Accessing psycho-acoustic perception and language-specific perception with speech sounds

MOLLY BABEL and KEITH JOHNSON

University of British Columbia
University of California, Berkeley

Abstract

In this paper we review the results of four speech perception experiments that explore the difference between language-specific perception and psycho-acoustic auditory perception. The first two experiments examine this difference with voiceless fricatives by American English and Dutch listeners. The second series of experiments explores the perception of consonant palatalization by American English and Russian listeners. These experiments examine this processing difference through two tasks: speeded AX discrimination ("same" or "different" response) and similarity rating. The fast-paced nature of the AX discrimination task is designed to bypass linguistic processing and hone in on pure auditory similarity. The similarity rating task asks listeners to compare two stimuli at a more leisurely pace and language-specific perception is evaluated. The results of these experiments suggest that psycho-acoustic perception can be evaluated apart from linguistic perception. Other work using this experimental paradigm, however, has found language effects in both AX discrimination and rating tasks (Boomershine et al. 2008; McGuire 2007). We reconcile our findings with the contrary results by demonstrating that language effects tend to appear in the longer response latencies that naturally allow for linguistic processing and are attenuated in the fast responses.

1. Introduction

Listeners do not process speech sounds with a blank slate. That is, speech perception is language-specific in nature. This is made apparent in neurological studies that find differences in vowel perception through electroencephalograms, with speakers of Estonian and Finnish (Näätänen et al. 1997) and in the frequency-following response in the brain stem to lexical tone with Mandarin- and English-speaking participants (Krishnan et al. 2005). Behavioral studies examining speech perception have also demonstrated that the lack of contrast in a language negatively influences perception of non-contrastive

JLP 1 (2010), 179–205
DOI 10.1515/LABPHON.2010.009

1836-6346/10/01–179

sounds. The classic example is that of Japanese listeners' difficulty in identifying the English liquids /l/ and /ɹ/ (Goto 1971; Miyawaki et al. 1975; McKain et al. 1981; Strange and Dittman 1984; Logan et al. 1991; Yamada et al. 1992; Lively et al. 1993; Flege et al. 1996). Going beyond basic inventory differences across languages, additional research has found that allophonic patterns within a language also influence the perception of the sounds in the allophonic relationship (Gandour 1983; Dupoux et al. 1997; Harnsberger 2001; Hume and Johnson 2003; Huang 2004; Boomershine et al. 2008). For example, Boomershine et al. (2008) examined the perceptual similarity of [d], [ɾ], and [ð] by English and Spanish listeners. In English [d] and [ɾ] are allophones of a single phoneme and [ð] represents a different phoneme, while in Spanish [d] and [ð] are allophones of one phoneme and [ɾ] represents a separate phoneme. In a similarity rating task, listeners rated the sounds that are allophones in their respective native languages as more similar sounding. This finding of subjective perceptual similarity was echoed in the results of a speeded AX discrimination task where listeners showed increased latency in responding "different" to pairs of VCV sequences with consonantal allophones (e.g. English listeners responded slower to the pair [ada] and [aɾa]) than between "different" pairs involving realizations of different phonemes (e.g. English listeners responded faster to the pair [ada] and [aða]). On the other hand, there are some sounds used phonemically in languages that listeners from other language groups who lack these sounds in their inventory have no trouble perceiving as distinct. This has been the result found with clicks (Best et al. 1988), although it has been suggested that listeners from non-click languages perceive clicks as non-speech sounds (Best and Avery 1999) which may explain their perceptual distinctiveness.

Despite this research demonstrating that language experience distorts auditory processing, speakers of language can and do learn new contrasts in both second language acquisition and laboratory training tasks (McGuire 2007). Moreover, non-speech sounds are not subject to language specific influences, suggesting that the auditory infrastructure in the human listener allows for unbiased perception. There is evidence for unbiased perception in the processing of speech sounds as well. Fujisaki and Kawashima's (1969, 1970) studies demonstrate that within-category discrimination of speech sounds is better than predicted from categorization. Pisoni (1973) described the difference between auditory and phonetic short-term memory codes; the former perceptual mode focuses on acoustic differences in the stimuli while the latter is a more abstracted level of perception that evaluates whether two stimuli are the "same" phonetic segments. Similarly, auditory and phonetic perceptual stages were proposed in Pisoni and Tash (1974). They demonstrate that listeners can use within category acoustic information to discriminate sounds.

If it is possible to experimentally measure auditory perception without influence of language specific phonetic perception using speech sounds with listeners from different languages, we should be able to measure the raw auditory similarity between phones. This auditory similarity can then be compared against perceived phonological similarity. The relationship between these two types of similarity—perceived phonological similarity and raw auditory similarity—is of great interest to linguists with respect to both the potential to explain perceptually-motivated sound changes and to provide perceptual motivation for phonological relationships.

In this paper we discuss a series of experiments that explore the difference between psycho-acoustic perception and linguistic perception of voiceless fricatives with Dutch and American English listeners (Johnson and Babel 2010) and palatalized consonants with Russian and American English listeners, which was briefly described in Babel and Johnson (2007). We make use of a discrimination task and a similarity rating task to probe these two different types of perception. For our purposes, discrimination tasks are preferred over labeling tasks as they allow listeners to judge the perceptual similarity of two tokens without having to make explicit judgments about the proper orthographic representation for the sound or whether such a symbol even exists within their repertoire (cf. Best et al. 2001; Gerrits and Schouten 2004). The first experiment for all subject populations is a similarity rating task. Listeners are asked to compare two stimuli presented with a 100 ms. interstimulus interval (ISI) on a 5-point equal interval scale ranging from "very similar" to "very different". This task allows for more leisurely processing of the tokens and has been shown to be sensitive to language-specific perception (Boomershine et al. 2008). At the same time it maintains the short ISI typical of discrimination tasks to minimize differences between the two tasks; this way the only crucial difference between the two tasks is the task to be completed by the listener. The second task is a speeded AX ("same" or "different") discrimination task where two stimuli are also presented consecutively with a 100 ms. ISI. Listeners are instructed to respond within 500 ms.; their response time and accuracy are presented on the screen between trials to motivate their performance. The fast-paced nature of this task is designed to circumvent linguistic processing and assess pure auditory similarity (Pisoni 1973; Pisoni and Tash 1974; Fox 1984).

Previous work that has employed this paradigm has found language effects to pervade all levels of perception (McGuire 2007; Boomershine et al. 2008). Admittedly, one of the chief purposes of this paper is to provide a post-hoc explanation for the results of McGuire (2007) and Boomershine et al. (2008). Therefore, in the last section of this paper, we demonstrate that the language effects reported by McGuire and Boomershine et al. are the consequences of long reaction times that allow for linguistic processing and do not unequivocally demonstrate language experience affects all levels of auditory perception.

2. Fricative perception

The work reported in this section is discussed in more depth in Johnson and Babel (2010). These experiments explore the perception of voiceless fricatives [f θ s ʃ x h] by Dutch and English listeners. The phonemic inventory of Dutch (Booij 1995; Cohen et al. 1972; De Groot 1968) includes the velar fricative [x], but Dutch does not use [θ] or [ʃ] contrastively. English on the other hand uses [θ] and [ʃ] to contrast words but not [x], although [x] may occasionally be a fronted articulation of /h/. We expect then, given these descriptions of the inventories of contrastive sounds and previous research on second language speech perception, that Dutch speakers will have greater sensitivity to [x] when compared to other fricatives than will English speakers and that English listeners will be more sensitive to the phonetic properties of [θ] and [ʃ]. These predictions depend though on whether listeners identify nonnative fricatives with fricatives that they are familiar with. For example, we might expect that English listeners would hear [x] as a variant of [h] because of the acoustic similarities between these sounds. The possibility that English /h/ is sometimes said with velar frication (e.g. in *who*) may also be a factor. If English listeners make this association then we would expect that [x] and [h] would be more phonetically similar to them than they are for Dutch-speaking listeners (see Best 1995 and Flege 1995 for more detailed discussion of these issues). Similarly, we might expect that Dutch listeners would have lowered perceptual sensitivity to [θ] particularly in contrast with [f], which they do have in their inventory.

