Categorization of non-native liquid contrasts by Cantonese, Japanese, Korean, and Mandarin listeners

James Smith and Alexei Kochetov
University of Toronto

This paper reports the results of a perceptual experiment involving the categorization of Russian liquid consonants by native speakers of Korean, Taiwanese Mandarin, Japanese, and Cantonese. Listeners were presented with stimuli containing plain and palatalized laterals (/l/ and /lʲ/) and plain and palatalized rhotics (/r/ and /rʲ/). The categorization task confirmed the results of an earlier discrimination experiment and revealed patterns of non-native perception that extend beyond simple segment-to-segment mappings.

1. Introduction

It is well accepted that native language experience affects the ability to perceive sounds that are not in the phonological inventory of the listener. Research in non-native speech perception has established that listeners have little difficulty perceiving native sounds that are contrastive in their native language, and adult listeners are most sensitive to their own native contrasts. Listeners commonly have problems identifying and discriminating sounds that are either not in the phonological inventory of their language or are not contrastive in their language (Best, McRoberts, & Sithole 1988).

Early investigations explained perception performance solely in terms of interference from native phonological contrasts (Miyawaki, Strange, Verbrugge, Liberman, Jenkins, and Fujimura 1975; MacKain, Best, & Strange 1981). The main finding of this early research is that perceptual difficulty is not universal and varies across languages, contrasts, and individuals, often despite consistent production by second-language learners (MacKain et al. 1981; Sheldon & Strange 1982). Some non-native contrasts are easier to perceive than others (Best, et al. 1988), even without prior experience, and perception improves with training (MacKain et al. 1981). Perceptual difficulty varies with syllable position and phonetic context (Mochizuki 1981; Sheldon & Strange 1982). Perception improves with early exposure to the non-native contrast before puberty (Goto 1971; Mochizuki, 1981; Werker & Tees 1984; Flege 1987). A key insight of more recent research has been that perception does not rely only on the properties of abstract phonological system. Listeners also respond to language-specific phonetic details at the same time as phonological structures in their

* We wish to thank the audiences of the 3rd annual Toronto Workshop on East Asian Languages and the Toronto Summer 2008 Phonetics/Phonology Workshop for their helpful questions and comments. Thanks also to Jaehee Bak for his help with the phonetic transcription of the Korean responses. Research for this study was supported by SSHRC grant 410-2006-1006 to Alexei Kochetov. Both authors contributed equally to the study.

Toronto Working Papers in Linguistics (TWPL), Volume 34
© 2009 James Smith and Alexei Kochetov
native system (Polka & Strange 1985; Underbakke, Polka, Gottfried, & Strange 1988; Takagi & Mann 1995). This work has focused on the internal structure of phonetic categories by making use of non-native contrasts that differ allophonically in the listener’s native system (Lisker & Abramson 1970; Kuhl 1991; Best & Strange 1992; Iverson & Kuhl 1995; Ingram & Park 1998).

Several researchers have investigated production and perception of lateral and rhotic consonants in second language learning, mainly of English sounds by Japanese speakers (Henly & Sheldon 1986; Best & Strange 1992; Takagi 1995; Ingram & Park 1998; Guion, Flege, Akahane-Yamada, & Pruitt 2000; Aoyama, Flege, Guion, Akahane-Yamada, and Yamada 2004). Borden, Gerber & Milsark (1983) provide one of the few early accounts of non-native perception of English /r – l/ by Korean listeners. Henly & Sheldon (1986) carried out one of the few perceptual studies that included Cantonese listeners, comparing the perceptual performance of Cantonese listeners to that of Japanese listeners in the identification of English /r/ and /l/ in various contexts. Schmidt (1996) explored the perceptual relationship between the consonants of Korean and English and examined language specific phonetic influences on perceived similarity between the two languages. Her results were consistent with Polka’s (1991) suggestion that acoustic-phonetic, articulatory phonetic, and phonemic factors together influence perception and perceptual variability, supporting theories of speech perception that claim perception is context-specific rather than phonemically abstract. Ingram & Park (1998) compared identification and discrimination performance for English /r/ and /l/ by both Japanese and Korean listeners. They tested the hypothesis that perception of non-native contrasts involves simultaneous operation of several levels of speech processing from the auditory-acoustic to the linguistic.

All theoretical models of non-native speech perception acknowledge systematic effects of phonetic experience on perception, and assume that the native phonological system affects discrimination of non-native contrasts (Best, McRoberts, & Goodell 2001). However, most research in non-native speech perception has tended to focus on perception of only one or two non-native (mainly English) contrasts by listeners from one or two language groups. To our knowledge, there have been few studies to date of perception of non-native consonants by listeners of several languages (Harnsberger 2000, 2001), and hardly any studies that examined cross-language perception of Russian consonants (Kochetov & So 2007).

This study aims to answer two questions. First, how does a listener’s native language background affect their categorization and discrimination of the non-native contrasts? Second, what is the role of native contrasts (segmental or segment-sequence) and phonetic differences in non-native perception? To address these questions, we examine how native speakers of four East Asian languages – Korean, Taiwanese Mandarin, Japanese, and Cantonese – categorize a complex set of non-native lateral-rhotic contrasts – Russian consonants /l/, /lj/, /r/, and /rj/. The goal of this experiment is also to clarify the patterns of discrimination of the same contrasts by listeners of these languages obtained in an earlier experiment (Kochetov & Smith, forthcoming).

