-different judgements. *Psychometrika, 48.3*, 393-423.

Wedel, A. (2004). Lexical category competition and contrast maintenance. In *Proceedings of the Special Interest Group in Phonology, annual meeting of the Association of Computational Linguistics*. Barcelona: Ms., University of Arizona.

Whalen, D. H. (1991). Perception of the English /s/ - /S/ distinction relies on fricative noises and transitions, not on brief spectral slices. *Journal of Acoustical Society of America, 90.4*, 1776-1785.

Wierzchowska, B. (1971). *Wymowa polska*. Warszawa: Pan'stwowe Zaklady Wydawnictw Szkolnych.

Żygis, M. (2003). The role of perception in Slavic sibilant systems. In P. Kosta, J. Blaszczak, J. Frasek, L. Geist & M. Zygis (Eds.), *Investigations into formal Slavic linguistics* (pp. 137-154). Berlin: Peter Lang Verlag.

Żygis, M. (2006). *Contrast optimisation in Slavic sibilant systems*: Habilitation, Humboldt University, Berlin.

Żygis, M., & Hamann, S. (2003). Perceptual and acoustic cues of Polish coronal fricatives. In *Proceedings of the 15th International Congress of Phonetic Sciences, August 3-9, Barcelona* (pp. 395-398).

Appendix I

Acoustics of stimuli

1. Duration

1a. CV *F* (3,27) = 1.383

sibilant	average duration (s)	s	ɕ	ʃʲ	ʂ
s	0.243	–	n.s	n.s.	n.s.
ɕ	0.239	n.s.	–	n.s.	n.s.
ʃʲ	0.254	n.s.	n.s.	–	n.s.
ʂ	0.263	n.s	n.s.	n.s.	–

1b. VC *F* (3,27) = 14.966

sibilant	average duration (s)	s	ɕ	ʃʲ	ʂ
s	0.336	–	n.s.	$p<.001$	n.s.
ɕ	0.323	n.s.	–	$p<.01$	n.s.
ʃʲ	0.247	$p<.001$	$p<.01$	–	$p<.001$
ʂ	0.352	n.s.	n.s.	$p<.001$	–

2. Center of gravity

CV

2a. Mean COG values (Hz)

sibilant	COG1	COG2	COG3	mean COG
s	8270	8780	8488	8512
ɕ	4728	5320	4966	5004
ʃʲ	3080	3738	3810	3542
ʂ	2683	2569	2605	2619

2b. COG within individual sibilants

sibilant	COG1	COG2	COG3	
s COG1	_	n.s.	n.s.	$F(2,20)=1.512$
COG2	n.s.	_	n.s.	
COG3	n.s.	n.s.	_	
ɕ COG1	_	$p<.05$	n.s.	$F(2,20)=7.932$
COG2	$p<.05$	_	n.s.	
COG3	n.s.	n.s.	_	
ʃʲ COG1	_	$p<.05$	$p<.05$	$F(2,20)=10.114$
COG2	$p<.05$	_	n.s.	
COG3	$p<.05$	n.s.	_	
ʂ COG1	_	n.s.	n.s.	$F(2,20)=.485$
COG2	n.s.	_	n.s	
COG3	n.s.	n.s.	_	

□

2c. COG 1 $F(3,27)=278.848$

COG1	s	ɕ	ʃʲ	ʂ
s	_	$p<.001$	$p<.001$	$p<.001$
ɕ	$p<.001$	_	$p<.001$	$p<.001$
ʃʲ	$p<.001$	$p<.001$	_	n.s.
ʂ	$p<.001$	$p<.001$	n.s.	_

2d. COG 2 $F(3,27)=381.859$

COG2	s	ɕ	ʃʲ	ʂ
s	_	$p<.001$	$p<.001$	$p<.001$
ɕ	$p<.001$	_	$p<.001$	$p<.001$
ʃʲ	$p<.001$	$p<.001$	_	$p<.001$
ʂ	$p<.001$	$p<.001$	$p<.001$	_

2e. COG 3 $F(3,27) = 410.307$

COG3	s	ɕ	ʃʲ	ʂ
s		$p<.001$	$p<.001$	$p<.001$
ɕ	$p<.001$		$p<.001$	$p<.001$
ʃʲ	$p<.001$	$p<.001$		$p<.001$
ʂ	$p<.001$	$p<.001$	$p<.001$	

2f. Mean COG $F(3,27)=490.444$

Mean COG	s	ɕ	ʃʲ	ʂ
s		$p<.001$	$p<.001$	$p<.001$
ɕ	$p<.001$		$p<.001$	$p<.001$
ʃʲ	$p<.001$	$p<.001$		$p<.001$
ʂ	$p<.001$	$p<.001$	$p<.001$	

VC

2g. Mean COG values (Hz)

sibilant	COG1	COG2	COG3	average COG
s	8290	8683	8344	8438
ɕ	4898	5064	4606	4856
ʃʲ	3004	3343	3374	3240
ʂ	2801	2883	2723	2803

2h. COG within individual sibilants

sibilant	COG1	COG2	COG3	
s COG1	–	n.s.	n.s.	$F(2,20)=1.246$
COG2	n.s.	–	n.s.	
COG3	n.s.	n.s.	–	
ɕ COG1	–	n.s.	n.s.	$F(2,20)=5.715$
COG2	n.s.	–	$p<.05$	
COG3	n.s.	$p<.05$	–	
ʃʲ COG1	–	n.s.	$p<.05$	$F(2,20)=4.667$
COG2	n.s.	–	n.s.	
COG3	$p<.05$	n.s.	–	
ʂ COG1	–	n.s.	n.s.	$F(2,20)=.138$
COG2	n.s.	–	n.s.	
COG3	n.s.	n.s.	–	

2i. COG 1 $F(3,27) = 825.230$

COG1	s	ɕ	ʃʲ	ʂ
s		$p<.001$	$p<.001$	$p<.001$
ɕ	$p<.001$		$p<.001$	$p<.001$
ʃʲ	$p<.001$	$p<.001$		n.s.
ʂ	$p<.001$	$p<.001$	n.s.	

2j. COG 2 $F(3,27) = 462.104$

COG2	s	ɕ	ʃʲ	ʂ
s		$p<.001$	$p<.001$	$p<.001$
ɕ	$p<.001$		$p<.001$	$p<.001$
ʃʲ	$p<.001$	$p<.001$		n.s.
ʂ	$p<.001$	$p<.001$	n.s.	