The case of [ʃ] in Dutch is different from [x] in English or [θ] in Dutch where the fricatives—[x] and [θ]—are not contrastive in one of the languages' phonological inventory. This is because phonetic [ʃ] is also found in Dutch. The CELEX Dutch lexicon encodes [ʃ] along with [f s x h] as one of the phonetically transcribed fricatives of Dutch (Baayen et al. 1993). Indeed, Nooteboom and Cohen (1976:144) transcribe Dutch with [ʃ], arguing that words such as *meisje* [ʃ] 'girl', *sjaal* [ʃ] 'shawl', *chef* [ʃ] 'chef, boss', and *sjouwen* [ʃ] 'carry' demonstrate that [ʃ] is a contrastive sound in Dutch—though one only found in borrowed words or words of Frisian origin (e.g. *sjouwen*). However, other analysts (Booij 1995; Cohen et al. 1972; De Groot 1968) have noted that [s] and [ʃ] alternate in diminutive forms, like 'girl' above (*poes* [s] 'cat'—*poesje* [ʃ] 'kitten', *tas* [s] 'bag'—*tasje* [ʃ] 'small bag') and in connected speech when [s] and [j] are adjacent across word boundaries (*was je* [ʃ] 'were you', *zes januari* [ʃ] 'January the 6th'). So, on the one hand, Dutch speakers do have extensive experience hearing [ʃ] and in some words [ʃ] does not alternate with [s] (eg. *meisje* [ʃ] 'girl', *sjaal* [ʃ] 'shawl', etc. cited above) making it reasonable to simply represent the word with an underlying phonemic /ʃ/. However, on the other hand, some instances of [ʃ] do alternate with [s] as if [ʃ] is a contextual variant of [s]. [ʃ] and [s] also alternate with each other to a limited extent in English (e.g., *oppress*, *oppression*; *confess*, *confession*). The main differences between the languages are that the [s]/[ʃ] contrast is more prevalent in the English lexicon while [s]/[ʃ] alternations seem to be more common in Dutch. We were interested in whether these linguistic differences would lead to any perceptual differences between Dutch and English listeners.

2.1 *Experiment 1: Rating perceptual similarity of voiceless fricatives*

2.1.1 *Method.* Sixteen American English speakers participated in the first experiment. Participants were college undergraduates who received course credit for their participation. None of the participants reported any speech, language, or hearing disorders.

Twelve Dutch speakers also participated in experiment 1. These participants were compensated $10 for their participation and none of them reported any speech, language, or hearing disorders.

The stimuli for this experiment consisted of 18 vowel-fricative-vowel sequences produced by the second author, a trained phonetician and native speaker of American English. They were composed of one of six voiceless fricatives [f θ s ʃ x h] embedded in one of three vowel environments [a_a], [i_i], or [u_u]. All of the tokens were produced with accent on the first vowel in a declarative intonation pattern.

The stimuli were presented in pairs binaurally with a 100 ms. ISI to listeners over headphones at a workstation using E-prime experiment software (Schneider et al. 2002). Listeners were instructed to rate the similarity between the two tokens on a 5-point scale. Participants logged their responses on a 5-point equal-interval button box where the buttons had the following labels: [1] very similar, [2] somewhat similar, [3] moderately different, [4] somewhat different, and [5] very different. Listeners had up to five seconds to respond before the presentation of the next set of stimuli. Same pairs of stimuli (eg. [asa] and [asa]) were presented twice in each block so that there were 42 trials per vowel (30 different pairs and 12 same pairs). Each of the AX pairs was presented three times for a total of 378 trials (42 per vowel for each of 3 vowels, each in 3 repetitions). While the presence of same pairs in similarity rating tasks is atypical, they were included to create parity between the rating task and the discrimination task. Note as well that the most similar label was “very similar” and not “identical”. Trials were presented in a unique random order for each participant.

2.1.2 *Results.* The rating scores for the different pairs were analyzed in a repeated measures analysis of variance with the between-listeners factor native language (English vs. Dutch) and the within-listeners factors vowel (/i/, /a/, or

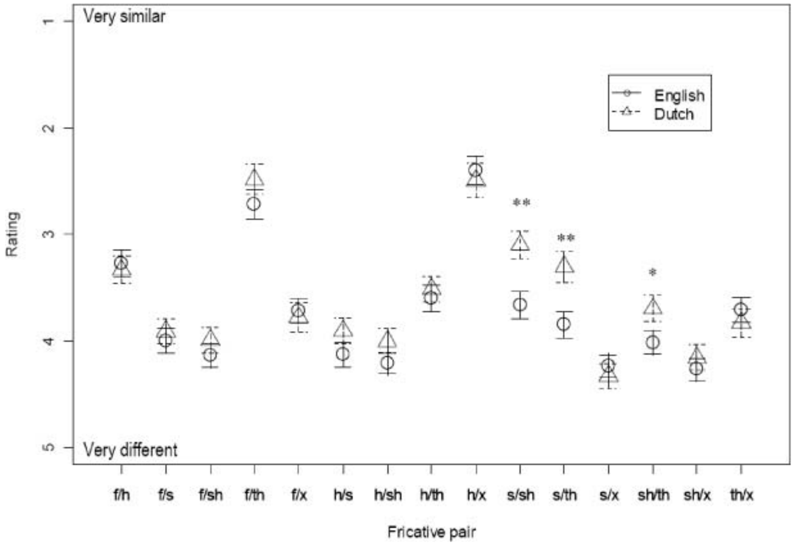
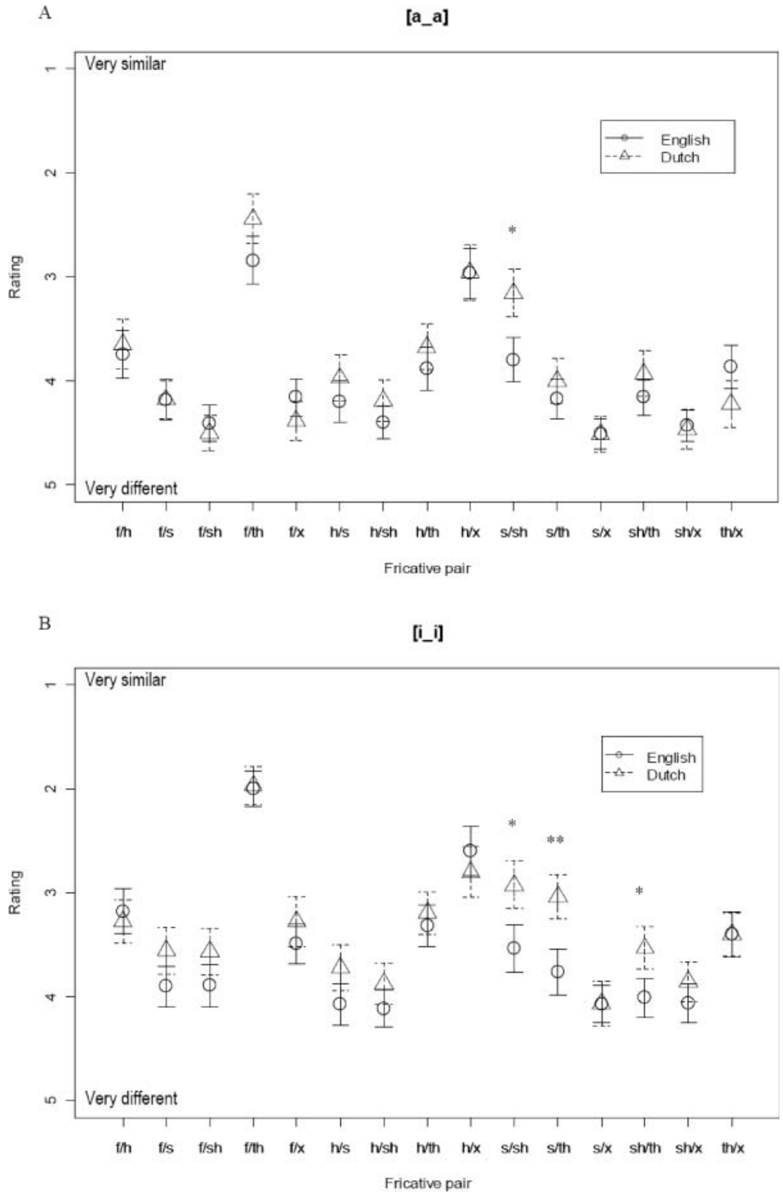


Figure 1. *Results of the first experiment's fricative pair by language interaction. Judgments of perceived similarity are shown on the vertical axis. A rating of [1] denotes a rating of "very similar" while a rating of [5] marks a pair as "very different". In the fricative pairs the label "sh" is used for [ʃ] and "th" is used for [θ]. Pairs marked with "**" or "*" showed a significant language difference in a planned comparison at $p < 0.01$ and $p < 0.05$, respectively. Error bars represent 95% confidence intervals.*

/u/) and fricative pair (15 comparisons). The vowel main effect was significant ($F[2, 52] = 65.3$, $p < 0.01$), as was the fricative pair main effect ($F[14, 364] = 94.5$, $p < 0.01$). The fricative pair by vowel interaction was also reliable ($F[28, 728] = 15.5$, $p < 0.01$). We also found a fricative pair by language interaction ($F[14, 364] = 3.8$, $p < 0.01$), and the three-way, pair by vowel by language, interaction was also significant ($F[28, 728] = 1.6$, $p < 0.05$). Figure 1 shows the fricative pair by language interaction, and in this figure we see that the main points of difference between the Dutch and English listeners were with pairs that contrasted [s] and [ʃ] (labeled "s/sh" on the horizontal axis of the graph), pairs that contrasted [s] and [θ] (labeled "s/th"), and pairs that contrasted [ʃ] and [θ] (labeled "sh/th"). Post-hoc testing revealed significant differences between listener groups only with respect to these three fricative pairs.