2. Cross-language liquid contrasts

The choice of Russian plain and palatalized lateral and rhotic consonants was motivated by two goals. First, we wanted investigate the non-native perception of the lateral-rhotic contrast, as both laterals and rhotics exist in some of the listeners’ native languages (phonemically in Mandarin and allophonically in Korean), while either laterals or rhotics are found in other languages (/r/ in Japanese and /l/ in Cantonese). Second, we wanted to explore the perception of a contrast that is absent phonemically from many of the investigated non-native languages: the contrast between plain and palatalized consonants (and liquids in particular). In this respect, Russian presents an interesting case study, as the language has a complex two-way lateral-rhotic consonant contrast, including plain and palatalized laterals /l/ and /lj/, and plain and palatalized rhotics /r/ and /rj/.
examples: xalat ‘gown’, parad ‘parade’, paljat ‘(they) scorch’, and narjad ‘costume’. Phonetically, the Russian /r/ is an alveolar tap [ɾ] or trill [r] intervocally, and /l/ is a velarized lateral approximant [ɿ]. /l/, /ɾ/, and /ɿ/ have apico-alveolar articulation, while /ɿ/ is described as apico-alveoprepalatal (Bolla 1981).

The languages under consideration in this study vary greatly in their lateral-rhotic contrasts, with the segment inventory falling into one of three classes: no contrast in Cantonese and Japanese, one-way segmental contrast in Mandarin, and segment-sequence in Korean. These differences are likely to play a significant role in the perception of the Russian consonants, and, therefore, require further discussion. Cantonese has only a lateral approximant consonant /l/, and no rhotic consonant. The sound is described as apicolateral dentalvelar or alveolar that does not show noticeable allophonic variation. The language has no phonemic palatalization, and no sequences of consonant + palatal glide or diphthong (*/C+j/ and */C+ia/) (Matthews and Yip 1994; Zee 2005). Japanese also has no lateral-rhotic contrast, with the single rhotic phoneme /r/, typically realized as a postalveolar flap [ɿ], among other realizations that are not clearly conditioned by phonological environment: [ɿ], [ɿ], [d], [ɿ], [r] or [ɿ] (Ingram & Park 1998; Okada 2005; Best & Strange 1992). Japanese also has phonemic palatalized consonants, including /ɿ/ (which can be also phonemically analyzed as a sequence /C+j/; Okada 2005). Mandarin Chinese contrasts a voiced lateral approximant /l/ and a retroflex approximant /ɿ/. Mandarin /ɿ/ is apical dentalvelar or apical alveolar, while /ɿ/ is an apical post-alveolar approximant [ɿ] with no lip rounding, and possible frication noise (Lee & Zee 2003). The language permits syllable onsets consisting of a consonant + glide or diphthong (/C+j/ or /C+ia/, depending on the analysis), in which the consonants are phonetically realized as palatalized (Duanmu 2007). Among the liquids, however, only /l/ can occur in such a sequence (i.e. /l+j/, but not */r+j/). Korean has a single liquid phoneme, /l/, which is realized as an alveolar flap [ɾ] intervocally (syllable initially) or before /h/, and as a lateral alveolar approximant [ɿ] syllable finally. It should be noted, however, that the liquid can occur as a geminate /ɿ/, realized as [ɿ] or [ɿ] depending on vowel context, thus potentially creating a segment-sequence rhotic/lateral contrast V[ɿ], [ɿ]V vs. V.[ɾ]V. Further, Korean phonotactics permits sequences of single or geminate /l/ with [ɿ]-initial diphthongs, as well as sequences /Ci+V/ (Sohn 1994; Ingram & Park 1998; Lee 2005).

In summary, Japanese and Cantonese show no lateral-rhotic contrast at all, Mandarin has a one-way contrast, and Korean shows a segment-sequence contrast involving the single phoneme /l/. The languages also vary in presence or absence of phonemic palatalization or similar consonant + glide or diphthong sequences. This variety of contrasts is expected to influence the perception of the two-way Russian liquid contrast by non-native listeners.

3. An earlier discrimination experiment

A perceptual experiment reported in Kochetov & Smith (forthcoming) investigated the discrimination of Russian liquid contrasts by non-native listeners from six language groups. This section summarizes the main findings of that study, and highlights research questions that motivate the categorization experiment presented in this paper. Listeners, 71 native speakers of Korean, Taiwanese Mandarin, Cantonese, Japanese, and Russian, were presented with randomized pairs of words containing one of the four Russian lateral and rhotic consonants in VCV context. Stimuli were two-syllable nonsense words of the form [taCap] where C is on of [ɿ], [ɿ], [r] or [ɾ]. The experiment used a same-different AX discrimination task designed to measure listener sensitivity to the lateral vs. rhotic and plain vs. palatalized contrasts. Listeners’ ‘different’ responses were analyzed in terms sensitivity (measured as D'); Macmillan & Creelman 2005) to lateral vs. rhotic and palatalized vs.
nonpalatalized contrasts and were compared to responses of native Russian listeners. Performance in the discrimination task was expected to vary according to the native contrasts of each language group. For the lateral-rhotic contrast, Cantonese and Japanese listeners were expected to show near chance discrimination, given the lack of contrast in these languages; Mandarin listeners were expected to show relatively good discrimination, given native /l/ vs. /r/ contrast. Expectations for Korean listeners were less clear: they could either show relatively poor discrimination, given the lack of segmental lateral-rhotic contrast, or relatively good discrimination if they categorize the contrast as segment-sequence. For the plain-palatalized contrast, Japanese listeners were expected to perform well, given the similar native contrast, while the other listener groups were expected to perform somewhat poorer given the apparent lack of such contrast. However, native knowledge of similar C+j or C+iα sequences (for Korean and Mandarin listeners) could facilitate the discrimination of the contrast.