2k. COG 3 $F(3,27) = 307.827$

COG3	s	ɕ	ʃʲ	ʂ
s		$p<.001$	$p<.001$	$p<.001$
ɕ	$p<.001$		$p<.001$	$p<.001$
ʃʲ	$p<.001$	$p<.001$		$p<.05$
ʂ	$p<.001$	$p<.001$	$p<.05$	

▯

2l. Mean COG CV $F(3,27)=742.534$

Mean COG	s	ɕ	ʃʲ	ʂ
s		$p<.001$	$p<.001$	$p<.001$
ɕ	$p<.001$			$p<.001$
ʃʲ	$p<.001$	$p<.001$		$p<.05$
ʂ	$p<.001$	$p<.001$	$p<.05$	

3. Formant rate (Hz/ms)

3a. CV $F(3,27)=22.913$

	Average $F2_{rate}$	s	ɕ	ʃʲ	ʂ
s	0.724	–	$p<.01$	$p<.001$	n.s.
ɕ	4.17	$p<.01$	–	n.s.	$p.<.05$
ʃʲ	6.3	$p<.001$	n.s.	–	$p <.001$
ʂ	1.59	n.s.	$p.<.05$	$p<.001$	–

▯

3b. VC $F(3,27)=7.187$

	Average $F2_{rate}$	s	ɕ	ʃʲ	ʂ
s	3.267	–	n.s.	n.s.	n.s.
ɕ	4.803	n.s.	–	$p<.05$	$p<.01$
ʃʲ	2.635	n.s.	$p<.05$	–	n.s.
ʂ	1.929	n.s.	$p<.01$	n.s.	–

Appendix II

1a. D-prime scores by fricative pair, position, and subject

Polish speakers

Subject	VC						CV					
	s ɕ	s ʃʲ	s ʂ	ɕ ʃʲ	ɕ ʂ	ʃʲ ʂ	sɕ	s ʃʲ	s ʂ	ɕ ʃʲ	ɕ ʂ	ʃʲ ʂ
1	3.44	4.55	4.16	4.40	3.44	1.28	4.56	2.73	3.44	2.23	3.44	4.56
2	4.75	4.55	3.44	4.27	2.93	1.88	3.44	5.06	4.16	4.56	4.27	2.64
3	4.75	3.44	5.06	4.56	4.56	0.88	4.40	4.27	4.40	4.56	4.76	2.93
4	6.18	4.55	6.18	4.55	6.18	2.03	4.16	4.27	2.77	4.56	4.75	5.06
5	2.13	1.77	3.80	1.78	4.40	0.06	2.53	2.53	2.73	4.27	3.44	2.77
6	3.44	4.56	4.40	4.75	4.55	1.97	6.18	5.06	3.44	5.06	5.06	4.76
7	5.06	4.40	5.06	2.93	6.18	2.53	6.18	5.06	6.18	5.06	6.18	5.06
8	5.06	5.06	4.75	4.76	4.56	1.29	6.18	6.18	6.18	6.18	6.18	6.18
9	5.06	5.06	5.06	4.76	4.76	2.46	4.56	5.06	5.06	3.44	5.06	4.76
10	3.44	3.13	2.93	4.56	2.77	1.97	5.06	3.44	4.76	6.18	5.06	5.06
11	6.18	4.27	5.06	4.27	5.06	2.53	6.18	5.06	6.18	3.44	6.18	5.06
12	6.18	6.18	5.06	6.18	5.06	2.54	5.06	6.18	6.18	5.06	5.06	6.18
13	6.18	4.75	6.18	3.13	6.18	2.23	6.18	6.18	6.18	6.18	4.56	6.18
Overall	4.76	4.33	4.70	4.22	4.66	1.82	4.97	4.70	4.74	4.67	4.92	4.71

1b. Bias scores by fricative pair, position, and subject

Polish speakers

Subject	VC						CV					
	s ɕ	s ʃʲ	s ʂ	ɕ ʃʲ	ɕ ʂ	ʃʲ ʂ	s ɕ	s ʃʲ	s ʂ	ɕ ʃʲ	ɕ ʂ	ʃʲ ʂ
1	0.87	-3.71	4.21	-3.92	0.87	1.05	3.70	0.81	0.87	1.22	0.87	-3.70
2	-3.39	-3.71	0.87	-4.08	0.00	0.70	0.87	-2.83	-4.20	3.70	-4.08	-0.38
3	-3.39	0.87	-2.83	-3.70	-3.70	1.08	-3.92	-4.08	-3.92	-3.70	-3.38	0.00
4	0.00	-3.71	0.00	-3.71	0.00	0.91	-4.20	-4.08	-0.22	-3.70	-3.39	-2.83
5	0.00	-1.03	-4.52	-0.32	-3.92	0.03	0.50	-0.50	0.81	-4.08	0.87	-0.22
6	0.87	-3.70	-3.92	-3.39	-3.71	0.38	0.00	-2.83	0.87	-2.83	-2.83	-3.38
7	-2.83	-3.92	-2.83	-0.01	0.00	0.50	0.00	-2.83	0.00	-2.83	0.00	-2.83
8	-2.83	-2.83	-3.39	-3.38	-3.70	1.06	0.00	0.00	0.00	0.00	0.00	0.00
9	-2.83	-2.83	-2.83	-3.38	-3.38	1.08	3.70	-2.83	-2.83	0.87	-2.83	-3.38
10	0.87	0.31	-0.01	-3.70	-0.22	0.38	-2.83	0.87	-3.38	0.00	-2.83	-2.83
11	0.00	-4.08	-2.83	-4.08	-2.83	-0.50	0.00	-2.83	0.00	0.87	0.00	-2.83
12	0.00	0.00	-2.83	0.00	-2.83	1.78	-2.83	0.00	0.00	-2.83	-2.83	0.00
13	0.00	-3.39	0.00	0.31	0.00	1.22	0.00	0.00	0.00	0.00	3.70	0.00
Overall	-0.97	-2.44	-1.61	-2.57	-1.80	0.74	-0.39	-1.63	-0.92	-1.02	-1.29	-1.72

1c. D-prime scores by fricative pair, position, and subject

English speakers

Subject	VC						CV					
	s ɕ	s ʃʲ	s ʂ	ɕ ʃʲ	ɕ ʂ	ʃʲ ʂ	s ɕ	s ʃʲ	s ʂ	ɕ ʃʲ	ɕ ʂ	ʃʲ ʂ
1	4.27	4.56	4.56	2.77	2.77	2.74	4.27	4.06	4.40	0.89	4.56	4.27
2	5.06	2.93	3.44	2.03	4.16	1.66	2.93	4.16	4.16	2.25	2.53	6.18
3	5.06	4.55	4.75	2.77	2.93	1.00	3.44	4.75	5.06	2.03	2.46	2.26
4	6.18	3.13	5.06	4.75	3.04	2.03	2.93	4.76	4.76	2.53	2.93	4.76
5	3.13	2.03	2.64	2.26	1.76	1.36	2.53	2.53	2.65	2.10	2.26	2.53
6	6.18	5.06	5.06	5.06	3.04	2.23	6.18	5.06	5.06	3.04	5.06	4.76
7	4.55	6.18	6.18	4.55	2.25	3.66	4.55	5.06	6.18	2.10	4.55	5.06
8	4.75	4.16	4.06	1.06	2.26	1.29	4.16	4.55	4.75	1.16	1.94	4.27
9	3.44	4.40	4.75	1.75	2.26	2.53	6.18	4.55	6.18	4.55	6.18	2.93
10	5.06	5.06	4.75	4.76	4.56	2.93	4.55	6.18	6.18	2.93	4.55	4.56
Overall	4.77	4.21	4.53	3.18	2.90	2.14	4.17	4.57	4.94	2.36	3.70	4.16