Figure 2 shows the fricative pair by vowel by language interaction. Post-hoc testing revealed the three pairs for which Dutch and English listeners differed in the pair by language interaction also differed in the [i_i] environment. In the [u_u] environment the only language effect was for the s/θ pair. In the [a_a] environment Dutch and English listeners differed only for the s/ʃ pair. Dutch listeners rated [s] and [ʃ] more similar to each other than did English



Figures 2a–c. *Results of the fricative pair by language by vowel interaction in experiment 1. Judgments of perceived similarity are shown on the vertical axis. A rating of [1] denotes a rating of "very similar" while a rating of [5] marks a pair as "very different". In the fricative pairs the label "sh" is used for [ʃ] and "th" is used for [θ]. Pairs marked with "**" or "*" showed a significant language difference in a planned comparison at $p < 0.01$ and $p < 0.05$, respectively. Error bars represent 95% confidence intervals.*

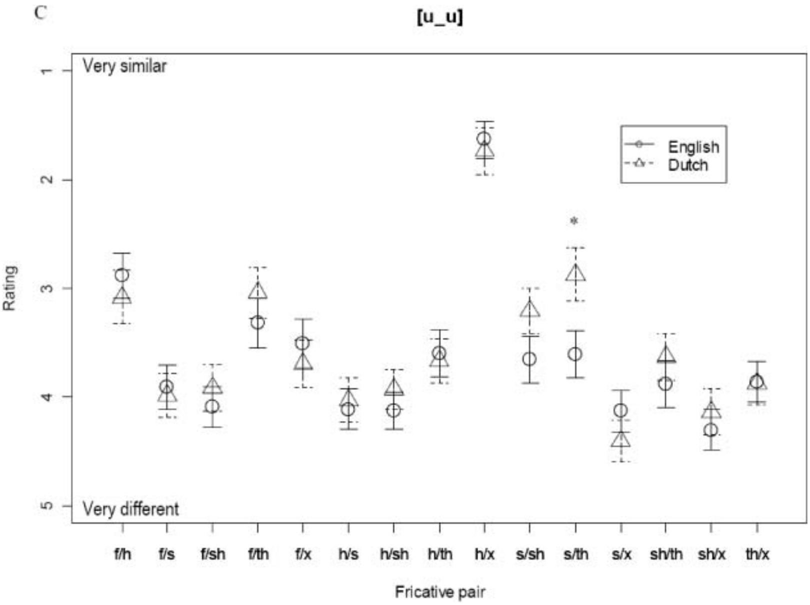


Figure 2 *(Continued)*

listeners. This pattern was reliably present for [isi]/[iʃi] and [asa]/[aʃa] and we observed a trend in this direction for [usu]/[uʃu]. This finding accords with Boomershine et al. (2008), which suggests that phonological alternation influences speech perception by reducing contrast between phones that stand in alternation with each other. Of interest here is the phonological ambiguity of [s] and [ʃ] and the low functional load of [ʃ] in Dutch, because they contrast in some lexical pairs and alternate in others. The results of this similarity rating experiment suggest that phonological structure has a powerful impact on speech perception. For our Dutch listeners, we found the perceived phonetic difference between [s] and [ʃ] to be reduced relative to the phonetic difference reported by English speakers both in a context where [s] becomes [ʃ] for Dutch speakers (e.g. a front vowel) and in environments where [s] and [ʃ] may contrast.

In addition to this effect of allophonic alternation, the different phonological inventories of Dutch and English are also related to differences between Dutch and English listeners' perception of [θ]. Dutch listeners rated [θ] as more similar to [s] than did English listeners both for the [isi]/[iθi] and the [usu]/[uθu] pairs. Dutch listeners also rated [iʃi]/[iθi] as sounding more similar than did English listeners. In these pairs, Dutch listeners found [θ] to be less distinct than did English listeners. It may be that our Dutch listeners were attending more to the vowel formant transitions of [θ] than to the fricative noise (McGuire 2007). Interestingly, this effect of phonological inventory was not found across the board for all pairs involving [θ]. For example, we might have expected [θ] and [f] to be more confusable for Dutch than for English listeners and though the average ratings do trend in this direction in [ufu]/[uθu] and [afa]/[aθa] these differences were not significant. It might be that for highly confusable pairs such as [f]/[θ] the rating task is not sensitive to language differences because a floor effect on rating scores obscures any differences between the Dutch and English listeners.

Another inventory difference that led us to expect perceptual differences did not have an effect. This is the presence of [x] in Dutch and the lack of a velar fricative in English. The pattern of [x] in the perception results is different from [θ]. While we saw no language difference for [f]/[θ] there were indications that the Dutch and English listeners differed in how they perceive [θ]. With [x] we found that none of the pairs involving [x] showed any difference between the Dutch and English listeners—not the highly similar pair [x]/[h] nor any of the less similar pairs. This suggests that the two groups of listeners were on an even footing when faced with these stimuli.

Experiment 2 uses the same stimuli but gives the listener a different task. Here, the task of the listener is judge whether the two stimuli are "same" or "different". The rapidity of the response is expected to force listeners to make judgments based on the auditory similarity of the sounds alone.

2.2 *Experiment 2: Voiceless fricative discrimination*

2.2.1 *Method.* Nineteen American English listeners participated in this experiment. Participants received partial course credit for their participation and they reported no speech, language, or hearing disorders. Data from two subjects were removed because English was not their native language.

Fifteen Dutch listeners participated in experiment 2. These participants were compensated $10 and none reported any speech, language, or hearing disorders. Nine of these Dutch volunteers also participated in experiment 1. A gap of approximately three months passed between their participation in the two experiments.

The same stimuli that were used in the first experiment were used again for experiment two.

The basic procedure for experiment two is the same as that reported for experiment one with the exception of the listener's task. Listeners were instructed to judge whether the two tokens were same or different by responding on a button box. Listeners were encouraged by the experimenter to keep their reaction times under 500 ms. In order to further monitor their response time and attempt to keep it under 500 ms, reaction time feedback was presented on the computer screen.

2.2.2 *Results.* Reaction time was measured from the onset of the second stimulus in the trial. Incorrect responses and correct "same" responses were removed from the data set prior to analysis. Log reaction times were used under the assumption that differences in participant responses would be greater among faster reaction times than among equidistant slower reaction times. In a repeated measures analysis of variance of the log reaction times for correct "different" responses, there was a vowel main effect ($F[2, 11] = 18.9$, $p < 0.01$), a fricative pair main effect ($F[14, 393] = 15.8$, $p < 0.01$), and a pair by vowel interaction ($F[28, 888] = 3.4$, $p < 0.01$). None of the interactions involving the language of the listener reached significance, though the pair by language interaction was close ($F[14, 393] = 1.6$, $p = 0.07$). Visual inspection of the pair by language interaction indicated that this trend was due to longer reaction times for American English listeners for the hardest pairs [f]/[θ] and [h]/[x], rather than any pattern that matches the language by pair interaction that was found in experiment 1 (Figure 1). This nominal difference between language groups may be an indication that the Dutch listeners were trying harder to keep their reaction times under 500 ms. even when the discrimination judgment was difficult to make.[1]

The experiment 2 pair by vowel interaction is shown in Figure 3. The parallels between the reaction time data shown in this figure and the average similarity ratings that were obtained in experiment 1 are striking. Pairs with longer

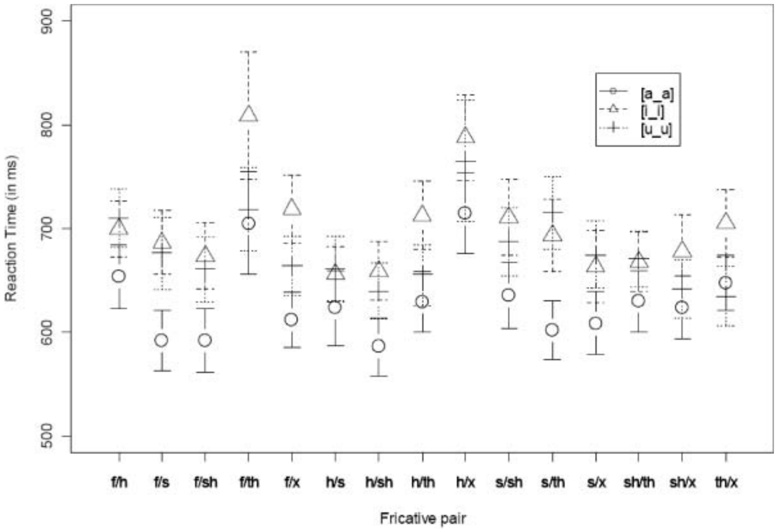


Figure 3. *Results of the fricative pair by vowel interaction in experiment 2. In the fricative pairs the label "sh" is used for [ʃ] and "th" is used for [θ]. Error bars represent 95% confidence intervals.*

mean reaction times are those that were explicitly judged to be more similar sounding in the rating experiment. Average reaction times and similarity rating judgments for each pair (collapsed across different vowel environments and across listener language) were strongly correlated [$t(13) = -15$, $p < 0.001$, Pearson's $R = -0.97$]. For example, in the rating data plotted by vowel (Figure 2) we found that the perceived difference between [uhu] and [uxu] was smaller than the perceived difference between [ufu] and [uθu], while in the [i_i] environment the opposite trend was found—the difference between [ihi]/[ixi] was larger than the difference between [ifi]/[iθi]. This pattern is also apparent in the reaction time data in Figure 3. Reaction times for "different" responses to [uhu]/[uxu] were longer than to [ufu]/[uθu] while reaction times for "different" responses to [ifi]/[iθi] was longer than to [ihi]/[ixi].