The results for the lateral-rhotic contrast, summarized in Figure 1a, showed that Cantonese, Japanese, and Mandarin listeners discriminated the contrast significantly poorer than Korean and Russian listeners, while Korean listeners’ discrimination was not as good as that of Russian listeners. The near-chance discrimination of both contrasts by Japanese and Cantonese groups was expected, since the two languages do not have the lateral-rhotic contrast. Interestingly, Cantonese and Japanese listeners performed well below chance (D' ≤ 0) with palatalized stimulus pairs. Iverson et al. (2003) observe that unlike native English speakers, Japanese listeners are more sensitive to F2 variation, which is largely irrelevant to the /l/ – /r/ categorization. This F2 sensitivity is an unreliable cue to categorization that distorts the perceptual space for Japanese listeners, and could interfere with correct category identification. Since F2 is a robust cue for palatalization, it is also possible that increased sensitivity to F2 not only accounts for poor Japanese performance with the /r/ – /l/ contrast, but may also have aided discrimination of the plain-palatalized contrast (discussed below). With respect to Cantonese, Henly and Sheldon (1986) observe that word-final English /l/ is difficult for Cantonese listeners because it is typically velarized. Similarly, it is likely that the velarized Russian intervocalic /l/ provided a poor “template” for the clear Cantonese /l/, making it harder for Cantonese listeners to perceive. However, since Cantonese lacks a rhotic consonant; evidence from other perceptual studies suggests that lack of contrast can be perceptually salient if listeners attend to purely auditory cues rather than phonological or phonetic knowledge (Best et al. 2001). This could account for the slightly better discrimination by Cantonese listeners of the plain lateral vs. rhotic contrast compared to Mandarin and Japanese listeners. Mandarin listeners, whose language contrasts laterals and rhotics, performed at near chance, well below expectation. One reason for the overall poor discrimination of the contrast might be that Mandarin listeners did not perceive the Russian trill /r/ as a rhotic, or similar enough to their native approximant category. The sensitivity of Korean listeners to the lateral vs. rhotic contrast showed somewhat unexpectedly good discrimination of both contrasts, with sensitivity to the lateral-rhotic contrast approaching the Russian control group. Though the Korean rhotic [ɾ] is an allophone of /l/ intervocally, and might be expected to be perceptually confused with [l], it appears that the Russian /l/ was assimilated to a consonant sequence, a geminate /ll/ (while the Russian /r/ was assimilated to a single /l/), thus suggesting that segment vs. sequence contrasts are relevant to non-native perception, and that perception involves more than a simple one-to-one mapping of non-native segment to a native category. Finally, the results also showed that discrimination was significantly better for plain pairs (/l/ vs. /r/) than for palatalized pairs (/lj/ vs. /rj/), consistent for all language groups, including native Russian listeners. Palatalization thus appears to hinder lateral-rhotic discrimination by reducing the perceptual salience of the contrast. This can be attributed to the greater acoustic difference between the non-palatalized consonants (lowering F2 for /l/ due to its velarization) compared to the palatalized pairs (raising F2 for both /lj/
vs. /r/), as well as to the greater familiarity of non-native speakers with the non-palatalized lateral-rhotic contrasts.

![Diagram of D-Prime values for different contrasts by language group.](image)

Figure 1. Mean D’ (a) for lateral vs. rhotic and (b) for plain vs. palatalized contrasts by language group

Turning to the discrimination of the plain-palatalized contrast (Figure 1b), accuracy discriminating palatalized from nonpalatalized stimuli showed considerably less variation across language groups than for discrimination of lateral vs. rhotic pairs. Discrimination rate for all languages was uniformly at near ceiling. Although discrimination performance differences between languages were not statistically significant, it is notable that Japanese listeners were nearly as accurate as Russian listeners in discriminating plain from palatalized stimuli, even performing slightly better discriminating palatalization in the rhotic pair. This is expected, given that Japanese has the plain-palatalized contrast, and, as observed above, sensitivity to F2 may have aided discrimination. All language groups showed higher D’ values discriminating lateral stimuli pairs, with slightly lower values for the rhotic stimuli pairs. In general, discrimination was significantly more accurate with lateral pairs ([talap/talajap]) than with rhotic pairs ([tarap/tarajap]), likely due to the enhancing role of velarization in the lateral contrast.

In summary, the discrimination experiment revealed significant differences across listener groups in the perception of the lateral vs. rhotic contrast, ranging from near-native performance to the apparent lack of any sensitivity to the contrast. The discrimination of the plain vs. palatalized contrast, however, was relatively similar across the groups, consistently showing high sensitivity to the contrast. The results were largely, but not fully consistent with predictions based on native language
contrasts. The discrimination experiment raises several questions that motivate the present study. All language groups showed relatively good discrimination of the plain vs. palatalized contrast, though only Japanese has a phonemic palatalized consonant comparable to those of Russian (but see above on alternative analyses for some of the languages). Why was the discrimination performance of this contrast relatively good across all language groups? For languages with no phonemic palatalization, better overall discrimination performance for the plain-palatalized contrast was likely due to the high acoustic salience of the contrast, given a difference in F2 of approximately 860 Hz for laterals and 200 Hz for rhotics in the stimuli (measured at the offset of the consonant). It also appears that listeners mapped palatalized consonants onto similar native consonant + glide or diphthong sequences (which phonetically may be realized as palatalized or palatal consonants). Since it is not possible from the discrimination results to tell exactly what native segments listeners are assimilating the non-native sounds to, the experiment reported in this study was designed to reveal categorization patterns among the language groups. One possible hypothesis is that perceptual mapping of non-native contrasts is not one-to-one (segment to segment), but perception of single non-native segments is attributed to more than one native segment. This suggests the need to revise the notion of “phonetic category” to accommodate this possible one-to-many mapping. Testing this hypothesis is the goal of present study.

4. Experiment – Categorization task

The goal of the categorization experiment was to relate the discrimination results of Kochetov & Smith (forthcoming) to specific native categories for the various language groups and to confirm the observed discrimination patterns.

