Perception of place and secondary articulation contrasts in different syllable positions: Language-particular and language-independent asymmetries

Alexei Kochetov, University of Toronto

Available at: http://works.bepress.com/alexei_kochetov/13/
Perception of Place and Secondary Articulation Contrasts in Different Syllable Positions: Language-Particular and Language-Independent Asymmetries
Alexei Kochetov
Language and Speech 2004 47: 351
DOI: 10.1177/00238309040470040201

The online version of this article can be found at:
http://las.sagepub.com/content/47/4/351

Published by:
SAGE
http://www.sagepublications.com

Additional services and information for Language and Speech can be found at:

Email Alerts: http://las.sagepub.com/cgi/alerts
Subscriptions: http://las.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav
Citations: http://las.sagepub.com/content/47/4/351.refs.html
Perception of Place and Secondary Articulation Contrasts in Different Syllable Positions:
Language-Particular and Language-Independent Asymmetries

Alexei Kochetov
Simon Fraser University, Burnaby, BC, Canada

Abstract
This study investigated the perception of place and secondary articulation contrasts in different syllable positions by Russian and Japanese listeners. The consonants involved in the study were the Russian plain (velarized) and palatalized labial and coronal voiceless stops in syllable-initial and syllable-final positions at word boundaries. The findings revealed substantial asymmetries in the perception of the contrasts by both groups of listeners: With respect to positions, consonants in syllable-final position were characterized by lower correct identification rates and (less consistently) longer reaction time than the same consonants in syllable-initial position. Positional differences were accompanied by differences in segment-specific contexts. With respect to individual consonants, the palatalized labial /pʲ/ and the plain coronal /tʲ/ showed a lower correct identification rate and smaller perceptual distance than the plain labial /p/ and the palatalized coronal /tʲ/. The results also showed some differences between Russian and Japanese listeners in the perception of the contrast. These differences can be explained by phonotactic differences between the two languages, as well as by differences in the phonetic realizations of the consonants. The results of the study provide evidence for the role of both universal and language-particular factors in speech perception.

Key words
palatalization
perception
phonotactics
place
syllable position

Acknowledgments: I am grateful to Keren Rice, Ron Smyth, Louis Goldstein, and Elizabeth Hume for the invaluable assistance and advice on this project. Many thanks to Sarah Hawkins and Elizabeth Zsiga for many helpful comments and suggestions on the earlier draft of this paper. Thanks to Maho Kobayashi for her assistance with the Japanese part of the experiment. The research was supported by a Post-doctoral Fellowship from the Social Sciences and Humanities Research Council of Canada (SSHRC) held at Haskins Laboratories, New Haven, CT and by a Doctoral SSHRC Fellowship held at the University of Toronto. All errors are my own.

Address for correspondence. Dr. Alexei Kochetov, Department of Linguistics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada: e-mail: <akochetov@sfu.ca>.
1 Introduction

It is well established that speakers’ language-specific phonological knowledge—the system of phonological contrasts and alternations—influences their perception of speech sounds. Studies of first-language acquisition (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; see also Vihman, 1996) have revealed that children’s speech perception is affected by language-specific phonetic/phonological properties of the ambient language almost as soon as they are exposed to it. This influence of phonology on perception becomes an even more important factor in adults’ perception, making perception or acquisition of novel phonological contrasts difficult. (See Best, McRoberts, & Sithole, 1988, on perception of non-native contrasts by second-language learners.)

In addition to language-specific characteristics, language-independent properties of speech perception have also been a focus of recent phonetic research. One aspect of this research is the investigation of universal properties of the perception of phonological contrasts in different phonological environments (Fujimura, Macchi, & Streeter, 1978; Hura, Lindblom, & Diehl, 1992; Kingston, to appear; Ohala, 1990; Redford & Diehl, 1999), and of the possible role of these properties in explaining common positional asymmetries in phonology (Jun, 1995; Steriade, 1997, among others). Some of these studies have revealed intriguing asymmetries in the perception of consonants in different contexts; for instance, English and Japanese listeners were found to perform well at identifying the place of articulation of consonants that are followed by vowels (CV sequences, or syllable-initial position), but to show substantially more errors at the same task in the absence of a following vowel (VC sequences, or syllable final position) (Fujimura et al., 1978). Interestingly, the results for English and Japanese listeners were very similar, despite the profound differences between the two languages in syllable structure and the phonotactics of place contrasts. It was suggested that this perceptual positional asymmetry might be a source of the regressive direction of phonological place assimilation (Ohala, 1990; Steriade, 2001; Jun, 1995). In addition to the positional asymmetry, differences in the perception of place (and manner) of articulation have also been observed. Thus, coronals in coda position were misidentified more often than labials or velars in the same position (Byrd, 1992; Kingston, to appear; Surprenant & Goldstein, 1998; Winters, 2001; Wright, 2001). These perceptual findings correlate with the general susceptibility of coronals to regressive place assimilation (Jun, 1995).

An explanation of positional and place asymmetries in perception, as well as in phonology, has been sought in the universal characteristics of speech perception, specifically in listeners’ greater attunement to CV transitions than to VC transitions (Ohala, 1990), and in the lesser perceptual salience of alveolars and nasals (Jun, 1995). They have also been explained by articulatory factors, namely by the overlap and perceptual “hiding” of articulatory gestures in sequences, especially the tongue tip gesture in coda (Byrd, 1992; Surprenant & Goldstein, 1998). It should be noted, however, that the number of studies which directly address the question of positional and place asymmetries is rather small; many of these studies deal with the perception of “major place contrasts” (labials, alveolars, and velars) by speakers of a single language, mainly English (e.g., most of the studies mentioned above; but see Fujimura...
et al., 1978). Unfortunately, this makes it hard to distinguish between truly language-specific and universal effects in each particular case, and weakens generalizations made about the universality of positional perceptual asymmetries.

The current study seeks to contribute to the investigation of cross-linguistic processing of phonological contrasts in different positions. The study investigates native and non-native perception of place and secondary articulation contrasts in Russian consonants — /p/, /t/, /pʲ/, /tʲ/ — in a number of syllable onset and syllable coda environments. It attempts to answer the following questions: To what extent are listeners’ responses affected by their native language? To what extent do these findings reflect language-independent, universal perceptual capabilities and biases? The findings reveal substantial similarities, as well as interesting differences in processing of the contrasts by native and non-native listeners. The findings thus provide strong evidence for a complex interaction of language-independent factors and language-particular factors.

2 Place and secondary articulation contrasts in Russian and Japanese

2.1 Russian /p/, /t/, /pʲ/, and /tʲ/

Russian has a phonological distinction between non-palatalized (velarized) and palatalized consonants (Avanesov, 1972; Jones & Ward, 1969; Timberlake, 1993). The consonants involved in this contrast, such as /p/, /pʲ/, /t/, and /tʲ/, tend to have roughly the same primary articulation (bilabial or dental/alveolar), but have different secondary gestures produced almost simultaneously with the primary gestures: some tongue backing for “plain” consonants (mainly labials: Kochetov, 2002), and substantial tongue fronting and raising for palatalized consonants. Note that the Russian palatalized alveolar stop /tʲ/ is also contrastive with the palatalized post-alveolar affricate /tʃʲ/.

Table 1 summarizes the phonotactics of the Russian plain/palatalized labial and coronal stops in five different syllable onset and syllable coda contexts: adjacent to a vowel (a; V) and adjacent to consonants homorganic with respect to place and secondary articulation (b; Cₕm-hₘ₃), hetero-organic with respect to place and homorganic with respect to secondary articulation (c; Cₗt-hₘ₃), homorganic with respect to place and hetero-organic with respect to secondary articulation (d; Cₗt-hₗ₃), and finally, hetero-organic with respect to place and secondary articulation (e; Cₕt-hₗ₃). To make the comparison between Russian and Japanese clearer, the examples given in each cell are nonsense words of the type aC₁(C₂)a, where C₁ and C₂ are either /p/, /pʲ/, /t/, or /tʲ/. The consonant of interest in each case is given in bold; the cells with ill-formed items are shaded. Well-formed items roughly correspond to real Russian words (with some possible differences in consonant voicing and manner, and vowel quality).