2.3 *Discussion*

In summary, experiment 2 has two main results. The first is that we found no consistent effect of the language of the listener. Dutch and English listeners showed similar reaction time patterns in this experiment, suggesting that this experiment was successful in measuring the raw auditory distances between these stimuli without influence from the listener's native language. The second main finding is that the reaction time patterns found in this experiment are highly correlated with the phonetic similarity ratings found in experiment 1. As noted above, average reaction times and similarity rating judgments for each pair were highly correlated. These findings suggest that the language differences found in experiment 1 are not due to low-level linguistic influence on the auditory system, though previous research clearly demonstrates that such distortion does occur in some circumstances (e.g. Näätänen et al. 1997; Krishnan et al. 2003; Kuhl et al. 1992), and that responses in both were highly influenced by raw auditory similarity. The strong correlation also demonstrates that the subjective responses in experiment 1 were motivated by auditory similarity; this follows from recent work that views more abstract phonological relationships as emerging from phonetic properties (eg. Mielke 2008). The next set of experiments explores the language-specific and psycho-acoustic perception of various degrees of consonant palatalization found in Russian.

3. Palatalization perception

A preliminary report of these experiments was given in Babel and Johnson (2007); however, the discussion of the tasks in this paper goes into much more depth. The consonant-vowel sequences discussed in this section and used as stimuli in the next set of experiments correspond to sequences taken from Russian lexical items. Examples of the degrees of palatalization, by which we mean amount of palatal constriction within a sequence, in consonant-vowel sequences are shown in (1). By degrees of palatalization we mean the amount of palatal constriction within a sequence, ranging from none in (1a) to a sequence of three segments involving an articulation within the palatal region, as in (1d).[2]

(1)	a.	CV	[mat]	'checkmate'
	b.	C^{j}V	[m^{j}at]	'rumpled'
	c.	C^{j}jV	[sud^{j}ja]	'judge'
	d.	C^{j}ijV	[zm^{j}iju]	'snake'

A series of experiments with native speakers of Russian and American English speaking Russian language learners reported in Diehm (1998) found striking effects of language experience on the perception of palatalized consonant sequences. The results were quite surprising and complex. In a labeling task Russian listeners performed worse than Russian language learners in identifying C^{j}jV and C^{j}ijV sequences. The C^{j}ijV sequence is rare in Russian and carries a low functional load. C^{j}jV and C^{j}ijV sequences are considered to be in a near-merger relationship in Russian. So, while speakers of Russian produced C^{j}jV and C^{j}ijV sequences differently, they are not very sensitive to the acoustic differences that distinguish the pair. The native speakers of English who were adult Russian language learners with only tentative knowledge of the Russian phonological system seem to have focused on the acoustic differences and perceived the system auditorily, and not linguistically. Interestingly, native speakers of English who were advanced Russian language learners performed worse on the labeling task than the novice language learners in perceiving the difference between C^{j}jV and C^{j}ijV. That is, they responded to the stimuli more like native speakers, demonstrating their "improved functional knowledge" of Russian (Diehm, 1998). From this we can predict that in our experiment Russian listeners should perceive C^{j}jV/C^{j}ijV pairs as less different in a similarity rating task than other pairs, since Diehm found this pair less discriminable for Russian listeners.

The phonemic inventory of Russian includes a series of palatalized consonants as a counterpart to every plain consonant (Jones and Ward 1969). Palatalized consonants do not occur phonemically in English, but some consonant-palatal glide sequences do arise occasionally in words like *music* [mjuzɪk] and *few* [fju], but the articulation of these sequences in English is different from actual palatalized sequences in Russian. In this case, we predict that American English listeners will generally be less sensitive than Russian listeners to the phonetic differences in palatalized consonants given that they have no experience perceiving and producing them in English.

The speeded AX discrimination experiment and the similarity rating task described in the following section investigate the effect of language on palatalized consonant sequences like those studied by Diehm (1998) with native

Russian listeners and American English listeners who have not had any experience with Russian.

3.1 *Experiment 3: Rating perceptual similarity of consonant palatalization*

3.1.1 *Method.* Thirteen American English listeners participated as listeners in experiment 3. Subjects reported no speech, language, or hearing disorders and were compensated $10 for their time.

Ten speakers of standard Russian participated as listeners in this experiment. They were compensated $10 for their time and reported no speech, language, or hearing disorders.

The stimuli were open syllables with the test consonants in their onsets.

Possible onsets were: /m/, /v/, /b/, /d/, /l/, and /r/ and their palatalized counterparts. Consonants were produced with varying degrees of palatalization, and followed by a vowel: /a/, /u/, or /i/. The degrees of palatalization can be divided into four levels; so, for example, a set of stimuli from a single consonant and vowel combination was da, d^ja, d^jja, and d^jija. As noted above,

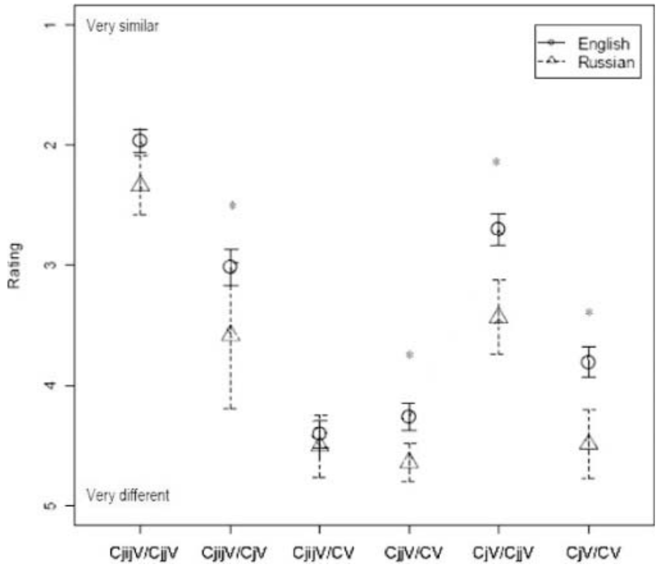


Figure 4. *Results of the experiment 3 palatalization by language interaction. Judgments of perceived similarity are shown on the vertical axis. A rating of [1] denotes a rating of "very similar" while a rating of [5] marks a pair as "very different". Pairs marked with "*" showed a significant language difference in a planned comparison at $p < 0.05$. Error bars represent 95% confidence intervals.*

throughout the course of this paper these types of syllables are referred to as CV, C^jV, C^jjV, and C^jijV, respectively. A female native speaker of Russian from Moscow produced the syllables ($6 \times 3 \times 4 = 72$) after saying real words containing the sequences in a carrier phrase.

The same basic procedure described for experiment 1 was used in this experiment. Listeners rated each pair six times throughout the task.

3.1.2 *Results.* A repeated measures ANOVA with listener rating responses from the different pairs as the independent variable and vowel context (/a/, /i/, or /u/), listener language (Russian vs. English), and palatalization pair (6 comparisons) as the dependent variables found main effects for language ($F[1, 12] = 5.26$, $p < 0.05$), palatalization pair ($F[5, 90] = 354.27$, $p < 0.001$), and vowel ($F[2, 30] = 20.75$, $p < 0.001$). There were significant interactions between palatalization pair and vowel ($F[10, 210] = 17.79$, $p < 0.001$) and palatalization pair and language ($F[5, 90] = 3.08$, $p < 0.05$).

Figure 4 shows the rated similarity of the pairs by language. The palatalization pair by language interaction indicates that Russians and naïve American English listeners perceive the pairs of palatalized tokens somewhat differently. Planned comparisons found significant differences in language groups' rating of all pairs except C^jijV/C^jjV and C^jijV/CV.[3] Post-hoc tests with the vowels show pairs in the context of /a/ were rated more dissimilar from those with /u/ or /i/ ($p < 0.001$). Rating averages were also significantly different between /u/ and /i/ ($p < 0.05$); pairs with /u/ were rated more different sounding than those with /i/. The rated similarity of the palatalized pairs by vowel is presented in Figure 5.