4.1 Method

Since the original participants from the discrimination experiment were not available for the orthographic categorization experiment, a different group of listeners was recruited from University of Toronto and the workplace of author JS. Listeners were 8 native speakers from the five non-native language groups: Korean (n=2), Mandarin (n=2), Cantonese (n=3), and Japanese (n=2). 4 listeners were University of Toronto graduate students, 5 were co-workers of author JS, all were resident in the Greater Toronto Area, and all were bilingual English speakers. None had a history of speech, language, or reading disorders. The mean age of listeners was 36 years (range 29-42 years), and mean length of residency for non-Canadians was 14 years (range 2-25 years). One Cantonese listener gave a different response for nearly every stimulus, and many responses violated the phonotactics of Cantonese – the results for that listener were excluded from further consideration.

Stimuli from the discrimination experiment consisted of minimal pairs of two-syllable tokens of the form [taCap] where C is on of /l/, /l j/, /r/ or /r j/ (Table 1), drawn from a corpus of Russian nonsense utterances collected using a magnetic articulometer (Kochetov & Goldstein 2002). Utterances were produced by a female speaker (in her mid 30s) of Standard Russian originally from Moscow. The vowel context was the same in all utterances: low central vowel /a/. The speaker was instructed to stress both syllables. The utterances were read five times in the carrier phrase [ˈɛtʌ_ˈʌpərət] “This is ___ again”. The second and fourth repetitions of each utterance were selected as stimuli tokens. The stimuli were randomized to create two lists of 16 word pairs and presented on a laptop computer in a Microsoft PowerPoint presentation. Testing took place in a quiet meeting room at the author JS’s workplace and in a classroom at the University of Toronto Linguistics Department. Listeners heard stimuli through Sennheiser HD280 Professional headphones. They were instructed to click each sound on the slide and listen to each pair of words. They were instructed to
write the words they heard in the orthography of their language (or English orthography for Cantonese listeners; see below) and proceed through the slides until they heard all 16 word pairs. They were told that the pairs of words were different, but were not told what language they were listening to or what sounds to listen for. Stimuli were presented in groups of four pairs per slide. Pairs had an interstimulus interval (ISI) of 500 ms. Listeners were allowed to listen to each pair of words as many times as they needed to, but once they had written their response, not to return to a previous pair. Stimuli pairs were presented in two blocks of 16 trials in random order. A total of 144 responses to stimuli were collected (16 trials X 9 listeners). Results of the second block of responses for the categorization experiment were compared to the results of the discrimination experiment. Listeners received no feedback on their responses, but were asked afterward to describe any responses that were unclear to the investigator. Most speakers attempted to reproduce what they heard as they described their transcriptions. Japanese and Korean listeners were asked for their interpretation of the Hiragana and Hangul responses. Responses were transcribed phonetically based on listeners’ subjective interpretation. Transcription of the Pinyin responses from Mandarin speakers was relatively straightforward.1 This procedure was subject to limitations, particularly for Cantonese and Japanese responses. Phonetic transcription of Cantonese responses was particularly necessary given the wide variation in vowel representation in English orthography and the lack of a standard Romanization for Cantonese. Transcribed Cantonese responses were further normalized so that each syllable was matched, whenever possible, to the closest phonotactically legal Cantonese syllable, with the exception of laterals and rhotics, which were be left as they were written by listeners. For the Japanese responses, Hiragana does not allow listeners to capture lateral-rhotic contrast, even if they were capable of discriminating the contrast. Also, since all listeners were second-language English speakers, some interference from English experience was likely.

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Order 1</th>
<th>Order 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>[l] vs. [r]</td>
<td>/talap/ vs. /tarap/</td>
<td>/tarap/ vs. /talap/</td>
</tr>
<tr>
<td>[l] vs. [l]</td>
<td>/talap/ vs. /talap/</td>
<td>/talap/ vs. /talap/</td>
</tr>
<tr>
<td>[r] vs. [r]</td>
<td>/tarap/ vs. /tarap/</td>
<td>/tarap/ vs. /tarap/</td>
</tr>
</tbody>
</table>

4.2 Results

The categorization results are summarized by language group. In the following tables, frequencies of unique responses are shown in parentheses. Rhotic ratio is the proportion of responses that showed sensitivity to the lateral-rhotic contrast, as indicated by responses phoneticized with a rhotic segment ([ɹ] or [ɾ]). Palatalization ratio is the proportion of responses that showed sensitivity to the plain-palatalized contrast, as indicated by responses phoneticized with a high front vowel ([i] or [ɨ]) or a high front glide [j].

Categorization by Korean listeners (Table 2) showed generally good sensitivity to both lateral-rhotic and plain-palatalized contrast. In particular, categorization of the lateral-rhotic contrast was generally more accurate for /l/ than for /r/ for both plain and palatalized pairs. Unique responses did not overlap between the two listeners, so frequencies for each listener are given in parentheses in each case.

1 Phoneticization of Pinyin responses follows correspondence between Pinyin and phonetic symbols in the appendix of Duanmu (2007). Phoneticization of the Cantonese responses follows the Yale Romanization system in the Appendix of Matthews and Yip (1994).
Table 2. Korean categorization – frequency of unique responses for listeners K1 and K2