1d. Bias scores by fricative pair, position, and subject

English speakers

Subject	VC						CV					
	s ɕ	s ʃʲ	s ʂ	ɕ ʃʲ	ɕ ʂ	ʃʲ ʂ	s ɕ	s ʃʲ	s ʂ	ɕ ʃʲ	ɕ ʂ	ʃʲ ʂ
1	-4.08	-3.70	-3.70	-0.22	-0.22	0.82	-4.08	-4.31	-3.92	0.55	-3.70	-4.08
2	-2.83	0.00	0.87	1.32	4.21	1.38	-0.01	4.21	4.21	0.75	0.50	0.00
3	-2.83	-3.71	-3.39	-0.22	0.00	0.68	0.87	-3.39	-2.83	0.91	1.08	0.76
4	0.00	0.31	-2.83	-3.39	1.37	0.91	0.00	-3.38	-3.38	0.50	0.00	-3.38
5	0.31	-0.10	-0.38	-0.76	0.15	-0.15	-0.50	-0.50	-0.38	0.54	0.76	0.50
6	0.00	-2.83	-2.83	-2.83	1.37	1.23	0.00	-2.83	-2.83	1.37	-2.83	-3.38
7	-3.71	0.00	0.00	-3.71	0.75	4.61	-3.71	-2.83	0.00	0.54	-3.71	-2.83
8	-3.39	-4.21	-4.31	0.37	-0.76	-0.19	-4.21	-3.71	-3.39	0.25	-0.19	-4.08
9	0.87	-3.92	-3.39	0.54	0.76	-0.50	0.00	-3.71	0.00	-3.71	0.00	-0.01
10	-2.83	-2.83	-3.39	-3.38	-3.70	0.00	-3.71	0.00	0.00	-0.01	-3.71	3.70
Overall	-1.85	-2.10	-2.34	-1.23	0.39	0.88	-1.53	-2.05	-1.25	0.17	-1.18	-1.28

2. Mean d-prime and post-hoc results for Fricative Pair levels averaged across position.

English speakers. Shaded cells are significant.

Fricative pair	Mean	s ʃʲ	s ʂ	ɕ ʃʲ	ɕ ʂ	ʃʲ ʂ
s ɕ	4.47	.705	.285	.001	.001	.001
s ʃʲ	4.39		.020	.001	.002	.003
s ʂ	4.73			.001	.001	.001
ɕ ʃʲ	2.77				.052	.271
ɕ ʂ	3.30					.570
ʃʲ ʂ	3.15					

3a. Mean RTs (ms) by Fricative Pair, Position, and Subject

Polish subjects

Subject	VC						CV					
	s ɕ	s ʃʲ	s ʂ	ɕ ʃʲ	ɕ ʂ	ʃʲ ʂ	s ɕ	s ʃʲ	s ʂ	ɕ ʃʲ	ɕ ʂ	ʃʲ ʂ
1	379	365	367	601	380	401	310	333	306	317	250	459
2	389	341	353	376	370	458	230	287	288	681	334	384
3	644	426	481	542	433	1132	686	793	409	377	430	492
4	680	513	397	608	472	1046	395	1060	316	784	345	367
5	325	396	360	587	342	267	357	504	371	438	254	226
6	616	742	1037	828	836	808	562	777	511	707	805	830
7	568	513	523	660	423	727	451	391	537	541	389	511
8	192	275	208	339	202	535	174	213	225	226	301	217
9	484	616	490	487	552	599	433	472	327	373	366	294
10	418	651	271	590	511	601	461	427	411	299	255	338
11	369	401	560	678	456	647	524	317	650	486	845	367
12	437	472	445	443	314	656	431	364	270	428	422	402
13	350	349	210	395	403	561	472	319	492	469	222	310
Overall	451	466	441	549	438	651	424	483	394	474	404	400

3b. Mean RTs (ms) by Fricative Pair, Position, and Subject

English subjects

Subject	VC						CV					
	s ɕ	s ʃʲ	s ʂ	ɕ ʃʲ	ɕ ʂ	ʃʲ ʂ	s ɕ	s ʃʲ	s ʂ	ɕ ʃʲ	ɕ ʂ	ʃʲ ʂ
1	1000	937	894	1250	841	847	689	717	653	1289	752	1117
2	1141	981	1042	1248	834	1260	521	904	613	1119	906	805
3	1146	1396	1220	1051	1599	1329	1541	1485	1187	1633	1696	1283
4	637	1075	607	806	1097	884	879	661	656	806	698	585
5	616	681	698	836	606	954	455	408	505	644	584	568
6	438	603	475	611	734	621	306	236	277	508	267	547
7	847	1089	678	1057	1217	1078	829	847	712	919	856	818
8	822	726	974	1123	695	1019	967	821	338	867	822	952
9	1450	1493	864	942	1104	1191	1236	1287	1016	1810	1161	1037
10	613	589	690	1044	876	896	597	677	592	641	521	737
Overall	869	960	813	981	961	987	801	806	657	1017	812	839