Post-hoc analyses confirmed our prediction that language experience affects the perceptual rated similarity of speech sounds. Native speakers of Russian generally hear greater contrast among degrees of palatalization; Russian listeners rated the sounds to be more different than American English listeners across most of the pairs. However, for two of the three pairs involving the functionally rare sequence C^jijV, Russians' ratings were not different from ratings given by American listeners. The near-merger pair C^jijV/C^jjV was rated as very similar by English and Russian listeners, although neither group of listeners rated them as identical (a rating of [1]). Conversely, the palatalization pairs involving CV, particularly the pairs C^jijV/CV and C^jjV/CV, were rated at near ceiling levels on the "very different" end of the scale.

3.2 *Experiment 4: Discrimination of palatalized consonants.*

The purpose of experiment 4 is to understand the underlying psycho-acoustic perception of palatalization. It is predicted that language background will not influence the results of this experiment as listeners will respond based on the

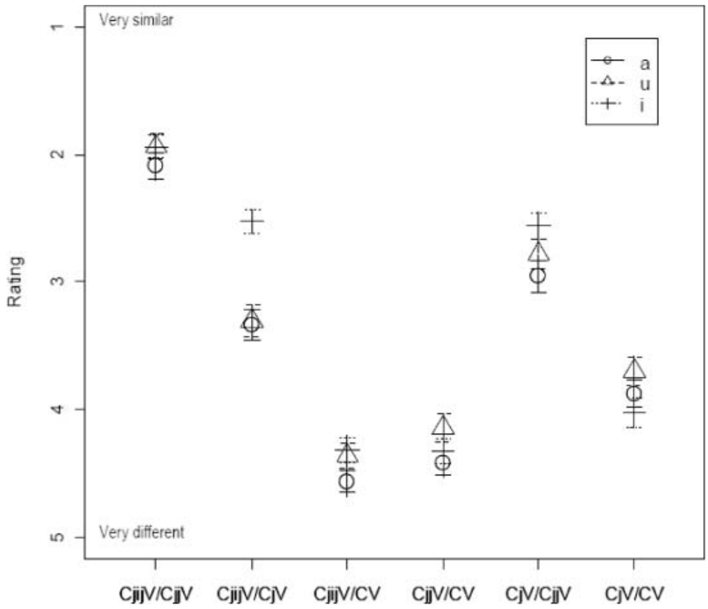


Figure 5. *Results of the palatalization by vowel interaction in experiment 3. Judgments of perceived similarity are shown on the vertical axis. A rating of [1] denotes a rating of "very similar" while a rating of [5] marks a pair as "very different." Palatalization pairs in an /a/ environment were rated as more different than those in an /i/ or /u/ environment. Pairs with /u/ were rated as more different sounding than those in an /i/ context. Error bars represent 95% confidence intervals.*

acoustic properties of the sounds as found for the voiceless fricatives in experiment 2.

3.2.1 *Method.* Fifteen native speakers of American English participated as listeners in experiment 4. Subjects reported no speech, language, or hearing disorders and were compensated $10 for their time.

Fourteen speakers of standard Russian were recruited to participate in experiment 4. Participants reported no speech, language, or hearing disorders and were compensated $10.

The data collected from one American English listener and three Russian listeners were excluded from the analysis due to excessively slow reaction times (mean reaction time > 700 ms.).[4]

This experiment used the same stimuli from experiment 3.

The same basic procedure described for experiment 2 was used in this experiment. Stimuli were blocked by consonant and vowel for a total of 21 blocks. In each block the four "same" pairs were presented three times each and the six "different" pairs were presented twice. This creates a total of 432 pairs.

3.2.2 *Results.* "Same" pairs of stimuli and incorrect responses were removed prior to the analysis. Log reaction time was entered as the dependent factor into a repeated measures ANOVA, and vowel, consonant, listener language, and palatalization pair were added as independent variables. The analysis revealed a main effect for palatalization pair ($F[5, 16] = 257.28$, $p < 0.001$) and consonant ($F[5, 4] = 13.63$, $p < 0.05$). There were significant interactions between language and palatalization pair ($F[5,16] = 4.52$, $p < 0.01$), palatalization pair and consonant ($F[25, 460] = 2.28$, $p < 0.001$), and palatalization pair and vowel ($F[10, 230] = 6.83$, $p < 0.001$). The ANOVA also returned a three-way interaction between palatalization pair, consonant, and vowel ($F[50, 1105] = 2.7$, $p < 0.001$).

Post-hoc tests found C^jijV/C^jjV pairs ($M = 645$, $SD = 178$) were labeled "different" more slowly than all other pairs, responses to C^jijV/C^jV and C^jV/C^jjV were slower than to C^jijV/CV, C^jjV/CV, and C^jV/CV. Planned comparisons of the language and palatalization interaction revealed that there were no significant differences between listener groups with the palatalization pairs. Figure 6 shows the main effect of palatalization pair for both Russian and

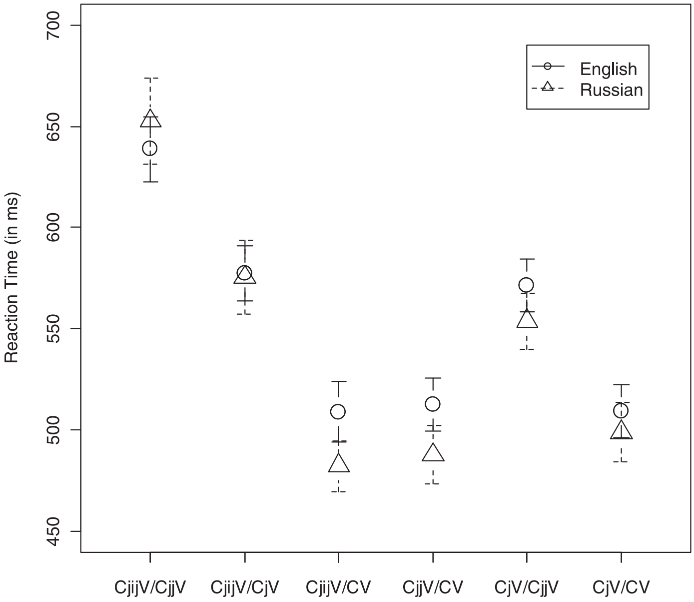


Figure 6. *Results of the palatalization main effect in experiment 4. The interaction of language by palatalization pair was not significant in planned comparisons. Error bars represent 95% confidence intervals.*

Table 1. *Mean response latencies and standard deviations in milliseconds for the six consonant contexts.*

	Mean	*SD*
/b/	554	170
/m/	545	159
/v/	549	172
/d/	563	156
/r/	537	161
/l/	519	152

English listeners, but the lack of differences between the two groups of listeners for the different pairs. The main difference between Russian and American listeners was that Americans were slower in responding to the easiest pairs (those with the overall fastest reaction times), a pattern that is quite different from the language by palatalization interaction found in the rating task.

Pair-wise post-hoc testing demonstrated that /l/-onset stimuli were responded to faster than the /v/, /m/, /b/, and /d/ stimuli ($p < 0.001$). Stimuli with /r/ onsets were also responded to faster than /d/ stimuli at a 0.001 level. Mean response latencies for the six consonant contexts are reported in Table 1.

For the vowel effects, post-hoc testing found reaction times in the context of /i/ was significantly slower than both /a/ ($p < 0.001$) and /u/ ($p < 0.05$) and reaction times with /u/ were also significantly slower than those with /a/ ($p < 0.05$). The reaction times in the vowel by pair interaction (Figure 7) mirror those seen in the rating data (Figure 5).

In this experiment, listeners' language background did not affect their responses to the acoustic properties of the stimuli. Listeners' processing of the stimuli was facilitated when one member of the pair did not have any palatalization. Having a simple CV token in a stimulus pair increased the acoustic dissimilarity of the pair and listeners responded faster. Consonant and vowel contexts were significant factors in listeners' response latencies as well.