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>/talap/</th>
<th>/tarap/</th>
<th>/talap/</th>
<th>/tarap/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses (n=16)</td>
<td>[tʰ.ʌ.p] (8, K1)</td>
<td>[tʰ.ʌ.p] (7, K1)</td>
<td>[tʰ.ʌ.p] (4, K1)</td>
<td>[tʰ.ʌ.p] (3, K1)</td>
</tr>
<tr>
<td></td>
<td>[tʰ.ʌ.p] (3, K2)</td>
<td>[tʰ.ʌ.p] (1, K1)</td>
<td>[tʰ.ʌ.p] (3, K1)</td>
<td>[tʰ.ʌ.p] (2, K1)</td>
</tr>
<tr>
<td></td>
<td>[tʰ.ʌ.p] (4, K2)</td>
<td>[tʰ.ʌ.p] (4, K2)</td>
<td>[tʰ.ʌ.p] (1, K1)</td>
<td>[tʰ.ʌ.p] (2, K1)</td>
</tr>
<tr>
<td></td>
<td>[tʰ.ʌ.p] (1, K2)</td>
<td>[tʰ.ʌ.p] (3, K2)</td>
<td>[tʰ.ʌ.p] (5, K2)</td>
<td>[tʰ.ʌ.p] (1, K1)</td>
</tr>
<tr>
<td></td>
<td>[tʰ.ʌ.p] (1, K2)</td>
<td>[tʰ.ʌ.p] (3, K2)</td>
<td>[tʰ.ʌ.p] (2, K1)</td>
<td>[tʰ.ʌ.p] (1, K2)</td>
</tr>
</tbody>
</table>

Rhotic ratio 0.00 0.69 0.00 0.81
Palatalization ratio 0.00 0.06 1.00 0.94

Both Korean listeners tended to categorize plain Russian /l/ and /r/ as onset and coda allophones of the single liquid phoneme /l/ – [l] and [ɾ] respectively (both transcribed as Hangul ᴹ). In a few cases, /l/ was interpreted as a hetero-syllabic sequence /l+l/ or /l+n/ (both [l.l]; K2 only). In some responses for /talap/ and /tarap/, both listeners attributed [l] to coda of the first syllable ([tʰ.ʌ.p]-K1 and [tal.ʌ.p] K2), and not onset [ɾ] of the second syllable. This is not expected, since single /l/ intervocically should be onset [ɾ]. It is possible that the Korean listeners heard the plain stimuli as two words ([tʰ.ʌ.p#] and [tal.ʌ.p#]) rather than a single two-syllable word, but this could not be confirmed. Both listeners also tended to distinguish the palatalized liquids, although differed in the syllabification of the stimuli. K1 consistently transcribed the palatalized stimuli as three syllables, rendering /l/ as a hetero-syllabic geminate /l/ followed by /i/, /l.i/ (e.g., [tʰ.ʌ.p] for /talap/, and /ɾ/ as a single /l/ followed by /i/, /ɾi/ (e.g., [tʰ.ɾi] for /tarap/). K2 transcribed the palatalized stimuli as two syllables, rendering /l/ as a coda /l/ (or a geminate /l/) followed by the diphthong /jʌ/ (Hangul ㅕ) and /ɾ/ as an onset /l/ followed by the same diphthong (e.g., [tal.ɾap] for /talap/ and [ta.ɾap] for /tarap/). As with the plain liquids, both listeners categorized laterals and rhotics correctly, though both listeners had more difficulty with /ɾ/ than /l/. Rhotic ratios and palatalization ratios for each listener showed accurate categorization for lateral stimuli, with some categorization confusion with plain rhotic stimulus /tarap/ (K2). Also notable is the difference in the initial consonant, where K1 transcribed /t/ as the Korean voiceless aspirated alveolar stop /tʰ/ (Hangul ᵮ), while K2 transcribed /t/ as the Korean voiceless tensed alveolar stop /t’/ (Hangul ᵯ). The first vowel was transcribed either as /e/ (Hangul ㅔ) (K1 only) or /a/ (Hangul ㅏ). The second vowel was rendered variously as /ʌ/ (Hangul ㅓ), /o/ (Hangul ㅗ), /a/ (Hangul ㅏ), and the diphthong /jʌ/ (Hangul ㅕ).

Table 3 shows the frequencies of unique responses for Mandarin listeners. Despite the larger number of unique responses, particularly for the /tarap/ stimulus, all liquids overwhelmingly assimilated to non-rhotics – mostly lateral /l/ or in a few cases stop /tʰ/ /t/ /t’. This resulted in the low rhotic ratios across all stimuli. Mandarin listeners also tended to assimilate palatalized liquids to sequences /l/ + diphthong /ia/, giving a generally high palatalization ratio. Notable is the lack of /r+ia/ responses, since such a sequence is phonotactically prohibited in Mandarin. The first vowel varied between /a/ and /e/. Velarization of the Russian /l/ was attributed to rounding of the following vowel, rendered as /o/ or /ou/.

Table 3. Mandarin categorization – frequency of unique responses for listeners K1 and K2

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>/talap/</th>
<th>/tarap/</th>
<th>/talap/</th>
<th>/tarap/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses (n=16)</td>
<td>[tʰ.ʌ.p] (8, K1)</td>
<td>[tʰ.ʌ.p] (7, K1)</td>
<td>[tʰ.ʌ.p] (4, K1)</td>
<td>[tʰ.ʌ.p] (3, K1)</td>
</tr>
<tr>
<td></td>
<td>[tʰ.ʌ.p] (3, K2)</td>
<td>[tʰ.ʌ.p] (1, K1)</td>
<td>[tʰ.ʌ.p] (3, K1)</td>
<td>[tʰ.ʌ.p] (2, K1)</td>
</tr>
<tr>
<td></td>
<td>[tʰ.ʌ.p] (4, K2)</td>
<td>[tʰ.ʌ.p] (4, K2)</td>
<td>[tʰ.ʌ.p] (1, K1)</td>
<td>[tʰ.ʌ.p] (2, K1)</td>
</tr>
<tr>
<td></td>
<td>[tʰ.ʌ.p] (1, K2)</td>
<td>[tʰ.ʌ.p] (3, K2)</td>
<td>[tʰ.ʌ.p] (5, K2)</td>
<td>[tʰ.ʌ.p] (1, K1)</td>
</tr>
<tr>
<td></td>
<td>[tʰ.ʌ.p] (1, K2)</td>
<td>[tʰ.ʌ.p] (3, K2)</td>
<td>[tʰ.ʌ.p] (2, K1)</td>
<td>[tʰ.ʌ.p] (1, K2)</td>
</tr>
</tbody>
</table>