---

1 Some results of the current experiment (viz., the perception of the plain/palatalized contrast in specific segmental contexts) have been presented in the context of a phonological study investigating the role of articulatory and perceptual factors in the emergence of cross-linguistic patterns of palatalization (Kochetov, 2002; a revised version of my Ph.D. thesis).
Table 1

Distribution of the contrasts between /p/, /t\^/, /t/, and /t/ in Russian; V = adjacent to a vowel, C = adjacent to a consonant; hm-hm = homorganic place, homorganic secondary articulation; ht-ht = hetero-organic place, hetero-organic secondary articulation, etc.; shading indicates phonotactically ill-formed word-internal contexts

<table>
<thead>
<tr>
<th>Context</th>
<th>Onset</th>
<th>Coda</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p</td>
<td>p^</td>
</tr>
<tr>
<td>a. V</td>
<td>pa</td>
<td>pa^</td>
</tr>
<tr>
<td>b. C_{hm-hm}</td>
<td>appa</td>
<td>ap^p^a</td>
</tr>
<tr>
<td>c. C_{ht-hm}</td>
<td>atpa</td>
<td>atp^a</td>
</tr>
<tr>
<td>d. C_{hm-ht}</td>
<td>*ap^pa</td>
<td>*ap^pa</td>
</tr>
<tr>
<td>e. C_{ht-ht}</td>
<td>at^pa</td>
<td>atp^a</td>
</tr>
</tbody>
</table>

The items in (a) show that the contrast between Russian plain and palatalized voiceless labial or coronal stops is maintained in the vowel context, both syllable-initially or word-initially (cf., /p/ at 'stalemate' versus /p\^/ at 'heal', gen. pl.; /t\^/ 'so' versus /t/ 'agad 'drought') and syllable-finally at the end of the word (cf., /uto[p] 'has sunk' versus /to[p\^] 'swamp'; /ma[t] 'mat' versus /ma[t\^] 'mother'). The items in (b–e) show the distribution of the consonants in four types of consonant clusters. Consonants in the clusters homorganic with respect to place (c) always agree in their secondary articulation (cf., /gru[p]a 'group', /gru[p\^]a 'group', gen. sg., /o[tt]\^a 'askivat' 'to drag away', /o[t\^] 'agivat' 'to pull away'). None of the hetero-organic-homorganic clusters are phonotactically possible (d). When such clusters arise through derivation, the phonological processes of palatalization (C\^C \rightarrow C\^C\^) and depalatalization (C\^C \rightarrow CC) are known to apply: /o[t]+[t\^] 'agivat' \rightarrow /o[t\^] 'agivat' 'to pull away'; /pja[t\^]+[n]\^a 'decap' \rightarrow /pj\^a[t\^] 'decap' '15') (Kochetov, 2002; Timberlake, 1993). Almost all consonant combinations are possible in clusters heterorganic with respect to place (c and e) (cf., /o[tp]a 'fell off', /sva[d\^b]a 'wedding', /le[pt]\^a 'mite', /a[pt\^]\^a 'pharmacy'), except the ones with the palatalized labial /p\^/ as C\_1 (*[p\^t\^] and *p\^t\^)). Thus, the contrast between the coronals /t/ and /t/ is maintained in syllable onset. It is worth noting that word-final stops and coda stops in hetero-organic clusters in Russian tend to be audibly released (Jones & Ward, 1969, p. 88; Kochetov & Goldstein, 2001; Zsiga, 2000), thus the plain/palatalized contrast in coda is acoustically conveyed by both VC transitions and releases (Kochetov, 2002).

In sum, the place of articulation contrast (labial vs. coronal) is maintained in all contexts in both syllable positions. The secondary articulation contrast (plain/velarized vs. palatalized) is maintained in single consonants syllable-initially and

---

2 Here and below, Russian words are given in the transliteration adopted in the North American literature on Russian; Japanese words are given in the traditional Romanization (cf., Ito & Mester, 2003); IPA symbols are used to transcribe phonemes of interest.
syllable-finally; it is also maintained in hetero-organic consonant clusters, unless C₁ is /p/>. It should be noted that no restrictions on consonant sequences across word boundaries has been reported (e.g., Jones & Ward, 1969, p. 94), thus all consonant combinations are assumed to occur here.

2.2 Japanese /p/, /p/, /t/, and /t̠c/

Japanese is described as having a contrast between plain and palatalized stops/affricates, for example, /p/ versus /p/ and /t/ versus /t̠c/ (Ito & Mester, 2003, p. 8–9; Tsujimura, 1996, pp. 11–16, 38–40; Vance, 1987, pp. 28–29). Unlike Russian, Japanese plain consonants are not considered to be velarized. The Japanese palatalized labial stop /p/ appears to be similar to the corresponding Russian prevocalic consonant, the Japanese coronal affricate /t̠c/ differs from the Russian palatalized coronal /t̠j/ in its primary place of articulation (alveolopalatal vs. denti-alveolar) and the degree of affrication (a “true” affricate vs. an affricated stop). It is thus more phonetically similar to the Russian palatalized post-alveolar affricate /t̠j/ than to /t̠j/.

Table 2

<table>
<thead>
<tr>
<th>Context</th>
<th>Onset</th>
<th>Coda</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. V</td>
<td>pa</td>
<td>*ap</td>
</tr>
<tr>
<td>b. C_hm-hm</td>
<td>appa</td>
<td><em>ap</em>apja</td>
</tr>
</tbody>
</table>
| c. C_ht-hm | *atpa | *atptca *
| d. C_hm-hm | *apja | *atptca *
| e. C_ht-hm | *atpja | *atptca * |

The phonotactics of the Japanese place and secondary articulation contrasts are summarized in Table 2. As with Russian (Table 1), examples given in each cell are well-formed or ill-formed nonsense words. Unlike Russian, place and secondary articulation contrasts in Japanese are limited to the syllable-initial position: in the vowel context (a) they are maintained in onset (cf., [p]oko-[p]oko ‘up and down movement’, [p]oko-[p]oko ‘jumping around imprudently’, [t]a ‘rice field’, [t̠c]a ‘tea’) and neutralized in coda, since the language does not permit word-final coda obstruents. Similarly, both contrasts are neutralized in consonant clusters: /p/, /p/, /t/, and /t̠c/ can occur here only when preceded by or followed by consonants of the same place (b; i.e., in geminates or in nasal-obstruent clusters; Ito & Mester, 2003, pp. 8–9). They never occur in the clusters hetero-organic with respect to either place or secondary articulation. Place and secondary articulation are thus fully predictable from the

It is important to note that Japanese treats word-final coda obstruents in loan-words as onset obstruents followed by an epenthetic vowel (e.g., loans from English: Vance, 1987). Thus, word-final coda consonants in the Russian words uto[p], to[pj], ma[t], and ma[v] are interpreted in Japanese as onset segments of an additional syllable: utoo[pu], too[pj], maa[to], and maa[ç]. Similarly, Russian consonants in clusters in o(tp)al, sva(d'b)a, and le[pt]a ‘mite’ are interpreted as o[top]aaru, suvaad[dz'å]a, ree[pu]ta (cf., Ando et al., 1997; Ito, 1993). Note that in Japanese high vowels tend to devoice word-finally after voiceless obstruents and word-internally between voiceless obstruents (Tsujimura, 1996, pp. 24–29). This makes, for instance, the Japanese sequences [pu] and [pu]t in utoo and reepta perceptually relatively similar to the corresponding Russian [p] and [pt] in utop and lepta, where the coda stops are often audibly released. The sound correspondences between Russian and Japanese were confirmed by a pilot study where two Japanese native speakers were presented with Russian words with plain and palatalized stops in different phonological contexts. They were asked to write down the words in the Katakana script and then to pronounce the items based on their writing.

In sum, the phonotactic similarities between Japanese and Russian are limited to the presence of place and secondary articulation contrasts in syllable-initial single consonants; both languages also allow all four consonants non-contrastively in homorganic-homorganic clusters. Unlike Japanese, Russian also contrasts place and secondary articulation in single syllable-final consonants, as well as in most consonant clusters hetero-organic with respect to place. The phonological and phonetic similarities and differences between Russian and Japanese are crucial for the current study. Investigating the perception of the same Russian stimuli by Russian and Japanese listeners will allow us to determine to what extent the results are due to language-particular influences (phonotactic knowledge or phonetic differences) and to what extent they are due to language-independent perceptual properties of position and the consonant contrasts involved.