Figures

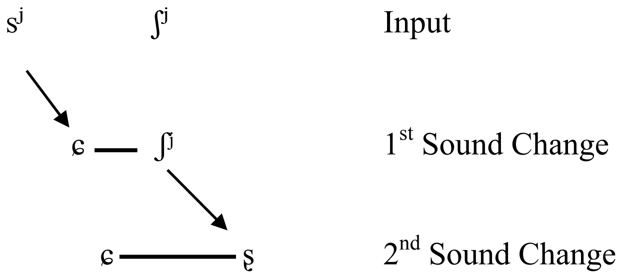


Figure 1: Sound changes in Polish

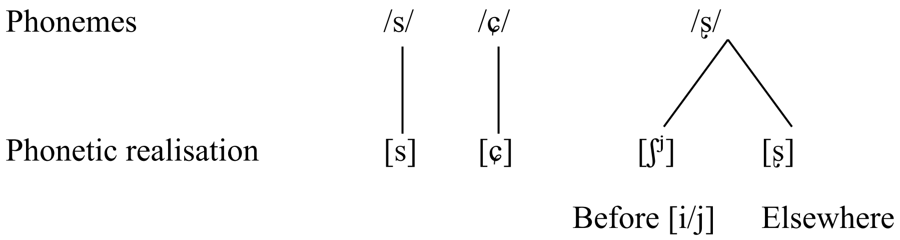


Figure 2: Polish phonemes and allophones

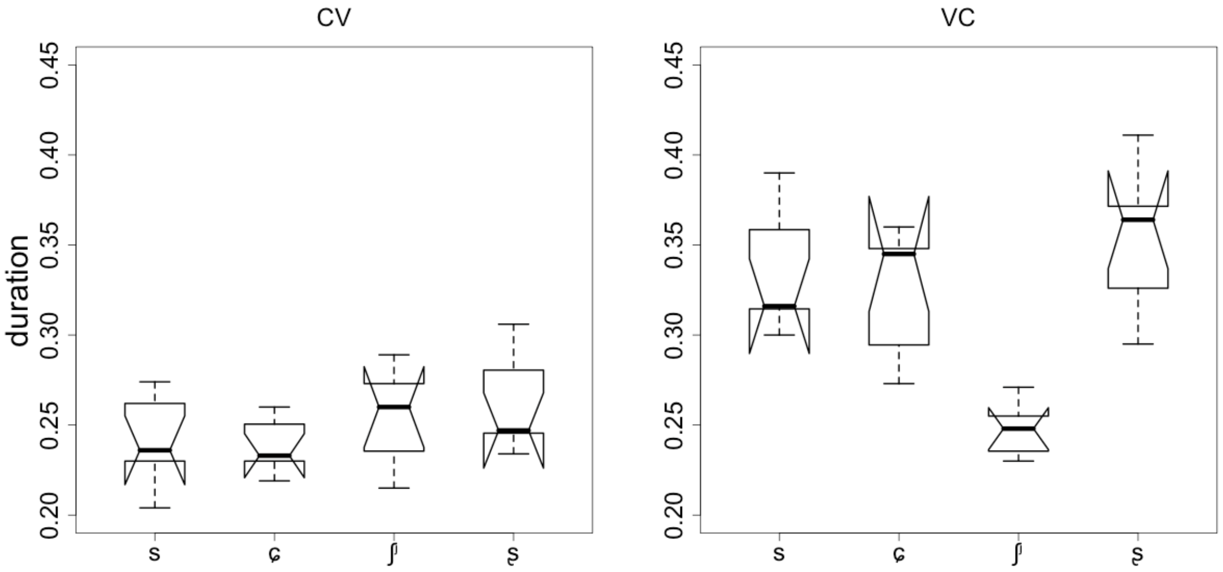


Figure 3: Duration (s) of the fricative portion of the stimuli, by position (CV vs. VC).

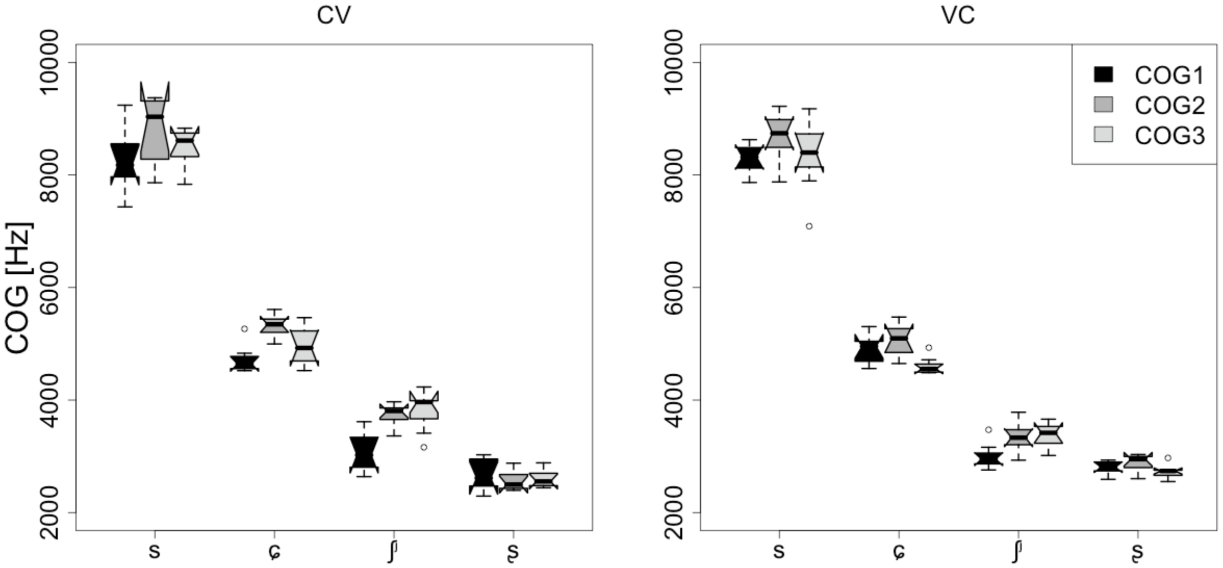


Figure 4: Center of gravity values (Hz) of the fricative portion of the stimuli, by position. For each fricative values are reported for the first third, middle third, and last third of the fricative duration.

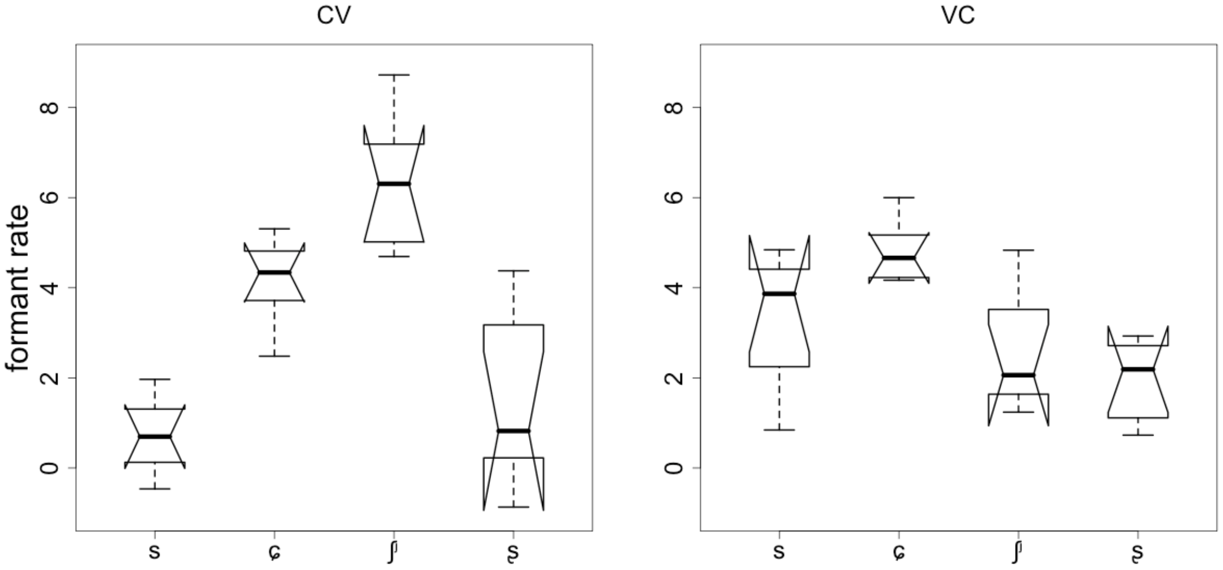


Figure 5: Rate of change of F2 (Hz/ms), by position.