Rhotic ratio 0.00 0.69 0.00 0.81
Palatalization ratio 0.00 0.06 1.00 0.94
### Table 3. Mandarin categorization – frequency of unique responses

<table>
<thead>
<tr>
<th>Stimulus /tala/</th>
<th>/tarap/</th>
<th>/talap/</th>
<th>/tarap/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses (n=16)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ta.loup] (8)</td>
<td>[tʰa.la] (4)</td>
<td>[ta.liap] (8)</td>
<td>[ta.liap] (6)</td>
</tr>
<tr>
<td>[ta.lou] (6)</td>
<td>[ta.la] (2)</td>
<td>[tʰa.liap] (4)</td>
<td>[ta.liap] (3)</td>
</tr>
<tr>
<td>[tʰa.lou] (1)</td>
<td>[ta.tʰap] (2)</td>
<td>[tʰe.liap] (2)</td>
<td>[tʰa.liap] (3)</td>
</tr>
<tr>
<td>[tʰa.ɾap] (1)</td>
<td>[ta.tʰap] (2)</td>
<td>[te.lia] (1)</td>
<td>[tʰe.liap] (2)</td>
</tr>
<tr>
<td>[ta.lou] (1)</td>
<td>[tʰa.la] (1)</td>
<td>[ta.tʰiap] (1)</td>
<td></td>
</tr>
<tr>
<td>[ta.lap] (1)</td>
<td>[tʰa.la] (1)</td>
<td>[ta.tʰiap] (1)</td>
<td></td>
</tr>
<tr>
<td>[ta.lap] (1)</td>
<td>[tʰa.la] (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[tʰa.ɾap] (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhotic ratio</td>
<td>0.06</td>
<td>0.31</td>
<td>0.03</td>
</tr>
<tr>
<td>Palatalization ratio</td>
<td>0.00</td>
<td>0.13</td>
<td>0.94</td>
</tr>
</tbody>
</table>

As shown in Table 4, the Japanese listeners were compelled by Japanese orthography to categorize both laterals and rhotics as either /ɾa/ (Hiragana ら) or /ɾa/ (Hiragana いら), while at the same time consistently distinguishing the plain-palatalized contrast according to their native phonology. The palatalization ratio for Japanese categorizations indicates expected good sensitivity to palatalized stimuli. Since Japanese orthography permits only categorization to the Japanese rhotic /ɾ/, rhotic ratio is not relevant for Japanese listeners. In one of the listener’s responses, the second vowel varied depending on the preceding consonant, with the syllable /ɾo/ (Hiragana ろ) corresponding to plain /l/ (a likely perceptual effect of velarization) and /ɾa/ or /ɾa/ corresponding to the other consonants. The first vowel was always transcribed as /a/. The unreleased Russian stop [p] was transcribed as glottal stop [ʔ] (Hiragana っ).

Table 5 shows the frequencies of unique responses for Cantonese listeners. These listeners produced the greatest number of unique responses, indicating extremely low sensitivity to all contrasts. Categorization of the lateral-rhotic contrast by Cantonese listeners was nearly random – both lateral and rhotic stimuli were categorized as either /l/ or /ɾ/ (which is not a phoneme of Cantonese), though /ɾ/ was considerably more frequent (with the most frequent response for all contrasts being [ta.ɾap]). Similarly, there was little indication of sensitivity to palatalized stimuli, indicated by a uniformly low palatalization ratio, with lower sensitivity for rhotic stimuli than for laterals. Cantonese listeners were sensitive to the final [p], which in all cases would be unreleased in Cantonese, as it was in the Russian stimuli. The listeners also seemed generally insensitive to...

---

2 The word-final [p] in the transcriptions from the Mandarin listeners violates Mandarin phonotactics, which only permits final /n/ or /ŋ/; however, it indicates sensitivity to the final unreleased /p/ in the Russian stimuli, and is noted here for phonetic accuracy. The response with a coda [ɾ] ([ta.tʰap]) may represent a rhoticized vowel. ([a]).