2.3 Predictions

Given the widely attested effect of language-specific phonotactic knowledge on speech perception (Best et al., 1988; Kuhl et al., 1992), it is reasonable to expect that listeners will perform well in identifying place and secondary articulation contrasts in the contexts phonotactically legal in listeners’ native languages. With respect to the segmental contexts, we may expect that Russian listeners would perform well in the identification of single consonants, regardless of their position (see Table 1, a), as well in the identification of onset and coda consonants in the clusters homorganic with respect to place and secondary articulation (b). Further, their performance with onset and coda consonants in the clusters hetero-organic with respect to place (c and e) should be as good, except for the clusters where C1 is /pj/. Finally, Russian listeners are likely to perform the worst with onset and coda consonants in the clusters homorganic with respect to place and hetero-organic with respect to secondary articulation (d). In sum, we may expect that the performance of Russian listeners will decrease from...
context to context in the following order: V, C_{hm-hm} > C_{ht-hm}, C_{ht-ht} (/pjtj/ and /pht/ only) > C_{hm-hl} (where “>” stands for ‘better than’: higher correct identification of consonants and shorter reaction time).

As for Japanese listeners, we may expect that they would perform well in the identification of single consonants in syllable-initial position (see Table 2, a, onset), as well in the identification of onset and coda consonants in the clusters homorganic with respect to place and secondary articulation (b). Their performance with single coda consonants (a, coda), as well as onset and coda consonants in clusters that are not homorganic either with respect to place or secondary articulation (c, d, and e) should be the worst. In sum, we may expect that the performance of Japanese listeners will decrease from context to context in the following order: V (onset), C_{hm-hm} > V (coda), C_{ht-hm}, C_{ht-ht}, C_{hm-hl}. Note that in Japanese all four consonants of interest are fully contrastive only when they are single and syllable-initial, while being fully predictable both in terms of place and secondary articulation in clusters; this could further differentiate the vocalic onset context from all the other contexts.

Given the previously observed language-independent syllable position perceptual effects (Fujimura et al., 1978), we might expect to find a similar positional asymmetry for both Russian and Japanese listeners: an overall better performance (higher correct identification of consonants and shorter reaction time (RT)) in syllable onset position and an overall poorer performance (higher correct identification of consonants and shorter RT) in syllable coda position (onset > coda). More interestingly, these general syllable-position effects may interact with the context-specific phonotactic effects in the following way. Both Russian and Japanese listeners will perform well with onset consonants, but their performance may be lower with the phonotactically ill-formed clusters (C_{hm-hl} for Russian and all except C_{hm-hm} for Japanese). Further, both Russian and Japanese listeners will perform poorer with syllable-final consonants, but their performance may be higher with the phonotactically well-formed clusters (almost all except C_{hm-hl} clusters for Russian and only C_{hm-hm} for Japanese).

In addition, perceptual effects of position and contexts may not be the same for individual consonants, given their phonotactic distribution (e.g., /pjt/ in Russian clusters), their general articulatory and acoustic differences, as well their different susceptibility to contextual variation. Language-specific differences in the production of consonants (e.g., the Russian /tj/ versus the Japanese /tc/ ) could also be a factor.

Finally, comparing the two languages, we may expect that Russian listeners perform overall better than Japanese listeners, at least given the language differences in the scope of phonotactic restrictions. Russian performance is expected to be better in the contexts that are phonotactically possible in Russian but not in Japanese (coda V and most C_{ht-hm} and C_{ht-ht} clusters). The two language groups are expected to show similarly good performance in the contexts that are phonotactically possible in both languages (onset V and C_{ht-hm} and C_{ht-ht} clusters), and similarly poor performance in the contexts that are phonotactically impossible within words in both languages (C_{ht-ht}, as well as C_{ht-hm} and C_{ht-ht} clusters with /pjt/ as C_{1}). However, given the possibility of such consonant sequences across word boundaries in Russian, we may expect Russian listeners to perform here somewhat better than Japanese listeners.
3 Method

3.1 Stimuli

Since phonotactic restrictions in Russian do not hold across word boundaries (at least categorically: see discussion below), it appeared reasonable to use word-boundary clusters instead of word-internal clusters as stimuli for the experiment. The stimuli, shown in Table 3, consisted of 16 nonsense utterances of the type \([\text{ta}(C_1)\#(C_2)\text{api}]\), where \(C_1\) and \(C_2\) were /\(p\)/, /\(p\, j\)/, /\(t\)/, and /\(t\, j\)/. The vowel context was the same throughout: the stressed low central vowel /\(a\)/. Five repetitions of each utterance were read in a carrier phrase \(\text{eto }\_\_\text{ opjat'}\) \([\text{eta }\_\_\text{p'at'j}']\) ‘This is \_\_ again’. The utterances were produced by a native speaker of Russian, the author, and collected using the EMMA magnetometer system (Perkell et al., 1992) at Haskins Laboratories. (Subsets of these and other nonsense and real utterances produced by four Russian speakers were used in a series of articulatory studies (Kochetov, 2002; Kochetov & Goldstein, 2001; Kochetov, to appear).)

<table>
<thead>
<tr>
<th>Context</th>
<th>Position</th>
<th>Stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. V</td>
<td>onset</td>
<td>ta papi</td>
</tr>
<tr>
<td></td>
<td>coda</td>
<td>tap api</td>
</tr>
<tr>
<td>b. Chm-hm</td>
<td>coda/onset</td>
<td>tap papi</td>
</tr>
<tr>
<td>c. Cht-hm</td>
<td>coda/onset</td>
<td>tat papi</td>
</tr>
<tr>
<td>d. Cht-hl</td>
<td>coda/onset</td>
<td>tap\text{i} papi</td>
</tr>
<tr>
<td>e. Chl-hlt</td>
<td>coda/onset</td>
<td>tat\text{i} papi</td>
</tr>
</tbody>
</table>

Some of the stimulus items were real Russian words (e.g., papy [papi] ‘dad’ gen. sg., ta [ta] ‘that’, fem. nom. sg., tjap(-l'jap) [t\text{i}ap] ‘sloppy’), while others were phonotactically well-formed nonsense words (e.g., tapy [tapi], tat’ [tat’], and tap’ [tap’]; compare lapy [lapi] ‘paw’, gen. sg., stat’ [stat’] ‘to become’, top’ [top’] ‘bog’). Using sequences of real words and nonsense words made it possible to keep all other segmental and prosodic contexts constant, while minimizing lexical interference in the phoneme identification task performed by native speakers. An attempt was made to produce the items as naturally as possible. (Some nonsense/real word production differences are discussed below).

Twenty stimuli (a–e, excluding the coda items in a) were used for the identification of onset consonants; 20 stimuli (a–e, excluding the onset items in a) were used...
for the identification of coda consonants. Thus, most of the stimuli (b–e) were used for the identification of both onset and coda consonants.

3.2 Participants

The participants were 20 native speakers of Russian and 10 native speakers of Japanese. The Russian participants were 11 males and 9 females, mostly graduate students at the University of Toronto who have lived in Canada for about three years prior to the experiment; their average age was 30. The Japanese participants were 1 male and 9 females, mostly ESL University of Toronto students who have lived in Canada for about one year; their average age was 27. The Japanese participants had no prior exposure to the Russian language. Both groups were paid for their participation. None of the participants reported any hearing disabilities.

3.3 Data collection and analysis

During the experiment, the stimuli were presented separately for syllable-initial and syllable-final consonants with 20 stimuli in each block using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993). Each utterance was repeated five times with each repetition corresponding to a different production token of the same utterance. Russian listeners were told that they would be presented with utterances consisting of two nonsense words in two blocks. In Block 1, the second word would be either papy [papi], pjapy [p’api], tapy [tapi], or tjapy [t’api]. The listeners’ task was to identify whether the first syllable of the second word was either pa, pja, ta, or tja by pressing the keys labeled correspondingly in Cyrillic (‘па, пя, та, тя’). They were told that in Block 2 the first word would be either tap [tap], tap’ [tap’], tat [tat], or tat’ [tat’]. The listeners’ task was to identify whether the last consonant of the first word was either p, p’, t, or t’ by pressing the keys labeled correspondingly (‘п, пь, т, ть’). The instructions were conveyed to the participants both in writing and orally.