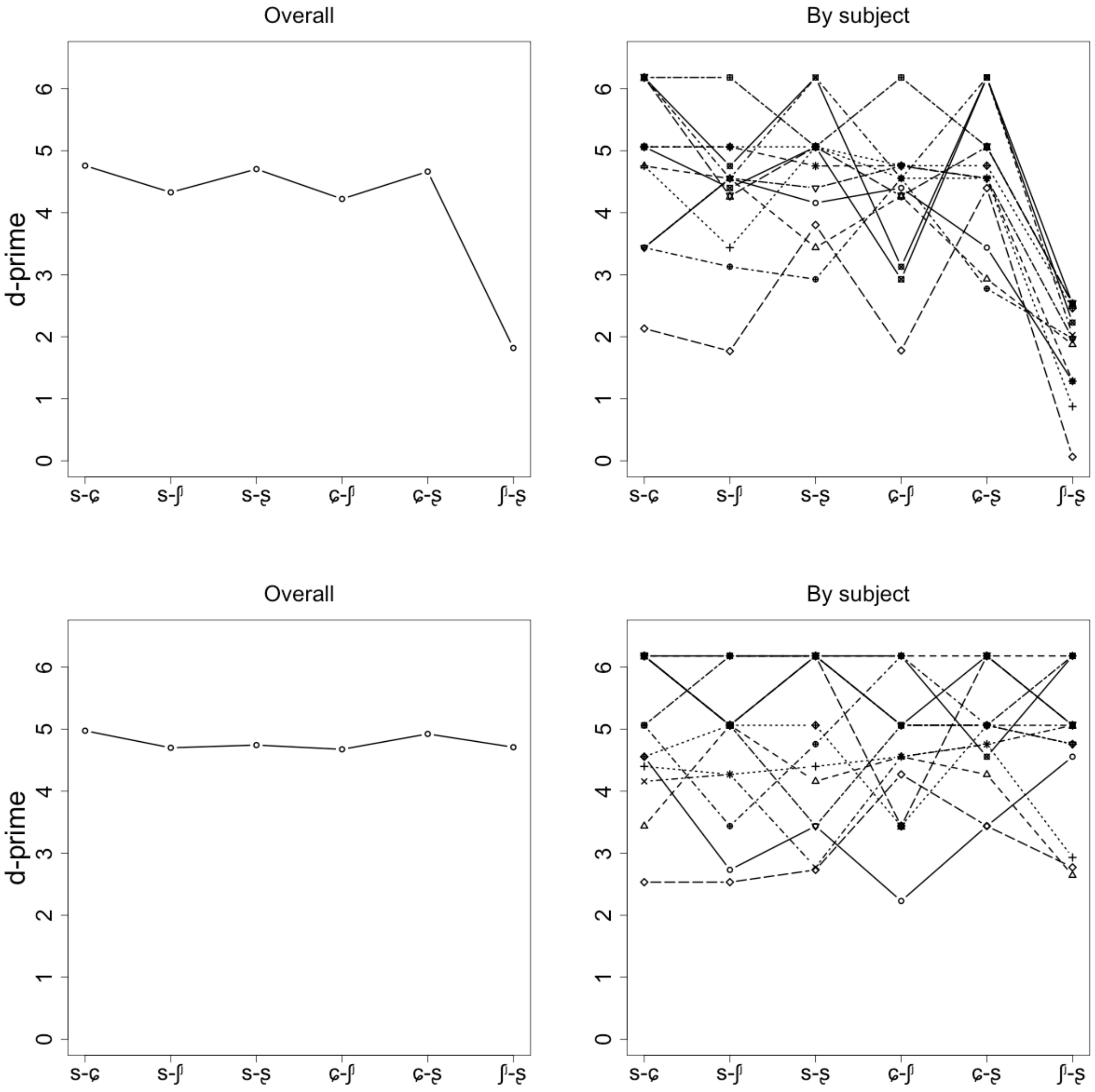


Figure 6: D-prime scores by fricative pair. All Polish speakers. Top panels: VC position; bottom panels: CV position. Left panels: overall means; right panels: all subjects.

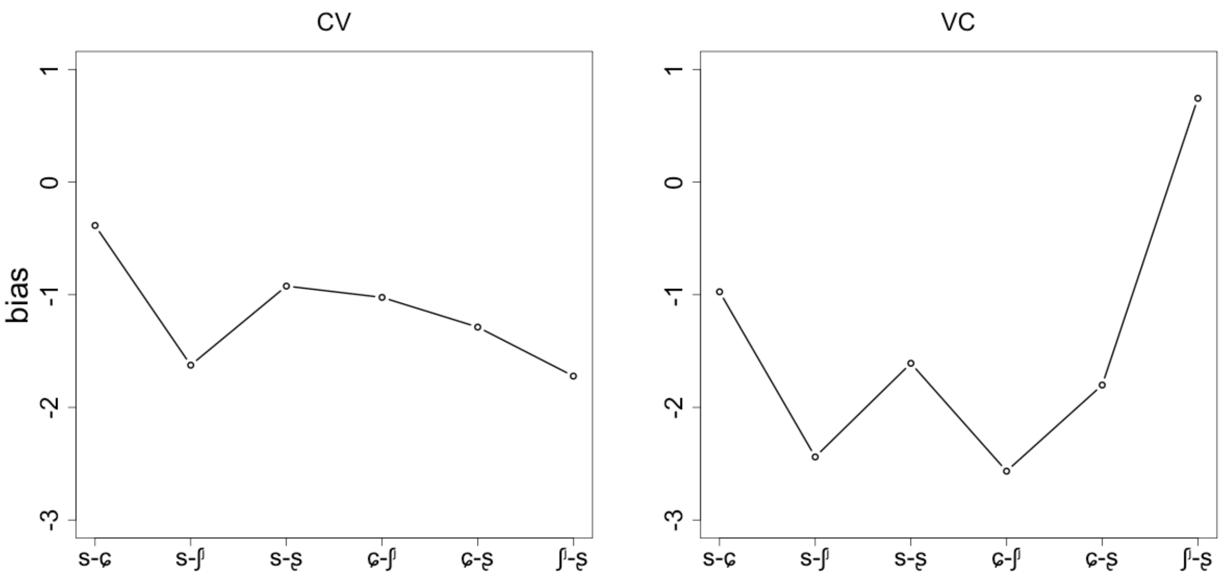


Figure 7: Mean bias scores by fricative pair. All Polish speakers.