3.3 *Discussion*

Experiments 3 and 4 have two main results that directly fall in line with the results from experiments 1 and 2. First, there was no clear effect of the listeners' language in the AX discrimination task. Russian and English listeners patterned similarly, which suggests they were providing responses based on raw auditory similarity and not using linguistic knowledge during the processing of the stimulus pairs. Again, in this second set of experiments we also found the pattern of reaction times in the discrimination task (experiment 4) mirroring the similarity rating responses in the rating task (experiment 3). Again, as was found in the first set of experiments, average reaction times and similarity

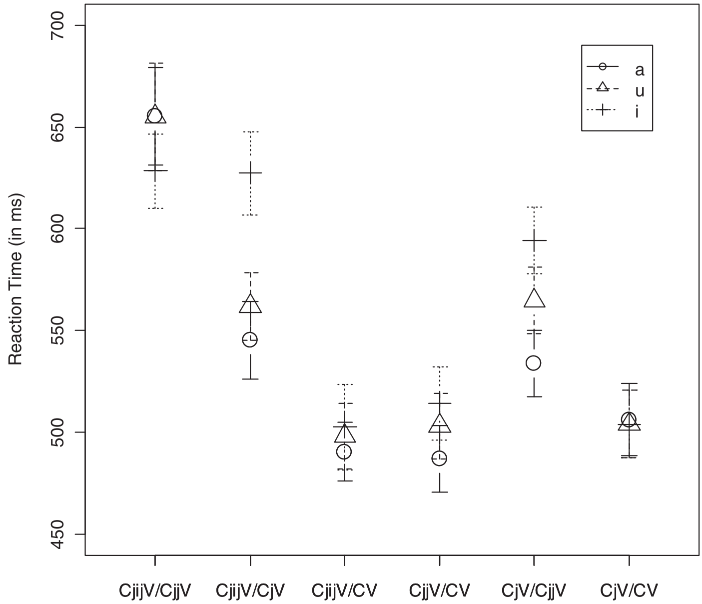


Figure 7. *Results of the palatalization by vowel interaction in experiment 4. Palatalization pairs in an /a/ environment were responded to more quickly than those in /i/ or /u/ contexts. Error bars represent 95% confidence intervals.*

rating judgments were strongly correlated [$t(4) = -5.7$, $p < 0.01$, Pearson's $R = -0.94$]. This suggests that the subjective similarity responses largely reflect the overall auditory similarity of the stimulus pairs. For example, the near-merger pair C^jijV/C^jjV was rated the most similar and also received the longest response latencies. In addition, the pairs with a plain CV member were responded to the fastest in the discrimination task and were rated as most different by both Russian and English listeners. Again, this finding indicates that the two tasks are not tapping dramatically different percepts, but the subtle differences in responses between the discrimination task and the rating task indicate comparatively more language-specific phonological knowledge affects responses in the rating tasks.

4. Depth of processing: from psycho-acoustic to linguistic knowledge

Other recent research has found effects of the listener's language in both a similarity rating task and in an AX discrimination task (Boomershine et al. 2008; McGuire 2007; Huang 2004; Huang and Johnson, under review), but in this study we found that the listener's language only affected the results of our similarity rating task—the AX discrimination reaction times showed no significant effect of listener's language. One experimental difference that may be important for understanding the difference between these earlier studies and those discussed here is that in our speeded discrimination experiments we strongly emphasized speed of responding. Therefore, our reaction times are much faster than those seen for American English and Spanish listeners in Boomershine et al. for consonant discrimination where a language effect was found in the AX discrimination task. The mean reaction time in the Boomershine et al. AX discrimination task was 781 ms. compared to 667 ms. for the Dutch experiment and 545 ms. for the Russian experiment. On the other hand, McGuire's listeners also produced fast reaction times (M = 578 ms.), but nonetheless showed a language effect.[5] More research is needed to understand the experimental situations that will result in language effects in experiments like these. In this section we attempt to address these contradictory results.

In the two speeded AX discrimination experiments reviewed above we report finding no language effect; there were no significant differences in the listener populations' responses in this behavioral task. In his dissertation, McGuire (2007) reports finding a language effect using the same experimental paradigm. The stimuli in his experiment were modified naturally produced tokens of [ʂa] and [ɕa] from a native speaker of Polish. For the AX discrimination task, the stimuli were re-synthesized combinations of correlated and conflicting fricative-vowel sequences such that the original productions [ʂa] = f0v0 and [ɕa] = f9v9 created the additional stimuli f0v9 and f9v0 with the spectral information from one fricative, but the formant transitions going into the vowel suggesting the other place of articulation. These pairs were presented to American English and Mandarin listeners in all possible combinations. (The [ʂa]/[ɕa] contrast is used in Mandarin, while in English all of the stimuli sound like variants of /ʃa/ "shah".) A repeated measures ANOVA using log reaction times found a significant interaction between pair and language. Post-hoc tests revealed significant differences between language groups for the pairs f0v0/f0v9 and f0v9/f9v0. This can be seen in Figure 8 (the equivalent of Figure 3.3 in McGuire [2007]).

Fox (1984) found that if reaction times were grouped into three bins—those less than 500ms., between 500 and 800 ms., and greater than 800 ms.—the development of the Ganong effect becomes apparent. The Ganong effect (Ganong 1980) is the lexical bias toward, for example, *task* in a VOT varying *task-dask* continuum. Ambiguous tokens in the middle of the continuum will be perceived as *task* because of its lexical status while *dask* is a non-word. In Fox's data, the Ganong effect was seen for responses with reaction time greater than 800 ms. Faster responses showed no evidence of lexical bias. Such a result suggests that a stage of phonetic processing takes place prior to lexical processing. If we consider this alongside the work of Pisoni (1973) and Pisoni and Tash (1974), then we may hypothesize that a stage of auditory

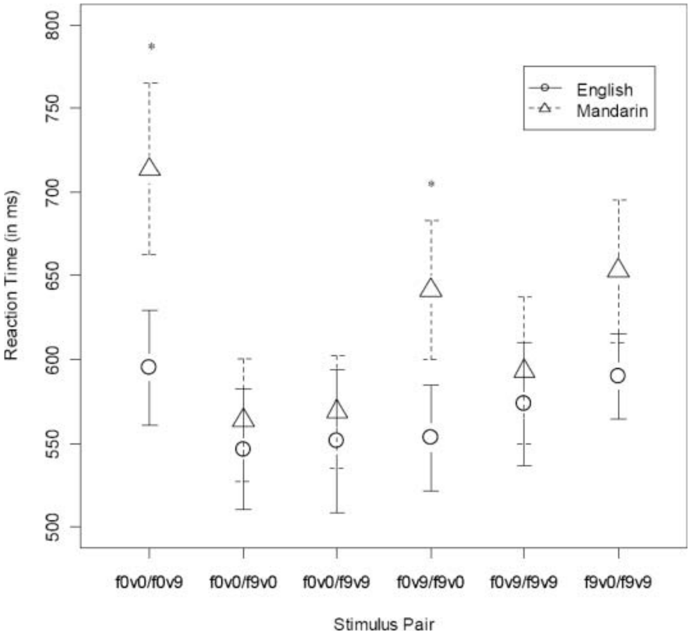


Figure 8. *Results reported in McGuire (2007) demonstrating a significant language by pair interaction. In the stimulus pairs f0v0 is the fricative vowel sequences [ʂa], f9v9 is the fricative vowel sequences [ɕa], and f0v9 and f9v0 are cross-spliced sequences of the fricative vowels in [ʂa] and [ɕa]. This figure corresponds to his Figure 3.3. Error bars represent 95% confidence intervals.*

perception precedes the phonetic processing stage. Evidence for this hypothesis would come in the way of more language influences surfacing in speeded AX discrimination tasks with longer response latencies. To explore this hypothesis, we conducted a median split analysis with our voiceless fricative and palatalization data. For both of these data sets, slow and fast responses were determined by calculating the median reaction time for each participant for each AX comparison.[6]

Figure 9 shows a median split of the voiceless fricative data. It is clear in this figure that within the fast set of responses, Dutch listeners performed overall much more quickly. However, the fast responses do not provide insight into the perceived phonological similarity that mirrors differences between Dutch and English phonology (i.e., the absence of /θ/ and the contested status of /ʃ/). The interesting results in terms of speech perception are in the set of slow responses. Recall that an interesting interaction in the Dutch rating task was with the s/θ pair. Dutch listeners rated this pair as more similar sounding than English listeners. In the slow responses in the median split, we see evi-

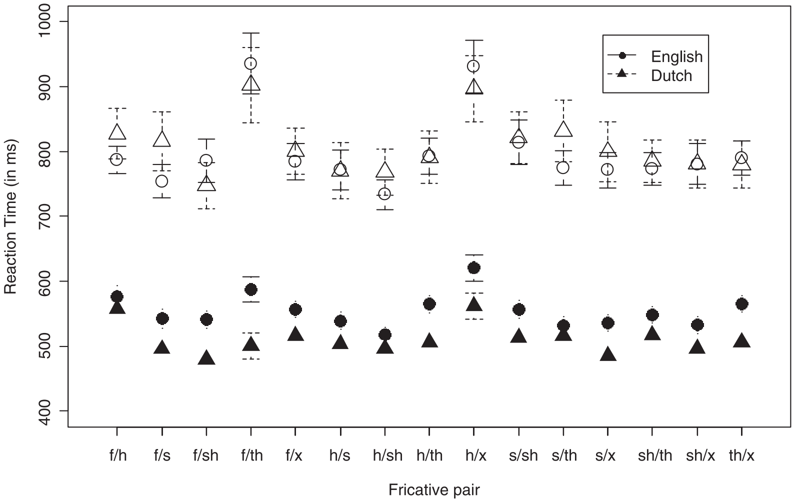


Figure 9. *Results of median split analysis with Dutch and English listeners for the fricative AX discrimination data. Error bars represent 95% confidence intervals.*

dence of this subjective rating which was not apparent in the pooled data; Dutch listeners are responding slower to this pair.