The instructions for Japanese listeners were similar: They were told that in Block 1 the second word would sound like either papui [papui], pyapui [p’apui], tapui [tapui], or chapui [chapui]. The listeners’ task was to identify whether the first syllable of the second word was either pa, pya, ta, or cha by pressing the keys labeled correspondingly in the Katakana script (“パ, ピャ, タ, チャ’). They were told that in Block 2 the first word would sound like either tapu [tapu], tapi [tap’i], tato [tato], or tachi [ta’ci]. The listeners’ task was to identify whether the last syllable of the first word was either pu, pi, to or chi by pressing the keys labeled correspondingly (“プ, ピ, ト, チ”). This choice of characters was based on the traditional transliteration of Russian words in Japanese (Ando et al., 1997; Ito, 1993), as well as on the perceptual pilot study mentioned above. The use of characters was further confirmed with each participant during the pre-test. The pre-test, conducted for both Russian and Japanese listeners, involved a writing task where listeners were presented with two tokens of the same stimuli and were asked to write down the words in orthography.
Confusion matrices obtained from the raw phoneme identification responses during the experiment were used to determine mean correct identification rates (proportions of correct responses out of 1.00). There were a total of 4000 identification responses (40 utterances × 5 repetitions × 20 listeners) for Russian listeners and 2000 responses of the same type (40 utterances × 5 repetitions × 10 listeners) for Japanese listeners. There were 31 cases (1%) where Russian listeners did not respond to a stimulus; there were 13 such cases (less than 1%) for Japanese listeners. Reaction times (RT), measured from the onset of an audio stimulus, were also collected for each response. Only positive responses were used in the analysis, resulting in 3140 RT responses for Russian listeners (1899 for syllable-initial and 1241 for syllable-final consonants) and 1345 responses for Japanese listeners (940 for syllable-initial and 405 for syllable-final consonants).

Both identification and RT responses for each listener (averaged over 5 repetitions) were analyzed statistically by running multivariate ANOVAs with the factors: Language (between listeners; Russian and Japanese), Position (within-listeners; onset and coda), and Consonant (within-listeners; /p/, /p̩/, /t/, and /t̩/), and Context (within-listeners; vowel, homorganic-homorganic, hetero-organic-homorganic, homorganic-hetero-organic, and hetero-organic-hetero-organic; where the first term corresponds to the place of an adjacent consonant and the second term corresponds to its secondary articulation). A Tukey HSD post-hoc test was used to determine differences between different levels of the factors Consonant and Context. In cases of three- or four-way significant interactions, additional ANOVAs were performed within each language to clarify the statistical significance of differences in question. Only significant correct identification and RT differences (p < .01) are discussed in the presentation of the results.

To visualize perceptual results, “perceptual distance” was calculated between the categories of place (/p/, /t/ versus /p̩/, /t̩/) and secondary articulation (/p/, /t/ versus /p̩/, /t̩/). This distance (d_{ij}) was calculated as the negative of the natural log of the similarity between a pair of categories /i/ and /j/: $d_{ij} = -\ln((P_{ij} + P_{ji})/(P_{ii} + P_{jj}))$, where $P_{ij}$ is the proportion of times that the tokens of some category /i/ were perceived as some category /j/; $P_{ii}$ is the proportion of times that /i/ tokens were perceived as /i/; $P_{ji}$ is the proportion of times that /j/ tokens were perceived as /i/; and $P_{jj}$ is the proportion of times that /j/ tokens were perceived as /j/ (Johnson, 2003, pp. 66–71; based on Shepard, 1972). For example, if listeners perceived the stimulus /p/ as /p/ 90 times out of 100 ($P_{ii} = 0.90$) and as /t/ 10 times out of 100 ($P_{ij} = 0.10$), and they perceived the stimulus /t/ as /p/ 40 times out of 100 ($P_{ji} = 0.40$) and as /t/ 60 times out of 100 ($P_{jj} = 0.60$), the perceptual distance between the categories /p/ and /t/ will be 1.1 ($= -\ln((0.1 + 0.4)/(0.9 + 0.6))$). In the case of near-perfect identification both of /p/ and /t/ (99 out of 100 for each), the distance will be high (4.6); in the case of chance-level identification both of /p/ and /t/ (50 out of 100 for each), the distance will be zero. In the calculations below, if any of the proportions were either 0.00 or 1.00 (0 out of 100 or 100 out of 100), they were changed to 0.01 and 0.99 respectively. This was done to set the maximum value of distance (4.6) and to avoid the un-interpretable calculation of the natural log of zero.
3.4
Production effects and perception

Some details of the production of the stimuli may be of importance to our perceptual experiment, and thus require special attention. These issues include the timing of secondary articulation gestures, gradient palatalization or depalatalization effects, the timing of primary gestures in stop clusters, stress, and the production of nonsense words.

Timing of palatalization gestures. As has been mentioned above, the production of palatalized consonants involves raising and fronting of the tongue body to the hard palate. The achievement of the two targets is not simultaneous: for the Russian syllable-initial /p\j/ the tongue body achieves its target at the release of the constriction of the lip aperture gesture (Kochetov, to appear; cf., Ladefoged & Maddieson, 1996, pp. 363–365); for the syllable-final /p\j/, the tongue body gesture achieves its target closer to the onset of lip aperture constriction rather than to its release. In addition, the tongue body gesture in syllable coda tends to be somewhat lower and less front than the same gesture in syllable onset (more for /p\j/ than for /t\j/). Given this timing, most of the acoustic information about the secondary articulation of the onset /p\j/ in [a#p\ja] is in the following vowel (as manifested in F2 lowering throughout the vowel); most of the acoustic information about the coda /p\j/ in [ap\j#a] tends to be in the preceding vowel (as manifested in F2 rising throughout the vowel; Kochetov, 2002). For both syllable-initial and syllable-final /t\j/ (in [a#t\ja] and [at\j#a]), the acoustic information is fairly evenly distributed across the preceding and following vowels (rising and lowering F2), given the near-simultaneous timing of the tongue tip and the tongue body; the tongue body gesture for /t\j/ is also significantly higher than the same gesture for /p\j/ (and thus having higher F2). These differences between positions and consonants may have an effect on consonant identification.

Palatalization/depalatalization effects. Although no apparent phonotactic restrictions apply to consonant clusters across word boundaries, we may expect some gradient co-articulation effects in the clusters hetero-organic in terms of their secondary articulation. This is because secondary articulation gestures tend to have relatively long duration: the movement towards the tongue body target of the /p\j/ in [ta p\japi] in the current stimuli begins as early as at the onset of the preceding vowel and it is about twice as long as the movement towards the lip aperture target of the same consonant. An analysis of the movement of the tongue body data for the current speaker (as well as for 2 other Russian speakers; Kochetov, 2002) did find a substantial gradient effect, reminiscent of the word-internal regressive palatalization process (CCj \rightarrow CjCj). On average, the tongue body measured at the onset of the C1 closure in CCj was significantly higher and more front than the same articulator measured at the onset of the C1 closure in CC, yet it was significantly lower and less front than the tongue body at the onset of C1 in CjCj. Based on the overall range of tongue body

---

3 I am particularly thankful to the reviewers for raising the important issues of palatalization assimilation in clusters, inter-gestural timing in consonant clusters, syllabification, and nonsense word production.
values in these clusters, the tongue body fronting and raising effect at the onset of $CC_j$ was about 55%, thus in terms of this parameter, $C_1$ in $CC_j$ was intermediate between the plain and palatalized categories. Other gradient effects, such as the “regressive depalatalization” of $C_1$ (tongue body lowering/backing by about 15%) in $C_1C$, as well as the “progressive palatalization” of $C_2$ in $C_1C$ (tongue body raising/fronting by about 15%) were also observed; yet these effects did not reach the significance level. We may expect these gradient effects, and particularly regressive palatalization, to affect the perception of consonants in these clusters.