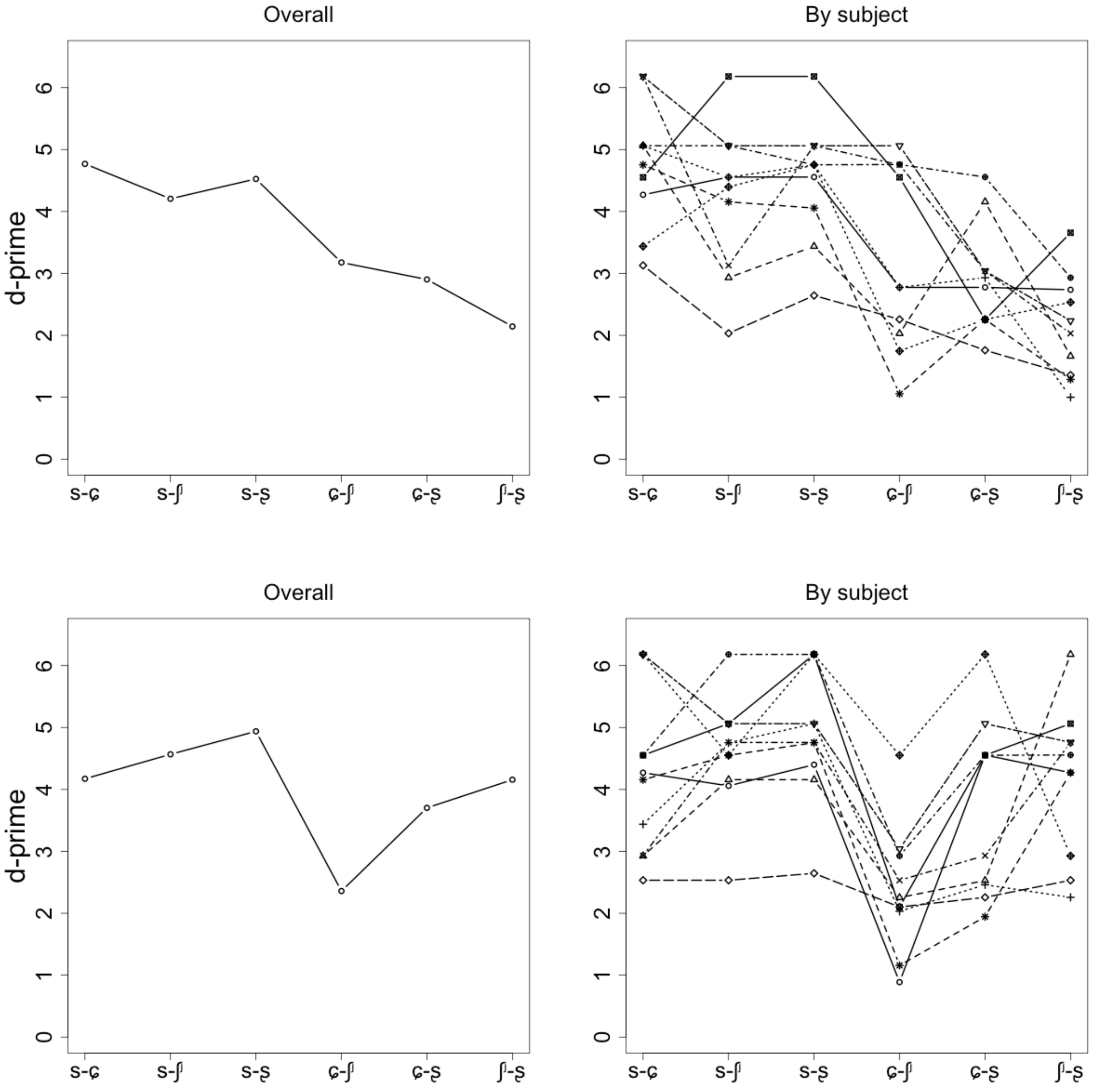


Figure 8: D-prime scores by fricative pair. All English speakers. Top panels: VC position; bottom panels: CV position. Left panels: overall means; right panels: all subjects.

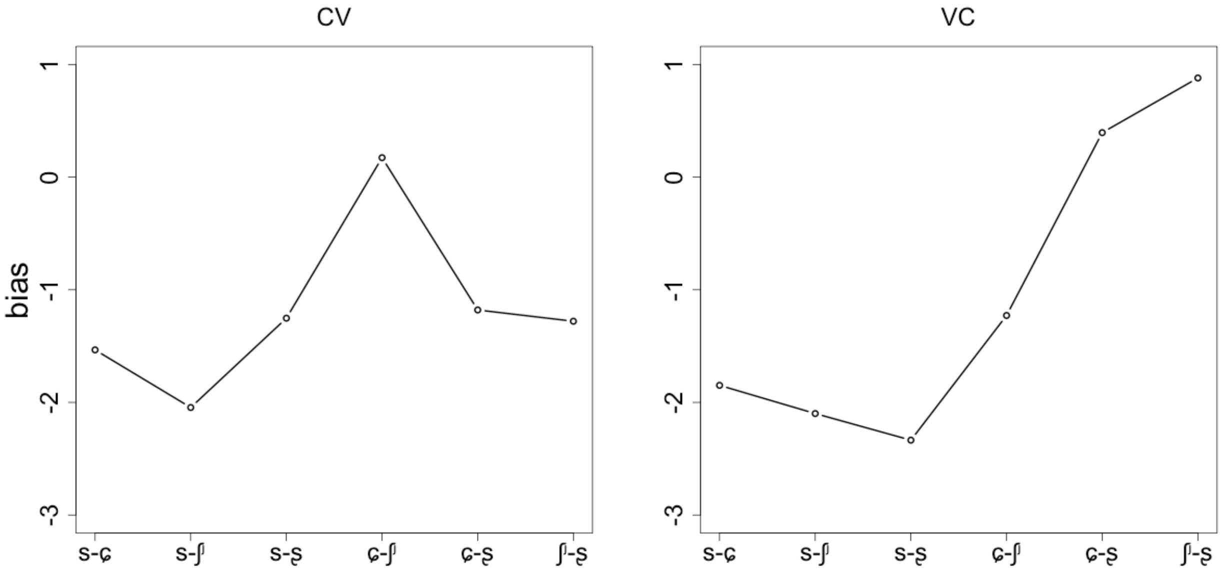


Figure 9: Mean bias scores by fricative pair. All English speakers.

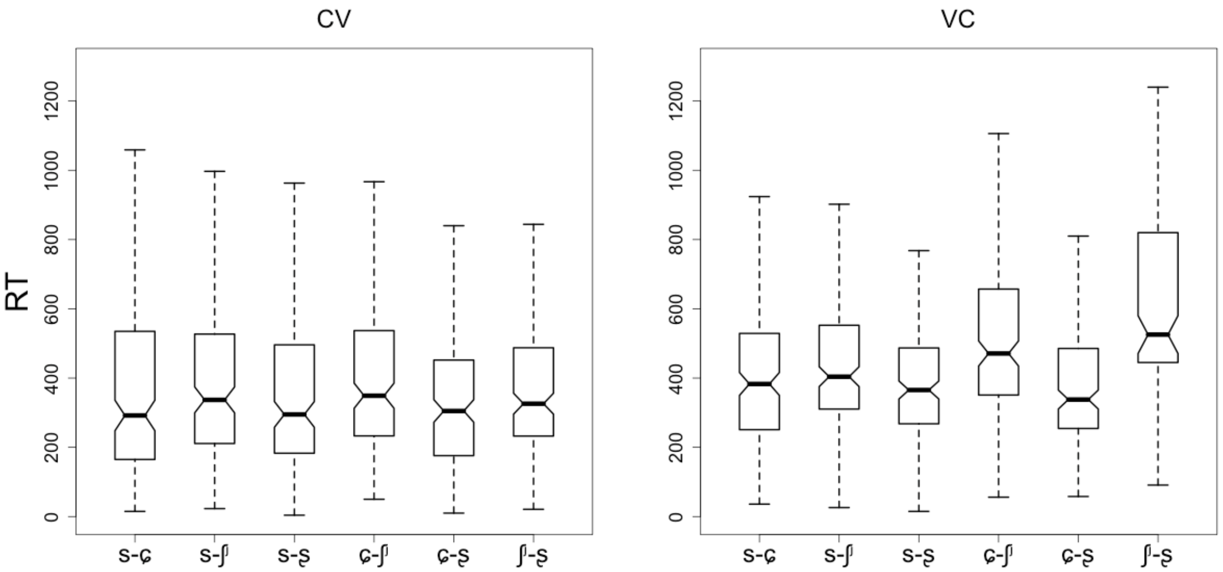


Figure 10: Reaction times by fricative pair and position (outliers not shown)
All Polish speakers