3 There was some confusion among Cantonese listeners about what they were supposed to transcribe. They were puzzled by the task of writing words that in many cases “sounded the same”, and were unsure about whether to write the words as if they were “English” words or “Cantonese” words. Lack of a standard Romanization orthography possibly interfered with categorization, and resulted in some phonotactically illicit responses, e.g., not all sequences of vowels and [p] in the Cantonese responses conform to the phonotactically well-formed rimes of Cantonese.
palatalization. For the most part, no responses could be said to assign palatalized liquids to a consistent native category. While front vowel [i] was used to render palatalized /l/ in a few cases ([ta.l] for /talap/), even more often it was used with plain /r/ ([ta.l] and [ta.r] for /talap/ and /tarap/). Vowel assignment consistently favored /a/ for the first vowel, but varied among /ɔ/, /ʊ/, /i/, and /u/ for the second vowel. This may be an artifact of attempting to transcribe “Cantonese” sounds in English orthography.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>/talap/</th>
<th>/tarap/</th>
<th>/talap/</th>
<th>/tarap/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses (n=16)</td>
<td>[da.ra] (8)</td>
<td>[da.ra] (8)</td>
<td>[da.r'a] (8)</td>
<td>[da.r'a] (8)</td>
</tr>
<tr>
<td></td>
<td>[taro?] (8)</td>
<td>[da.ro?] (1)</td>
<td>[da.r'a?] (4)</td>
<td>[da.r'a?] (3)</td>
</tr>
<tr>
<td></td>
<td>[tara?] (4)</td>
<td>[ta.r'a?] (4)</td>
<td>[ta.r'a?] (5)</td>
<td></td>
</tr>
<tr>
<td>Palatalization ratio</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>/talap/</th>
<th>/tarap/</th>
<th>/talap/</th>
<th>/tarap/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses (n=16)</td>
<td>[ta.l] (1)</td>
<td>[ta.l] (1)</td>
<td>[ta.l] (1)</td>
<td>[ta.l] (1)</td>
</tr>
<tr>
<td></td>
<td>[ta.l] (1)</td>
<td>[ta.l] (1)</td>
<td>[ta.l] (1)</td>
<td>[ta.l] (1)</td>
</tr>
<tr>
<td></td>
<td>[ta.l] (1)</td>
<td>[ta.l] (1)</td>
<td>[ta.l] (1)</td>
<td>[ta.l] (1)</td>
</tr>
<tr>
<td></td>
<td>[ta.l] (1)</td>
<td>[ta.l] (1)</td>
<td>[ta.l] (1)</td>
<td>[ta.l] (1)</td>
</tr>
<tr>
<td></td>
<td>[ta.l] (1)</td>
<td>[ta.l] (1)</td>
<td>[ta.l] (1)</td>
<td>[ta.l] (1)</td>
</tr>
<tr>
<td></td>
<td>[ta.l] (1)</td>
<td>[ta.l] (1)</td>
<td>[ta.l] (1)</td>
<td>[ta.l] (1)</td>
</tr>
<tr>
<td>Palatalization ratio</td>
<td>0.13</td>
<td>0.25</td>
<td>0.50</td>
<td>0.19</td>
</tr>
</tbody>
</table>

In summary, categorization by Korean listeners showed generally good sensitivity to both lateral-rhotic and plain-palatalized contrast. Mandarin listeners consistently assimilated the lateral-rhotic contrast to /l/ rather than /r/, while showing good sensitivity to the plain-palatalized contrast. Japanese listeners categorized the plain-palatalized contrast accurately to /r/ or /rj/, though the previously mentioned limitation of the Japanese orthography could not fully reveal how Japanese listeners actually assimilated the lateral-rhotic contrast. Categorization of the lateral-rhotic contrast by Cantonese listeners was nearly random – both lateral and rhotic stimuli were categorized as /r/, though most frequently as /l/, suggesting that there was little actual discrimination of the
lateral-rhotic contrast. Similarly, there was little indication of sensitivity to palatalized consonants, indicated by a uniformly low palatalization ratio, with lower sensitivity for rhotic stimuli than for laterals.

4.5 Discussion

The results of the categorization experiment have revealed some interesting differences in responses among listener groups and experimental conditions. These are summarized in the following figures: the rhotic ratio of categorization responses by language group (Figure 2a; excluding Japanese listeners) and the palatalization ratio of categorization responses (Figure 2b). Recall that rhotic ratio indicates generally accurate assimilation of plain laterals and rhotics to the appropriate native categories and palatalization ratio is the proportion of responses that showed sensitivity to the plain-palatalized contrast, as indicated by responses phoneticized with a front vowel ([i] or [ɪ]) or a front glide [j].

As seen in Figure 2a, Russian plain /l/ was almost consistently categorized as non-rhotic by Cantonese, Mandarin, and Korean listeners, while plain /r/ was categorized as either non-rhotic or rhotic, with the latter response more common for Korean listeners. Russian palatalized /l j/ was categorized as rhotic almost half of the time by Cantonese listeners, while almost always considered a non-rhotic by Mandarin and Korean listeners. Russian palatalized /r j/ was infrequently categorized as rhotic by Cantonese and Mandarin listeners (less than 20%), while Korean listeners assimilated it to its rhotic allophone more than half of the time. This shows that listeners overall tended to favour the non-rhotic response, in many cases assimilating the Russian plain and palatalized trills/taps to their native lateral category. The relatively low categorization of /r/ as rhotic by Mandarin listeners suggests that they perceive the Russian consonant as sufficiently different from their native rhotic ɹ, and more similar to their lateral /l/. The relatively high and consistent categorization of both /l/ and /r j/ as rhotic and of /l j/ and /l j/ as non-rhotic shows that Korean listeners are using their native knowledge of consonant/sequence contrasts (mapping them onto single /l/ or geminate /ll/ and /l j/), thus overcoming the limitation imposed by their segmental system of contrasts.

Figure 2b shows near-ceiling palatalization ratios for palatalized consonants by all listener groups, with the exception of Cantonese listeners. This reflects overall accurate categorization of the Russian plain and palatalized contrasts, as either the most similar native consonants or consonant + glide or diphthong sequences. Particularly, Korean listeners attributed palatalization of the liquids to the following high front vowel /i/ or the diphthong /jʌ/, thus assimilating the non-native segments as sequences /(C,)C+i(V)/ or /{(C,)C+jʌ}/. The nonpalatalized liquids were rendered as segments or sequences (C+C) that characteristically lacked following front vowels or diphthongs. Similarly, Mandarin listeners tended to assimilate palatalized liquids to sequences of /l/ + diphthong /iV/, in which the lateral is phonetically palatalized. Evidently, the lack of phonemic palatalization in the native language had little or no influence on Mandarin listeners’ responses. Not surprisingly, Japanese listeners were highly accurate categorizing Russian palatalized stimuli to their own palatalized segments (with 100% accuracy). Cantonese listeners showed the lowest palatalization ratios for all stimuli, as their responses to the palatalized (and plain) liquids rarely contained front vowels and glides. Although lower than average accuracy for this group was expected (given the lack of the native plain-palatalized contrast or /C+j/ sequences), the results are somewhat puzzling in light of the relatively good discrimination performance of Cantonese listeners in the discrimination experiment. The Cantonese categorization results may not be therefore fully representative of these listeners’ native phonological or phonetic knowledge, but rather reflect the difficulty of the
categorization task (given the lack of appropriate native orthography) and likely interference of English.