*Timing of primary gestures.* The analysis of the stimuli showed that the lip aperture and tongue tip gestures were somewhat overlapped in the word-boundary labial-coronal clusters /pt, pt/, /tj, tj/, /tt, tjt/, of the stimuli, while no overlap was found in coronal-labial sequences /tp, tp/, /tj, tj/ (cf., Kochetov & Goldstein, 2001; Zsiga, 2000). (The degree of overlap was measured as the relative timing of the offset of the $C_1$ constriction and the onset of the $C_2$ constriction). The presence or absence of overlap correlated with the presence or absence of audible releases: $C_1$ in coronal-labial tokens tended to be audibly released, while no release was more common for labial-coronal tokens. There was no $C_1$ release in clusters homorganic with respect to place (pp, pp/, tt, tt/, etc.), which had single lip or tongue tip constrictions, almost twice as long as constrictions of single stops. The fact that coronals were released in hetero-organic clusters could contribute to their better perceptibility in this context.

*Syllable affiliation and vowel length.* Although no deliberate attempt was made by the speaker to produce intervocalic syllable-initial and syllable-final consonants differently ([ta Capi] versus [ta C api]), an analysis of the data has revealed some articulatory and acoustic differences in position. In addition to the positional differences in the timing of the tongue body discussed above, the tongue body for /pi/ and /ti/, the tongue tip for /tj/ and /tji/, and the lips for /p/ had somewhat different magnitude and duration values in onset and coda (see Kochetov, 2002). An additional acoustic examination of the stimuli revealed that the first vowel in [ta Capi] was consistently shorter than the first vowel in [ta C api] by about 30ms; the second vowel in [ta C api] was occasionally pre-glottalized (at least in one token of each utterance); it was also longer than the first vowel in [ta C api] by about 20ms. Vowel duration differences were not limited to single consonants. The first vowel in [ta C1 C2 api] tended to be shorter than the second vowel by about 20ms, possibly due to a higher level of stress on the second vowel (the vowel intensity values, however, did not appear to be different). It is not clear whether such vowel duration differences would have any effect on the identification of onset and coda consonants.

*Production of nonsense words.* It has been noted that speakers’ degree of familiarity with stimuli may be a factor in speech production, and that the production of less familiar or nonsense words may differ from the production of more familiar or real words (Wright, 2003). These differences may involve duration and magnitude of gestures or vowel spectra. An examination of articulatory properties of some of the nonsense utterances used in the current study and real word utterances collected from the same speaker (Kochetov, to appear) showed a number of significant differences: nonsense words showed less overlap in stop clusters than real words (e.g., ta[t#p]apy versus bra[t#p]adaja ‘brother, while falling’); they also had somewhat longer lip
aperture or tongue tip constrictions in coda, as well as higher and more front tongue body gesture for [p]. These differences suggest that nonsense words were to some degree hyper-articulated compared to the corresponding real words (cf., Wright, 2003), which may be expected to have a positive effect, if any, on the identification of consonant in the stimuli.

We will return to some of the above described production effects in the discussion of the perceptual results.

4 Results

Recall that language-independent influences on perception are likely to manifest themselves in significant differences in terms of syllable position (Position), while language-specific influences are likely to manifest themselves mainly in significant differences in terms of segmental contexts (Language × Context interactions). Significant interactions of Language, Position, and Context would point to the combined effects of language-independent and language-specific influences. Significant differences and interactions involving the Consonant factor may suggest either language-independent or language-specific differences between consonants or their positional/contextual variants. Below, the differences in Position will be presented first, followed by differences in Context, separately for correct identification and reaction time.

4.1 Differences between positions: Correct identification

The phoneme identification results are summarized in Table 4 as confusion matrices: the numbers of tokens for which a given consonant was perceived as ‘p’, ‘pj’, ‘t’, or ‘tj’. The results are given separately for onset and coda consonants and for Russian and Japanese listeners. The ratio of correct responses is calculated for each consonant (as a proportion of correct responses out of 1.00). Table 5 presents the ANOVA results for correct identification: main effects and interactions. Significant effects and interactions (p < .05) are given in bold.

Correct identification was significantly different for the two language groups (Language): Overall, Russian listeners showed higher correct identification rate than Japanese listeners (on average, 0.79 vs. 0.68). However, these differences were significant mainly in syllable-final position (Language × Position): While correct identification rates in onset were only slightly better for Russian listeners than for Japanese listeners (0.96 vs. 0.94; p < .05), the rates in coda were much higher for Russians than for the Japanese (0.62 vs. 0.41; p < .001). The effect of Position was highly significant for both language groups, indicating an abrupt drop in correct responses in coda by 0.34 for Russian listeners and by 0.53 for Japanese listeners (both p < .001). There were also significant Language × Consonant × Position interactions (discussed below).
Table 4
Confusion matrices of responses for /p/, /p[^j]/, /t/, and /t[^j]/ and mean correct identification rates and SD for these consonants in syllable onset and coda for Russian (N = 20) and Japanese listeners (N = 10)

<table>
<thead>
<tr>
<th></th>
<th>Russian responses</th>
<th></th>
<th>Japanese responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>onset</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/p/</td>
<td>/p[^j]/</td>
<td>/t/</td>
</tr>
<tr>
<td>onset</td>
<td>468 10 12 4 0.95 (0.11)</td>
<td>244 1 4 0 0.98 (0.06)</td>
<td></td>
</tr>
<tr>
<td>/p[^j]/</td>
<td>3 475 1 10 0.97 (0.08)</td>
<td>8 234 2 5 0.94 (0.13)</td>
<td></td>
</tr>
<tr>
<td>/t/</td>
<td>11 0 458 24 0.93 (0.13)</td>
<td>5 3 221 20 0.89 (0.22)</td>
<td></td>
</tr>
<tr>
<td>/t[^j]/</td>
<td>0 1 1 498 1.00 (0.02)</td>
<td>1 8 0 241 0.96 (0.10)</td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>0.96 (0.10)</td>
<td>0.94 (0.14)</td>
<td></td>
</tr>
</tbody>
</table>

|        | coda              |        |                  |
|        | /p/               | /p[^j]/ | /t/   | /t[^j]/ | correct |        | /p/               | /p[^j]/ | /t/   | /t[^j]/ | correct |
| coda   | 348 94 37 19 0.70 (0.31) | 162 37 18 31 0.65 (0.30) |
| /p[^j]/| 183 222 29 64 0.45 (0.32) | 120 60 23 45 0.24 (0.25) |
| /t/    | 53 33 276 135 0.56 (0.33) | 99 27 70 48 0.29 (0.24) |
| /t[^j]/| 26 27 52 395 0.79 (0.28) | 60 34 35 113 0.47 (0.33) |
| mean   | 0.62 (0.34)       | 0.41 (0.33) |

Table 5
ANOVA results for correct identification (main effects and interactions), Position and Consonant; significant effects and interactions (p < .05) are given in bold

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>DF</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>1</td>
<td>89.840</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Consonant</td>
<td>3</td>
<td>61.620</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Position</td>
<td>1</td>
<td>1359.568</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Language × Consonant</td>
<td>3</td>
<td>11.338</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Language × Position</td>
<td>1</td>
<td>61.367</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Consonant × Position</td>
<td>3</td>
<td>39.875</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Language × Consonant × Position</td>
<td>3</td>
<td>3.820</td>
<td>.01</td>
</tr>
</tbody>
</table>

The differences between consonants (Consonant) were also significant: overall, /p/ and /t[^j]/ showed substantially higher correct identification rates than /p[^j]/ and /t/ (all p < .001). There were, however, significant Consonant × Position, Language ×
Consonant, and Language × Consonant × Position interactions. These were analyzed using additional ANOVAs separately for each language, position, and consonant.