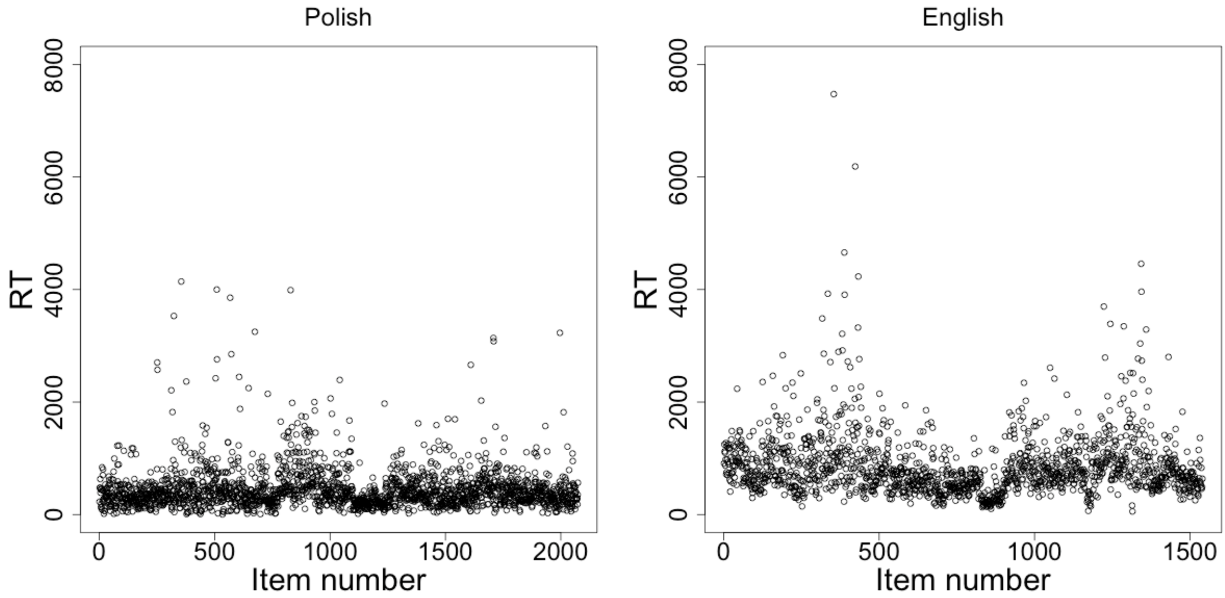


Figure 11: All individual RT scores for Polish and English subjects.

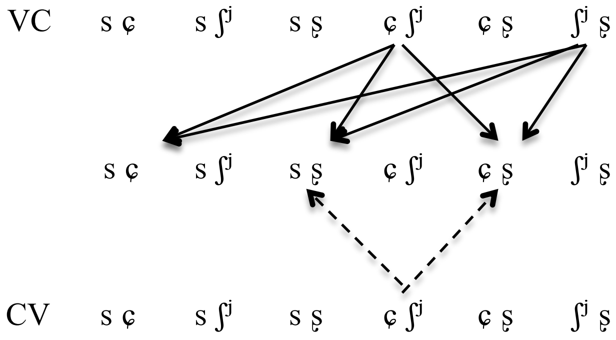


Figure 12: Diagram indicating RT differences that are statistically significant (solid lines) or nearly so (dashed lines) for positions VC and CV. An arrow from pair *x* to pair *y* means that the RT for *x* was significantly worse than the RT for *y*.

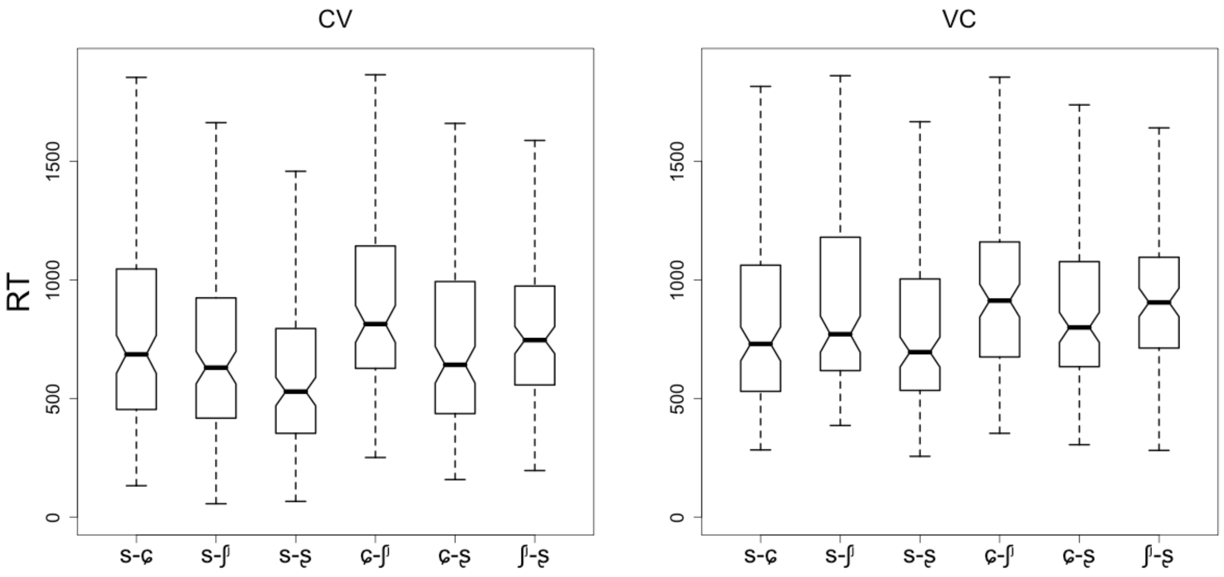


Figure 13: Reaction times by fricative pair and position (outliers not shown)