Figure 2. Rhotic ratio (a) and palatalization ratio (b) by language group

5. General discussion

The results of the categorization experiment confirm the results of the earlier discrimination experiment (Kochetov & Smith, forthcoming) and shed light on how listeners actually assimilate the non-native contrasts. Categorization results for all listeners (leaving apart the Cantonese responses to the palatalized stimuli) matched well the discrimination performance of other listeners belonging to the same language groups, with respect to both lateral-rhotic and plain-palatalized contrasts. In particular, the higher than expected sensitivity to the lateral-rhotic contrast by Korean listeners in the previous experiment can be explained by different assimilation strategies for the two consonant types: while liquids are mapped onto a single /l/ in syllable coda ([l]) or a geminate /l.l/ ([lː]), rhotics are mapped onto a single /r/ in syllable onset [ɾ]. The lower than expected discrimination of the lateral-rhotic contrast by Mandarin listeners can be attributed to their overwhelming assimilation of both Russian laterals and rhotics to their lateral category, likely due to the perceived substantial difference between the Russian tap/trill and the Mandarin retroflex approximant /ɹ/. The poor discrimination of the contrast by Japanese and Cantonese listeners also matched the categorization performance of both groups in the current experiment.

The overall very high sensitivity to the Russian plain-palatalized contrast in the discrimination experiment are also consistent with the patterns of assimilation shown in the current
experiment: assimilation of palatalized consonants as corresponding native segments (Japanese) or the sequences of consonants and high front vowels or diphthongs (Korean and Mandarin). The overall inaccurate categorization of the plain-palatalized contrast by Cantonese listeners, however, did not match their high discrimination of the contrast, and may be attributed to the relative difficulty of the categorization task (as discussed in the previous section). In addition, the results of the categorization experiment did not consistently show higher sensitivity to the lateral-rhotic contrast among plain liquids compared to palatalized ones. Likewise, among plain-palatalized laterals compared to plain-palatalized rhotics, the categorizations did not consistently show the differences (albeit small) revealed by the discrimination experiment. These apparent mismatches require further investigation.

Overall, results of both studies show that native language background clearly affects discrimination and categorization of non-native phonological contrasts. In addition, the studies also reveal patterns of non-native perception that extend beyond the target consonants. It is notable that all categorizations show striking effects that the two contrasts appear to have on the surrounding vowels. While some of this is attributable to inconsistencies in the native orthography (e.g., Cantonese responses based on English orthography), it is also clear that perception of consonants is not independent of the surrounding vowels, and in fact some properties of a consonant can be attributed to vowels (which is often reflected in the listeners’ transcriptions). The results therefore suggest that perceptual mapping of non-native categories is not necessarily one-to-one; rather, where a language lacks a single category for a given non-native contrast, as is the case for phonemic palatalization in Korean and Mandarin, listeners consistently transferred or distributed phonetic input across several segments. The same applies to the categorization of the lateral-rhotic contrast by Korean listeners, whose interpretation of segments as sequences allowed them to overcome the lack of a native segmental contrast. Taken together, the findings of the study provide support for the hypothesis that perceptual mapping of non-native contrasts is not necessarily one-to-one (segment to segment). Rather, a single segment can be mapped onto either a segment or a sequence of segments, depending on the phonetic properties of the non-native segment and the native sound patterns.

It should be noted that the unconstrained categorization method used in this experiment has limitations, particularly in the case of Japanese, which does not distinguish the large phonetic variation of its /r/ phoneme in orthography. As already noted, the categorizations that use English orthography suffer from inconsistent vowel representation (as in the case of Cantonese responses). Perhaps constraining the categorization by having listeners choose responses from a list of choices would overcome these biases, although this would not solve the Japanese problem.

More research is clearly needed to explore the full extent of cross-language and within-group variation observed in the current study. Of particular interest is the degree to which target non-native segments are apparently attributed to a more than one native segment. Perception of phonetic properties of the non-native segment seems to spread to surrounding native segments to build a one-to-many perceptual mapping. It is possible that native phonology acts as a perceptual “template” into which attributes of non-native sounds are slotted.

6. Conclusion

Emphasis in more recent non-native perception research on the fine-grained phonetic details of speech perception has proven to be a fruitful direction of inquiry (Best et al. 2001; Harnsberger 2001; Aoyama et al. 2004). The lack of phonetic detail has made it difficult to generalize how particular native consonant inventories affect perception. For example, Harnsberger (2000) claims that allophonic, rather than phonemic, representations of perceptual categories are more successful in predicting the native category that listeners used to label a non-native sound. Even so, he concludes that perceptual categories described only in terms of abstract units such as phonemes or allophones
do not have enough articulatory and acoustic-phonetic detail to describe what is available to listeners (Harnsberger 2000). The present study reinforces the crucial finding of this research: the failure of abstract representations alone to predict patterns of discrimination, identification, and categorization of non-native sounds. Perceptual categories described only in terms of abstract segmental units such as phonemes or allophones cannot account for non-native perceptual sensitivities in traditional perceptual tests. Clearly, more detailed, cross-language descriptions of acoustic cues and articulatory gestures are needed to get a sufficiently detailed description of perceptual categories. Furthermore, attention to language-specific differences at the acoustic phonetic feature level rather than the phoneme or allophone level are essential to predicting and explaining non-native identification and discrimination and enabling the necessary detail in the description of actual perceptual processes.

References


CATEGORIZATION OF NON-NATIVE LIQUID CONTRASTS