**Consonant × Position by Language.** In onset, Russian listeners showed better identification rates for palatalized consonants than for plain consonants, specifically, better rates for /t\i/ than for /p\i/ and /t/ (p < .001 – .01) and better rates for /p\i/ than for /t/ (p < .05). Note that there were only two incorrect responses out of 500 for the onset /t\i/ (the rate of 0.996), while there were 35 of incorrect responses and seven non-responses for /t/ (the rate of 0.929). Japanese listeners showed better identification rates for /t\i/ and /p\i/ than for /t/ (p < .01 – .05). In coda, Russian listeners showed better identification rates for /t\i/ than for /p\i/ and /t/ (both p < .001), as well as better identification rates for /p\i/ than for /p\i/ and /t/ (p < .001 – .01). So did Japanese listeners: Identification rates were better for /t\i/ than for /p\i/ and /t/ (p < .001 – .01). In addition, Japanese listeners showed better identification rates for /p\i/ than for /t\i/ (p < .01). Notably, the rates for the poorer identified consonants /p\i/ and /t/ in coda in Japanese responses were near or below the chance level of 0.25. In sum, consonant differences in both language groups mainly involved poorer identification of /t/ both in onset and coda, and /p\i/ in coda. All the Consonant effects showed significant Consonant × Context interactions (described in the next section). Position was significant for all consonants in both languages (all p < .001), showing significant interactions with Context, with the exception of /p\i/ for Japanese listeners.

**Language × Consonant by Position.** The differences between languages mentioned above were modified by Consonant. In onset, Russian listeners showed better correct identification than Japanese listeners for the coronals, /t/ (p < .05; Language × Context interaction; see below) and /t\i/ (p < .01); however, Japanese listeners did better than Russian listeners in the identification of /p\i/ (p < .05). In coda, Russian listeners showed better correct identification than Japanese listeners for /p\i/ (p < .001), for /t/ (p < .001; Language × Context interaction), and for /t\i/ (p < .001; Language × Context interaction).

The graphs (a) and (b) in Figure 1 show perceptual maps for the consonants as perceived by Russian and Japanese listeners, averaged across positions. They are based on the measure of perceptual distance that takes into account both correct and incorrect responses (hits and false alarms in Table 4). Some two categories are maximally apart in the listener’s “perceptual space,” if the distance between them is as high as 4.6; the distance value of 0 reflects the categories’ complete confusability. The four consonant categories in Figure 1 are fairly dispersed in the perceptual space, however, more so for Russian (a) than for Japanese (b) listeners. The distances between plain and palatalized consonants (the horizontal dimension) are somewhat smaller than the distances between labials and coronals (the vertical dimension) for Russian listeners (/p\i/ vs. /p\i/: 1.65; /t/ vs. /t\i/: 2.03; /p/ vs. /t/: 2.62; /p\i/ vs. /t\i/: 2.74). This reflects the overall higher rate of confusions in secondary articulation compared to confusions in place. For Japanese listeners, the two contrasts are of almost the same magnitude (/p\i/ vs. /p\i/: 1.44; /t/ vs. /t\i/: 1.83; /p/ vs. /t/: 1.71; /p\i/ vs. /t\i/: 1.95), reflecting the relative frequency of both secondary articulation and place confusions.

Similarities between the two languages become more apparent when we consider the change in perceptual distance across different positions in the graphs (c) and (d) in Figure 1. Here the secondary articulation contrast (combined /p/ versus /p\i/ and
Perception of place and secondary articulation contrasts

Figure 1
The perceptual maps indicating distance between the Russian stops /p/, /pʲ/, /t/, /tʲ/ (mean across positions) as perceived by Russian (a) and Japanese (b) listeners. The consonants in both maps are rotated so that the plane between /p/ and /pʲ/ is horizontal. Positional differences in perceptual distances between plain and palatalized stops (combined /p/ versus /p__;/ pʲ/ and /t/ versus /t__;/ tʲ/) and between labial and coronal stops (combined /p/ versus /t/ and /p__;/ t__) in syllable onset and coda as perceived by Russian (c) and Japanese (d) listeners. (Distance values are calculated based on the confusion data in Table 4)

/t/ versus /t__;/ t__) and the place contrast (combined /p/ versus /t/ and /p__;/ t__) are plotted in onset and coda for both Russian and Japanese listeners. Perceptual distance values in onset are almost the same for both language groups, and are almost as high as the maximum distance value of 4.6 (Russian: 3.96 for secondary articulation and 3.91 for place; Japanese: 3.82 for secondary articulation and 3.64 for place). For both languages, distance values in coda are substantially lower, with somewhat lower values for Japanese (Russian: 1.06 for secondary articulation and 1.93 for place; Japanese: 0.58 for secondary articulation and 0.76 for place). The smaller distance between
plain and palatalized consonants in coda is in large part due to the high rate of mis-
identification of /pʲ/ as /p/ (Russian: 37% of all coda /pʲ/ tokens; Japanese: 48% of
all coda /pʲ/ tokens) and /t/ as /tʲ/ (Russian: 27% of all coda /t/ tokens; Japanese:
20% of all coda /t/ tokens) (see Table 4). The smaller distance between labials and
coronals in coda reflects the overall high confusability of the place distinction in
this position. This was particularly true for Japanese: Japanese listeners mis-
identified both /t/ and /tʲ/ as /p/ substantially more often (52% of all coda /t/ tokens and 25%
of all coda /tʲ/ tokens) than Russian listeners (10% of all coda /t/ tokens and 5% of
all coda /tʲ/ tokens).

4.2
Differences between contexts: Correct identification

In Table 6, the ratio of correct responses is calculated for each segmental context
(a–e), arranged by position and language (adapted from Kochetov, 2002: pp. 137, 149).
Table 7 presents the ANOVA results for correct identification.

<table>
<thead>
<tr>
<th>Context</th>
<th>Onset</th>
<th>Coda</th>
<th>Onset</th>
<th>Coda</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. V</td>
<td>0.99 (0.05)</td>
<td>0.75 (0.31)</td>
<td>0.96 (0.12)</td>
<td>0.52 (0.32)</td>
</tr>
<tr>
<td>b. C_{hm-hm}</td>
<td>0.99 (0.04)</td>
<td>0.76 (0.26)</td>
<td>0.96 (0.12)</td>
<td>0.55 (0.32)</td>
</tr>
<tr>
<td>c. C_{ht-hm}</td>
<td>0.95 (0.11)</td>
<td>0.66 (0.31)</td>
<td>0.97 (0.10)</td>
<td>0.41 (0.35)</td>
</tr>
<tr>
<td>d. C_{hm-ht}</td>
<td>0.95 (0.10)</td>
<td>0.35 (0.29)</td>
<td>0.98 (0.06)</td>
<td>0.30 (0.29)</td>
</tr>
<tr>
<td>e. C_{ht-ht}</td>
<td>0.94 (0.14)</td>
<td>0.60 (0.33)</td>
<td>0.86 (0.22)</td>
<td>0.33 (0.31)</td>
</tr>
</tbody>
</table>

There was a main effect of Context, indicating significant differences between all
the contexts, except the vocalic (V) and the homorganic-homorganic (C_{hm-hm}) contexts.
Overall, correct identification responses for contexts decreased in the following order:
V, C_{hm-hm} > C_{ht-hm} > C_{hm-ht} > C_{ht-ht} (p < .001 – .01). The Context effects, however,
showed significant interactions with Language, Position, and Consonant. Significant
effects of Context examined separately by Position, Consonant, and Language are
presented below.

Context × Position by Language. In onset, Russian listeners showed higher correct
identification rate in the vocalic and homorganic-homorganic contexts than for
consonants in the hetero-organic-hetero-organic context (i.e., V, C_hm-hm > C_h-t-h; p < .01 – .05; not significant for C_hm-hm and C_h-t-h). For Japanese listeners, correct identification was significantly lower in the hetero-organic-hetero-organic context compared to all the other contexts (i.e., V, C_hm-hm, C_h-t-h, C_hm-h-t > C_h-t-h; p < .001 – .01). In coda, Russian listeners showed significantly lower correct identification rate in the homorganic-hetero-organic context than in the other four contexts (i.e., V, C_hm-hm, C_h-t-h, C_hm-h-t > C_h-t-h; all p < .001). Japanese listeners showed significantly lower correct identification rate in the contexts hetero-organic with respect to secondary articulation than in the vocalic and homorganic-homorganic contexts (i.e., V, C_hm-hm > C_hm-h-t, C_h-t-h; all p < .001). Position was significant for each of the contexts in both languages (all p < .001), while showing some significant interactions with Consonant. Russian listeners showed significantly better correct identification than Japanese listeners (Language) in each of the contexts in both positions (all p < .001), except the homorganic-hetero-organic context (C_hm-h-t) in onset and coda and the hetero-organic-homorganic context (C_h-t-hm) in coda, where neither of the language groups performed better than the other. There were significant interactions with Consonant in some of the contexts (V and C_hm-hm in onset; C_h-t-hm and C_h-t-h in coda).