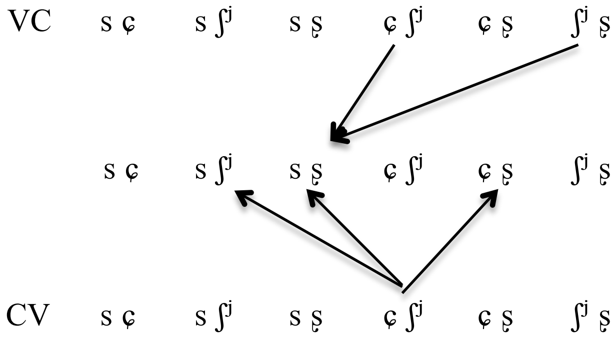
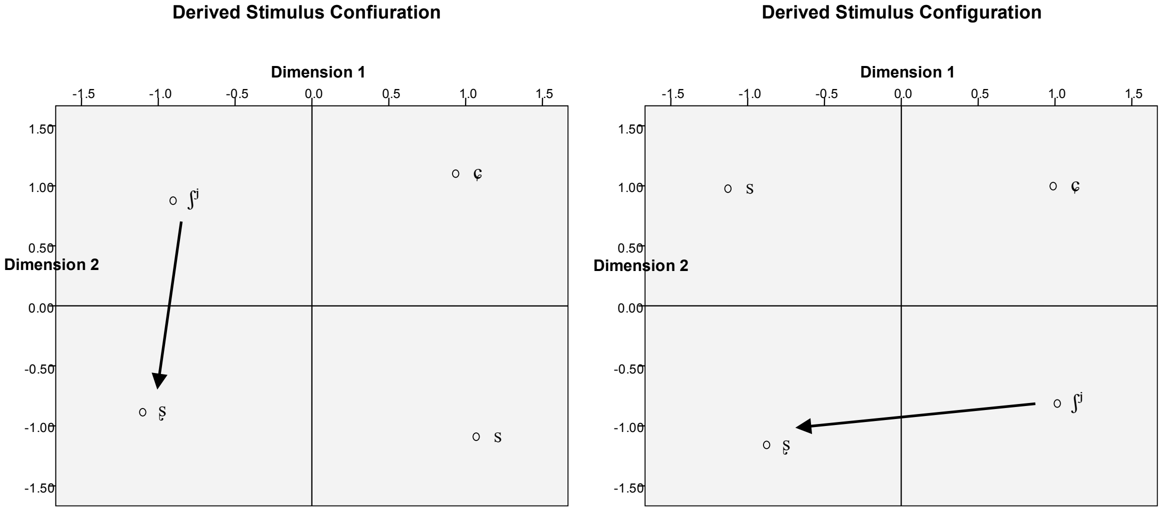


Figure 14: Diagram indicating RT differences that are statistically significant for positions VC and CV. An arrow from pair *x* to pair *y* means that the RT for *x* was significantly worse than the RT for *y*.

a. Polish VC
Stress = .170, RSQ = .296

b. Polish CV
Stress = .179, RSQ = .148



c. English VC
Stress = .180, RSQ = .369

d. English CV
Stress = .178, RSQ = .215

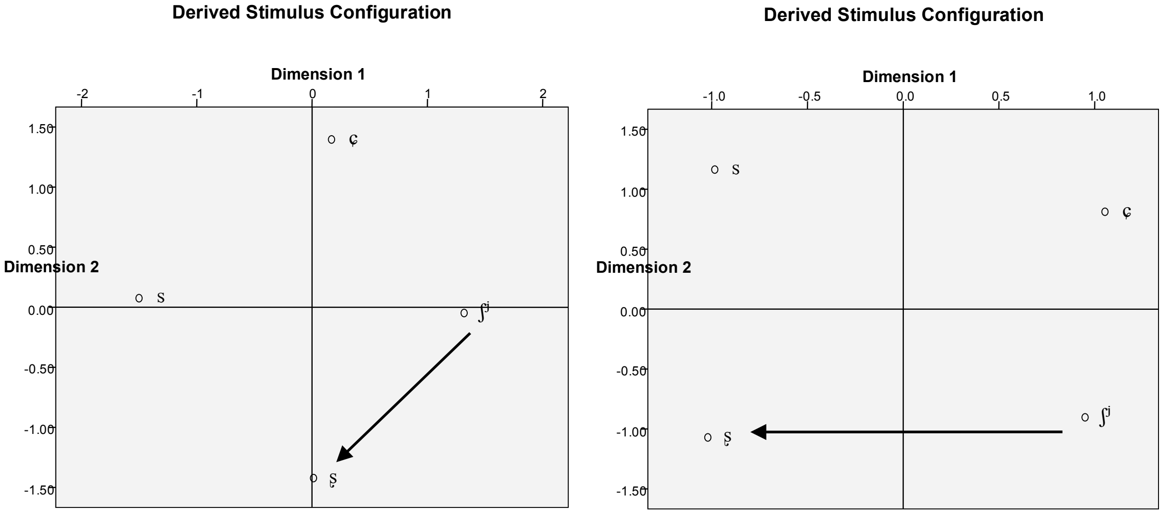


Figure 15: Multidimensional scaling plots using 1/ln(RT) as distance, for Polish VC and CV position (a-b resp.) and English VC and CV position (c-d resp.). The arrows (discussed in section 5.2) indicate the effect of the Polish sibilant retroflexion sound change on perceptual distances.

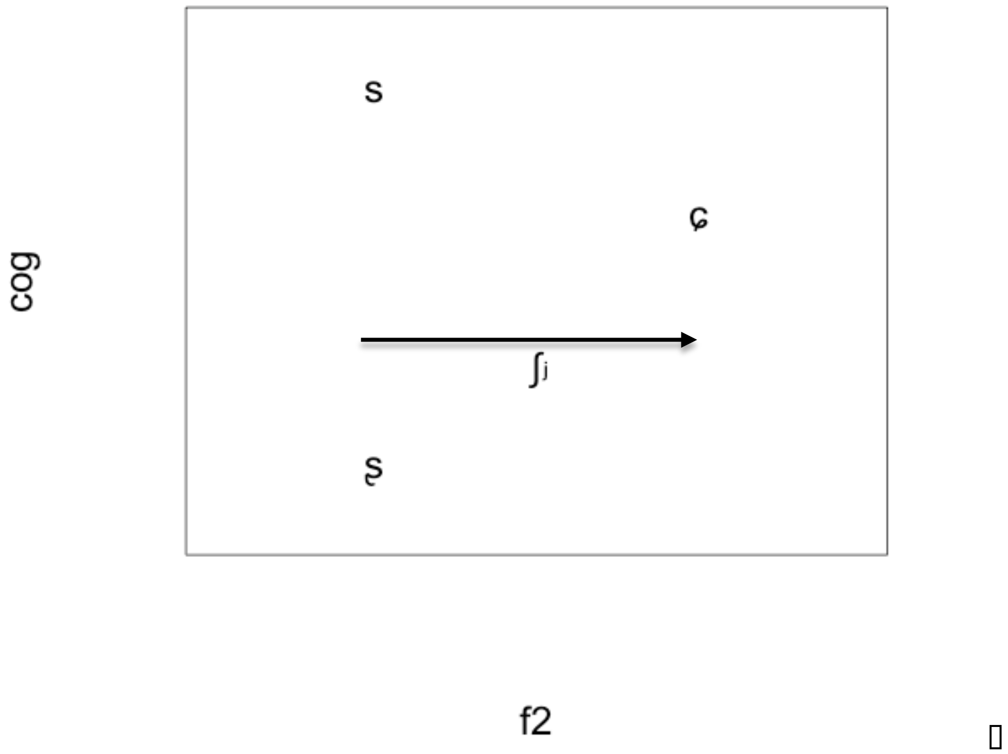


Figure 16: Schematic diagram of relative positioning of Polish sibilants based on COG and F2 properties. F2 of [ʃʲ] is significantly lower at onset than at offset, as the arrow is meant to suggest.