In Figure 10, the median split of the Russian palatalization data is presented. In the fast responses, Russian and English listeners are behaving similarly. Language differences become apparent in the slow responses. We see differences between the listener populations arising in these longer reaction times as the mean response times become more separated. These differences largely arise in Russian listeners taking longer to identify C^jijV/C^jjV, the low functional load pair, as "different" and being quicker to identify C^jijV/CV as "different". This longer response latency in identifying C^jijV/C^jjV as "different" for Russian listeners is what we would predict in a more language-influenced response.

With our data, the median split analysis seems to suggest that the longer response latencies demonstrate more effects of language experience. This is even when no language effect was found in the pooled data. We also performed a median split analysis with the McGuire (2007) data. As briefly described above, the stimulus pair f0v0 is the fricative vowel sequences [ʂa] and f9v9 is the fricative vowel sequences [ɕa]. The pairs f0v9 and f9v0 are cross-spliced sequences of the fricatives and vowels in [ʂa] and [ɕa]. Compare this median split analysis in Figure 11 to that of Figure 9. In the fast responses, the language effects are reduced considerably, particularly for the f0v9/f9v9 pair, although they do not disappear completely. In these data, Mandarin listeners are able to make greater use of the fricative noise information in

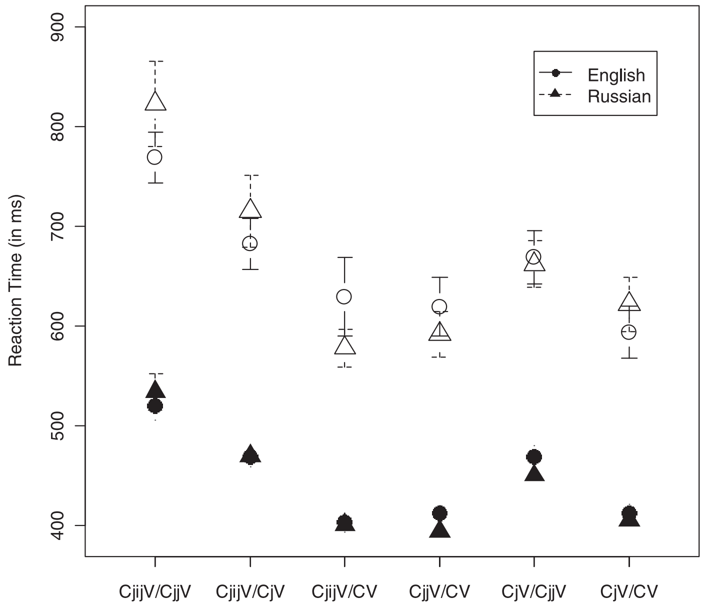


Figure 10. *Russian palatalization results with the median split analysis. Error bars represent 95% confidence intervals.*

discriminating pairs while English listeners attend more to the cues in the formant transitions (McGuire 2007). The fact that in these CV pairs the cues that Mandarin listeners are sensitive to occur earlier in the syllable may allow for the effect of language experience to appear in the fast set of responses.

As mentioned above, Boomershine et al. (2008) report a consonant pair by language effect when examining allophonic relationships of [d], [ð], and [ɾ] with native speakers of Spanish and American English. A median split analysis with this data would not be very insightful because of the overall longer reaction times in this data set. The median split analyses on the fricative data, the palatalization data, and the McGuire data indicate, however, that longer response latencies demonstrate more of a language effect than the fast sets of responses. The increased depth of processing that inherently accompanies delayed responses in an AX discrimination paradigm allows for more language knowledge to influence responses. This suggests that on some level a stage of auditory perception that is unaffiliated with language experience precedes language-specific processing.

A brief caveat is in order. In our research, we follow the assumption that reaction time is a proxy for processing time. One issue with this assumption is that speech processing may involve interactive and cascaded processing, as opposed to serial processing (McClelland and Elman 1986). This would allow

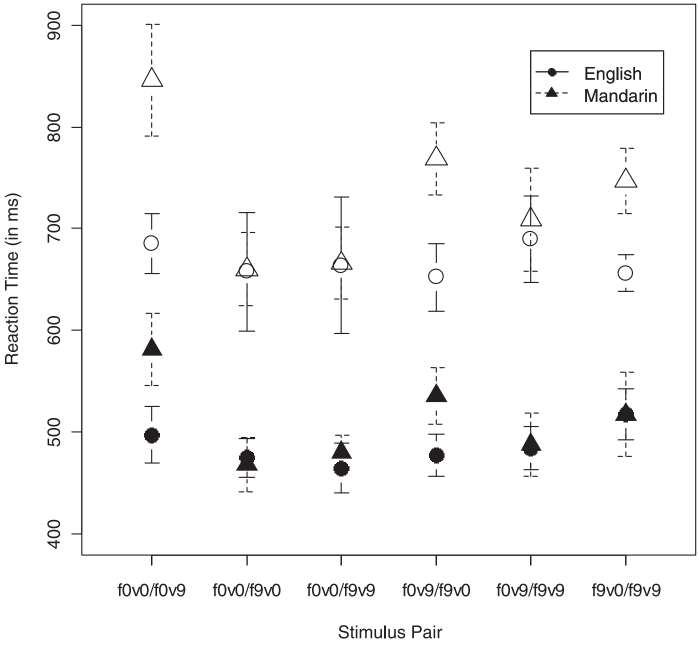


Figure 11. *Median split analysis with Mandarin and English perception data from McGuire (2007). Error bars represent 95% confidence intervals.*

simultaneous processing of acoustic and language-specific phonological information, allowing language effects to surface quickly. In addition, another issue is that the processing of natural speech may be so quick and efficient that language effects enter into the quickest responses.

5. Discussion and Conclusion

In this paper we provide evidence that psycho-acoustic perception can to some extent be evaluated apart from language-specific perception. This supports the work of Pisoni (1973), Pisoni and Tash (1974), and Fox (1984) who argue for multiple stages in speech perception in which a stage of auditory perception precedes language specific perception. Previous work by Boomershine et al. (2008) and McGuire (2007) did not find this stage of linguistically unbiased perception when the same methodology that we used in experiments 1 and 3 was employed. By looking at fast and slow responses within these data sets, however, we find that faster responses in a speeded AX discrimination task show less influence of language experience while longer response times provide evidence of language experience. We consider this to be evidence that a stage of auditory perception that is less influenced by language experience may be accessed prior to language-specific phonetic or lexical processing. Regarding this difference between phonetic and lexical processing, Johnson (2004) models responses to the perception tasks discussed in this paper using probability from the lexicon. Under this model, all linguistic memory and knowledge is intrinsically tied to the lexicon where language-specific phonological patterns emerge.

In this paper we argue that auditory processing can be evaluated apart from language-specific processing in fast response times. This argument by no means contradicts other work reviewed in the introduction that demonstrates neurological differences in speech processing (eg. Krishnan et al. 2005). It is crucial to recognize, however, that not all language processing can take place in the brain stem because perceptual categorization can change as a function of the language set (Caramazza et al. 1973; Elman et al. 1977) and in response the social expectations about the talker (Johnson et al. 1999). On a behavioral level, we have provided some evidence that listeners are able to make judgments about particular sounds, in our case voiceless fricatives and degrees of consonant palatalization, that are not affected by language experience. Theories of speech perception that envision language pervading the entire perceptual system from the cochlea to the brainstem to the auditory cortex have no way to account for the lack of a difference between listener groups in our data. Such theories predict that all levels of perception are influenced by language experience. The experimental data provided here suggests that in a speech behavioral discrimination task, listeners from different language groups can evaluate speech sounds using raw acoustic similarity. Exploring these types of perceptual differences has implications for phonological models of the role of perceptual similarity such as the P-map (Steriade 2001) and for understanding perceptually-motivated sound change.

Acknowledgements

This research has been supported by NIH grant R01 DC004421-05 awarded to the second author. Earlier versions of work reported here were presented at the Fall 2006 and the Spring 2007 meetings of the Acoustical Society of America and the 2007 ICPhS meeting. Parts of this research have been published in Johnson (2004), Babel and Johnson (2007), and Johnson and Babel (2010). Thanks to Grant McGuire, Amanda Boomershine, Kathleen Currie Hall, and Elizabeth Hume for sharing their raw data with us.