Context × Consonant by Language. Consonants in both languages varied in their sensitivity to different contexts. For Russian listeners, /t/ showed significantly different identification rates in almost all contexts (decreasing in the following order: t > tt > t[p, pt, tp], p[t] > t[t]; here and below the consonants of interest in clusters are shown in bold; p < .001 – .01); correct identification rates for /p/ and /t/ were significantly lower in the homorganic-hetero-organic context than in all the other contexts (i.e., p, pp, pt, tp, t[p], p[t], t[p], p[t] > p[t], t[t]; all p < .001 – .01); and correct identification for /p/ was higher in the homorganic-homorganic context than in all the other contexts (i.e., p[p] > p[t], p[t], t[p], t[p], p[p], p[t], p[t]; all p < .001 – .01). For Japanese listeners, /p/ was not significantly affected by Context; correct identification for /t/ was lower in the hetero-organic-hetero-organic context than in all the other

### Table 7

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>DF</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td>4</td>
<td>32.999</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Language × Context</td>
<td>4</td>
<td>5.342</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Consonant × Context</td>
<td>12</td>
<td>7.620</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Position × Context</td>
<td>4</td>
<td>24.826</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Language × Position × Context</td>
<td>4</td>
<td>1.927</td>
<td>.104</td>
</tr>
<tr>
<td>Language × Consonant × Context</td>
<td>12</td>
<td>3.280</td>
<td>.001</td>
</tr>
<tr>
<td>Consonant × Position × Context</td>
<td>12</td>
<td>4.591</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Language × Consonant × Position × Context</td>
<td>12</td>
<td>2.395</td>
<td>.005</td>
</tr>
</tbody>
</table>
contexts (i.e., t, tt, tp, pt, ttj, tj; all p < .05); for /tj/, it was lower in the hetero-organic-homorganic context than in the vocalic and homorganic-homorganic contexts (i.e., tj, tjt > tpj, pjt; both p < .05; not significant for the other two contexts: tj, ptj, tj, tj; for /pj/, it was higher in the homorganic-homorganic context than in all the other contexts (i.e., pjp > pj, pij, tjp, ppj, pjt, tpj; p < .01 – .05).

Thus in both languages, /t/ was identified poorer next to consonants hetero-organic with respect to place and secondary articulation (tpj and pjt) than next to a vowel or next to a consonant homorganic with respect to place and secondary articulation (ttj); /pj/ was identified better only next to a consonant homorganic with respect to place and secondary articulation (pjt). Similarly, /tj/ (t; in Russian and Japanese) and /pj/ (pp; in Russian) showed better correct identification rates in the homorganic-homorganic clusters than in the clusters hetero-organic with respect to either place or secondary articulation. Russian listeners showed particularly low correct identification rates for /t/, /p/, and /t/ in homorganic-hetero-organic clusters (tj, tj, ppj, pjp, tpj, tpj). Japanese listeners showed particularly low correct identification rates for the coronals /t/ and /t/ in hetero-organic-hetero-organic clusters (tpj, pjt, tjp, pjt).

Context × Consonant × Position by Language. Recall that in onset, Russian listeners showed better correct identification rate for the palatalized consonants /t/ and /p/ compared to the plain /t/; these differences turned out to be significant in the homorganic-hetero-organic and hetero-organic-hetero-organic contexts only (i.e., tj, ptj, ppj, tj > tj, pjt; p < .001 – .01). The better performance of Japanese listeners with /t/ and /p/ in onset compared to /t/ in onset was significant in the hetero-organic-hetero-organic context only (i.e., ptj, tj > pjt; both p < .001). In coda, both groups of listeners showed better identification rates for /t/ and /p/ than for /p/ and /t/. In terms of Context, however, the following significant differences were found for Russian listeners: in the vocalic context, /p/, /t/, and /t/ were identified better than /p/ (i.e., coda p, tj > coda p; all p < .001); in the homorganic-homorganic context, there were no significant differences between consonants; in the homorganic-hetero-organic context, /t/ was identified better than /t/ (i.e., tj > t; p < .001); in the hetero-organic-homorganic and hetero-organic-hetero-organic contexts, /p/ and /t/ were identified better than /p/ and /t/ (i.e., pt, ptj, tjp > pjp, tpj; p < .001 – .05). The following significant differences were found for Japanese listeners: in the vocalic context, /p/ and /t/ were identified better than /p/ and /t/ (i.e., coda p, t > coda t, p; p < .001 – .01); in the homorganic-homorganic context, /t/ was identified better than /t/ (i.e., tj > tj; p < .05); in the hetero-organic-homorganic context, /p/ was identified better than all the other consonants (i.e., pt > tp, tjp, pjt; p < .001 – .01); in the hetero-organic-hetero-organic context, /p/ was identified better than /p/ and /t/ (i.e., ptj > pjt, tpj; both p < .01); there were no significant differences in the homorganic-hetero-organic context.

For Russian listeners, Position was significant for all consonants in all contexts (p < .001 – .05), except for /t/ in the vocalic context (i.e., onset t > coda t), and for /p/ in the hetero-organic-homorganic context (i.e., tp vs. pt). For Japanese listeners, Position was significant throughout: for all consonants in all contexts (p < .001 – .05).
To summarize the differences between consonants by Context and Position, both Russian and Japanese listeners showed relatively poor identification of /t/ in hetero-organic clusters with respect to place: as the onset consonant in /pt/ and as the coda consonant in /tp/ and /tp'/; it was also poorly identified by Russian listeners in homorganic-hetero-organic clusters: as the onset consonant in /tt/ and as the coda consonant in /tt'/.

Unlike, the Japanese, Russians had a near-perfect identification of the single /t/ in coda (0.97 for Russian and 0.41 for Japanese). Both Russians and the Japanese showed relatively poor identification of the single /p'/ in coda, as well as of the coda /p'/ in the clusters hetero-organic with respect to place (p'p and p'p'). Note should be taken of the rather poor identification of the single /p'/ in coda by Russian listeners (0.38 compared to 0.97 for /t/, 0.80 for /p/, and 0.83 for /t'/ in the same context). The errors for /p'/ involved confusions with /p/ (32%), /t'/ (20%), and /t/ (10%). An examination of individual responses showed a high inter-speaker variation (SD of 0.28); there were listeners whose mean correct identification rate was as high as 1.00, and listeners whose rate was as low as 0.00.

Finally, both Russian and Japanese listeners showed relatively high correct identification for /p/ in clusters hetero-organic with respect to place (pt and pt'); the same was found for /t'/ for Russian listeners only (tp' and tp').

Language × Consonant × Context by Position. In onset, the previously mentioned better correct identification of /t/ by Russian listeners, compared to Japanese listeners, was limited to the hetero-organic-hetero-organic context (i.e., pt; p < .003), while the Japanese showed better identification of /t/ in the homorganic-hetero-organic context (i.e., t't; p < .05). In coda, the better correct identification of /t/ by Russian listeners, compared to Japanese listeners, was significant in the vocalic, homorganic-homorganic, and hetero-organic-hetero-organic contexts only (i.e., coda t, tt, tp'; p < .001–.01); the better correct identification of /t'/ by Russian listeners than Japanese listeners was significant in the hetero-organic-homorganic and hetero-organic-hetero-organic contexts only (i.e., t'p, t'p'; both p < .001).