Correspondence e-mail address: mbabel@interchange.ubc.ca
keithjohnson@berkeley.edu

Notes

1. As noted by a reviewer, this difference in reaction times could stem from Dutch listeners receiving monetary compensation for their participant, while American English listeners received course credit. Paid participants may have been more motivated to follow task instructions for this reason.
2. See Halle (1959) for an early acoustic study of consonant palatalization in Russian, and Diehm (1998) for an acoustic study of the sequences in (1).
3. The reader may note increased variability among the Russian participants. This may be due to Russian listeners coming from a wider range of ages, while English listeners were all of traditional undergraduate age.
4. In a post-experiment debriefing, the American English listener whose data was removed from the analysis commented that he had found a way to "beat the experiment". We have interpreted this as a strategy that included slowing the response time. Therefore, this participant and all others who had mean reaction times greater than his were removed from the data set prior to the analysis.
5. A partial explanation of McGuire's results may exist in the stimuli themselves. The stimuli were fricative-vowel sequences where auditory differences between the tokens are apparent from the onset.
6. We did not conduct any statistical analyses on the median split data because in splitting the data, we have reduced the power of any statistical test. We feel the descriptive statistics (means, standard deviations) presented in the figures adequately depict our position.

References

Baayen, R. Harald, Richard Piepenbrock, & Hedderik van Rijn. 1993. The CELEX lexical database, CD-ROM. Linguistic Data Consortium, University of Pennsylvania.

Babel, Molly & Keith Johnson. 2007. Cross-linguistic differences in the perception of palatalization. *Proceedings of the 16th International Congress of the Phonetic Sciences*.

Best, Catherine T. 1995. A direct realist perspective on cross-language speech perception. In Winifred Strange (ed.), *Speech perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-language Speech Research*, York: Timonium, MD. 167–200.

Best, Catherine T. & Robert A. Avery. 1999. Left-hemisphere advantage for click consonants is determined by linguistic significance and experience, *Psychological Science* 10. 65–69.

Best, Catherine T., Gerald W. McRoberts, & Elizabeth Goodell. 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(2). 775–794.

Booij, Geert. 1995. *The Phonology of Dutch*. Oxford: Oxford University Press.

Boomershine, Amanda, Kathleen Currie Hall, Elizabeth Hume, & Keith Johnson. 2008. The impact of allophony vs. contrast on speech perception. In Peter Avery, Elan Dresher & Keren Rice (eds.), *Phonological Contrast: Perception and Acquisition*. New York: Mouton de Gruyter.

Caramazza, Alfonso, Grace H. Yeni-Komshian, Edgar B. Zurif, & Ettore Carbone. 1973. The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America* 542. 421–428.

Cohen, A., C. L. Ebeling, K. Fokkema, & A. G. F. van Holk. 1972. *Fonologie van het Nederlands en het Fries: inleiding tot de moderne klankleer* [Phonology of the Netherlands including the language of Frisia: Introduction to the modern acoustics]. 's-Gravenhage: Martinus Nijhoff.

Diehm, Erin. 1998. *Gestures and linguistic function in learning Russian: Production and perception studies of Russian palatalized consonants*. Ph.D. Dissertation, The Ohio State University.

Dupoux, Emmanuel, Christophe Pallier, Nuria Sebastian, & Jacques Mehler. 1997. A destressing "deafness" in French? *Journal of Memory and Language* 36. 406–421.

Elman, Jeffrey L., Randy L. Diehl, & Susan E. Buchwald. 1977. Perceptual switching in bilinguals. *Journal of the Acoustical Society of America* 62. 971–974.

Flege, James E. 1995. Second language speech learning: Theory, findings, and problems. In Winifred Strange (ed.), *Speech perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-language Speech Research.* York: Timonium, MD. 167–200.

Flege, James E., Naoyuki Takagi, & Virginia Mann. 1996. Lexical familiarity and English-language experience affect Japanese adults' perception of /r/ and /l/. *Journal of the Acoustical Society of America* 99. 1161–1173.

Fox, Robert A. 1984. Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance* 10. 526–540.

Fujisaki, H. & T. Kawashima. 1969. On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute*, Vol. 28, Faculty of Engineering, University of Tokyo, Tokyo. 67–73.

Fujisaki, H. & T. Kawashima. 1970. Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute*, Vol. 29, Faculty of Engineering, University of Tokyo, Tokyo. 207–214.

Gandour, Jack T. 1983. Tone perception in Far Eastern languages. *Journal of Phonetics* 11. 149–176.

Ganong, William F. 1980. Phonetic categorization in auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 6. 110–125.

Gerrits, Ellen & M. E. H. Schouten. 2004. Categorical perception depends on the discrimination task. *Perception & Psychophyiscs* 66. 363–376.

Goto, Hiromu. 1971. Auditory perception by normal Japanese adults of the sounds "l" and "r." *Neuropsychologia* 9. 317–323.

De Groot, A. W. 1968. *Inleiding in de algemene taalwetenschap* [Introduction of the common linguistics]. Groningen: Wolters Noordhof.

Halle, Morris (1959). *The Sound Pattern of Russian: A Linguistic and Acoustical Investigation*. Mouton: 's-Gravenhage.

Harnsberger, James. 2001. The perception of Malayalam nasal consonants by Marathi, Punjabi, Tamil, Oriya, Bengali, and American English listeners: A multidimensional scaling analysis. *Journal of Phonetics* 29. 303–327.

Huang, Tsan. 2004. *Language specificity in auditory perception of Chinese tones*. Ph. D. dissertation, The Ohio State University.

Huang, T. & Johnson, K. (under review) Language-specificity in speech perception: a study on Mandarin tones with Chinese- and English-speaking listeners.

Hume, Elizabeth & Keith Johnson. 2003. The impact of partial phonological contrast on speech perception. *Proceedings of the 15th International Congress of Phonetic Sciences*. 2385–2388.

Johnson, Keith. 2004. Cross-linguistic perceptual differences emerge from the lexicon. In Augustine Agwuele, Willis Warren & Sang-Hoon Park (eds.), *Proceedings of the 2003 Texas Linguistics Society Conference: Coarticulation in speech production and perception*. Sommerville, MA: Cascadilla Press. 26–41.

Johnson, K. & Babel, M. 2010. On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics*, 38, 127–136.

Johnson, Keith, Elizabeth Strand, & Maria Paola D'Imperio. 1999. Auditory-visual integration of talker gender in vowel production. *Journal of Phonetics*, 274. 359–384.

Jones, Daniel & Dennis Ward. 1969. The Phonetics of Russian. Cambridge University Press.

Krishnan, Ananthanarayan, Yisheng Xu, Jackson Gandour, Peter Cariani. 2005. Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research* 25. 161–168.

Kuhl, Patricia K., Karen A. Williams, Francisco Lacerda, Kenneth Stevens, & Björn Lindblom. 1992. Linguistic experiences alter phonetic perception in infants by 6 months of age. *Science* 255. 606–608.

Lively, Scott E., John S. Logan, & David B. Pisoni. 1993. Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America* 94. 1242–1255.

Logan, John S., Scott E. Lively, David B. Pisoni. 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America* 89. 874–886.

MacKain, Kristin S., Catherine T. Best, & Winifred Strange. 1981. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics* 2. 369–390.

McClelland, James L. & Jeffery L. Elman. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18. 1–86.

McGuire, Grant. 2007. *Phonetic category learning*. Ph.D. Dissertation, The Ohio State University.

Mielke, Jeff. 2008. *The Emergence of Distinctive Features*. Oxford University Press.

Miyawaki, Kuniko, Winifred Strange, Robert Verbrugge, Alvin M. Liberman, James J. Jenkins, & Osamu Fujimura. 1975. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics* 18. 331–340.

Näätänen, Risto, Anne Lehtoski, Mietta Lennes, Marie Cheour, Minna Houtilainen, Antti Iivonev, Martti Vainio, Paavo Alku, Risto Iimoniemi, Aavo Luuk, Jüri Allik, Janne Sinkkonen, & Kimmo Alho. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385. 432–434.

Nooteboom, S. G. and Cohen, A. 1984. *Spreken en verstaan. Een nieuwe inleiding tot de experimentele fonetiek*. Assen: Van Gorcum.

Pisoni, David B. 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics* 132. 253–260.

Pisoni, David B. & Jeffrey Tash. 1974. Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics* 152. 285–290.

Schneider, Walter, Amy Eschman, & Anthony Zuccolotto. 2002. *E-Prime: User's Guide, version 1.0*. Psychology Software Tools.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume & Keith Johnson (eds.), *The Role of Speech Perception in Phonology*, New York: Academic Press. 219–250.

Strange, Winifred & Sibylla Dittmann. 1984. Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics* 36. 131–145.

Yamada, Reiko, Yoh'ichi Tohkura, & Noriko Kobayashi. 1992. Effect of word familiarity on non-native phoneme perception: Identification of English /r/, /l/, and /w/ by native speakers of Japanese. In Allan James & Jonathan Leather (eds.), *Second Language Speech*, The Hague: Mouton de Gruyter.