Overall, contextual differences tended to be consonant-specific; they were substantially less robust than differences in position. Graphs in Figure 2 compare perceptual distance values between plain/palatalized (/a, b/) and labial/coronal (/c, d/) categories in five segmental contexts — adjacent to /l, l/, /l, l/, and /t/ — arranged by position (onset and coda). The graphs are given separately for Russian (/a, c/) and Japanese (/b, d/) listeners. It is noteworthy that, distance values are affected by certain contexts (e.g., the effect of the preceding /p'/ on the secondary articulation contrast in onset); yet, these effects are far less robust or consistent across both languages than the effect of syllable position. In the onset contexts, both secondary articulation and place contrasts are almost maximally dispersed in the

4 The analysis of written responses during the pre-test showed that both Russian and Japanese listeners exhibited errors in the syllabification and/or segmentation of single syllable-final stops, particularly the palatalized labial. Thus, the single /p'/ was at times interpreted as a sequence of the hetero-syllabic /l/ and /p/. That is /tap' api/ was rendered as taj papy (in Cyrillic) and as taipapui (in Katakana), instead of tap' api and tapi apui. In a few cases, the second articulation of the syllable-final labial in /tap' api/ was attributed to the preceding non-adjacent /l/ (written as tja papy in Cyrillic and chapapui in Katakana) (Kochetov, 2005).
listeners’ perceptual space; in the coda contexts, the contrasts are substantially lower. Note the similarity between Russian and Japanese in perceptual distance values for the plain/palatalized contrast, reflecting relatively similar confusion patterns for the contrast. Note also the difference between the two languages in perceptual distance values for the labial/coronal values in coda, reflecting rather higher confusion rates for Japanese than Russian. (Note that the distance values in the vocalic context in coda for Russian listeners would have been somewhat higher, if not for the high rate of confusions involving /p/; compare the values in the vocalic context in (a) and (b) to the distance between the single /t/ and /t/-, 2.42, and to the distance between the single /p/ and /t/-, 2.69, in the same context.)

Figure 2
Perceptual distances between plain and palatalized stops (combined /p/ versus /pj/ and /t/ versus /tj/) by position and context (before/after /a/, /p/, /t/, /pj/, /tj/), as perceived by Russian (a) and Japanese (b) listeners. Perceptual distances between labial and coronal stops (combined /p/ versus /t/ and /pj/ versus /tj/) by position and context (before/after /a/, /p/, /t/, /pj/, /tj/), as perceived by Russian (c) and Japanese (d) listeners.

4.3 Differences between positions: Reaction time
Mean RT values for each consonant in a given position are presented in Table 8 separately for Russian and Japanese listeners. The ANOVA results for reaction time are given in Table 9.
Table 8

Mean reaction times (in ms) and SD for /p/, /p̆/, /t/, and /t̆/ in syllable onset and coda for Russian (N = 20) and Japanese (N = 10) listeners

<table>
<thead>
<tr>
<th>Position</th>
<th>Russian RT</th>
<th></th>
<th></th>
<th></th>
<th>Japanese RT</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p</td>
<td>p̆</td>
<td>t</td>
<td>t̆</td>
<td>mean</td>
<td>p</td>
<td>p̆</td>
<td>t</td>
</tr>
<tr>
<td>onset</td>
<td>1478</td>
<td>1416</td>
<td>1641</td>
<td>1494</td>
<td>1507</td>
<td>1302</td>
<td>1332</td>
<td>1430</td>
</tr>
<tr>
<td></td>
<td>(319)</td>
<td>(278)</td>
<td>(364)</td>
<td>(327)</td>
<td>(333)</td>
<td>(235)</td>
<td>(266)</td>
<td>(373)</td>
</tr>
<tr>
<td>coda</td>
<td>1509</td>
<td>1602</td>
<td>1496</td>
<td>1458</td>
<td>1516</td>
<td>1521</td>
<td>1565</td>
<td>1864</td>
</tr>
<tr>
<td></td>
<td>(387)</td>
<td>(512)</td>
<td>(384)</td>
<td>(374)</td>
<td>(418)</td>
<td>(477)</td>
<td>(564)</td>
<td>(650)</td>
</tr>
</tbody>
</table>

Table 9

ANOVA results for reaction time, Position and Consonant; significant effects and interactions (p < .05) are given in bold

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>DF</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>1</td>
<td>0.477</td>
<td>.490</td>
</tr>
<tr>
<td>Consonant</td>
<td>3</td>
<td>7.469</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Position</td>
<td>1</td>
<td>42.392</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Language × Consonant</td>
<td>3</td>
<td>4.389</td>
<td>.004</td>
</tr>
<tr>
<td>Language × Position</td>
<td>1</td>
<td>37.927</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Consonant × Position</td>
<td>3</td>
<td>0.575</td>
<td>.632</td>
</tr>
<tr>
<td>Language × Consonant × Position</td>
<td>3</td>
<td>5.347</td>
<td>.001</td>
</tr>
</tbody>
</table>

The overall language differences in RT were not significant (Language). RT was significantly different in two positions (Position); there was, however, a significant Language × Position interaction, pointing to highly significant position differences for the Japanese (p < .001), but not for Russian listeners (p > .05). Specifically, it took Japanese listeners substantially longer to correctly identify consonants in coda, compared to the same consonants in onset (by about 325 ms; p < .001). Although the overall language differences were not significant (Language), separate tests for onset and coda position revealed the following significant differences: Japanese listeners showed lower RT values in syllable onset (by about 140 ms; p < .001) and higher RT values in syllable coda (by about 180 ms; p < .001) compared to Russian listeners. All these differences were significant for some consonants and contexts only (as described below).

There were significant differences between consonants (Consonant): overall, RT values for the coronal /t/ were higher than the values for labials /p/ and /p̆/ (p < .01 – .05). There were, however, significant Language × Consonant and Language × Consonant × Position interactions. An examination of the effects within the three-way...
interaction revealed the following. In \textit{onset}, Russian listeners showed higher RT values for \(/t/\) compared to the other consonants, \(/p/\), \(/p^j/\), and \(/t^j/\) \((p < .001 - .01)\) (Consonant \(\times\) Context interaction; see below). The Consonant effect for Japanese listeners did not reach the significance level \((p = .055)\), while showing a significant Consonant \(\times\) Context interaction. In \textit{coda}, there were no significant consonant differences for Russian listeners, while Japanese listeners showed higher RT values for the \textit{coda} \(/t/\) compared to the \textit{coda} \(/p/\) \((p < .05)\).

The shorter RT for Japanese listeners than for Russian listeners in onset (as determined by the Language \(\times\) Position interaction) was limited to the plain consonants \(/p/\) and \(/t/\) \((p < .001 - .01)\); the longer RT in coda for Japanese listeners than for Russian listeners was limited to the coronals \(/t/\) and \(/t^j/\) \((p < .001)\). There were no significant Language \(\times\) Context interactions. In sum, consonant differences within both languages involved higher RT for \(/t/\), either in onset (Russian), or in coda (Japanese). The differences between the languages involved the onset and coda \(/t/\), onset \(/p/\), and coda \(/t^j/\).

Note that the relatively weak effect of Position in RT results, compared to correct identification results, could be due to the fact that the acoustic events relevant to the identification of onset and coda consonants are, obviously, not distributed in the acoustic signal at the same intervals of time: Most of the information about coda consonants would appear earlier in the signal than most of the information about onset consonants.

4.4 \textbf{Differences between contexts: Reaction time}

Mean RT values for each segmental context (a–e), arranged by position and language are presented in Table 10 separately for Russian and Japanese listeners. The ANOVA results for reaction time are given in Table 11.

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|c|c|}
\hline
\textbf{Context} & \textbf{Onset} & \textbf{Coda} & \textbf{Onset} & \textbf{Coda} \\
\hline
\textbf{Russian} & & & & \\
\hline
a. V & 1425 (297) & 1589 (464) & 1250 (224) & 1840 (575) \\
\hline
b. C_{hm-hm} & 1479 (312) & 1377 (325) & 1337 (228) & 1749 (653) \\
\hline
c. C_{ht-hm} & 1553 (353) & 1492 (408) & 1395 (280) & 1628 (585) \\
\hline
d. C_{hm-ht} & 1525 (347) & 1563 (459) & 1360 (194) & 1659 (546) \\
\hline
e. C_{ht-ht} & 1556 (342) & 1562 (402) & 1495 (428) & 1582 (589) \\
\hline
\end{tabular}
\end{table}

\textit{Language and Speech}
The overall differences in Context were not significant. There were no significant Language × Context and Consonant × Context interactions, while there were significant Position × Context and Consonant × Position × Context interactions. The first interaction indicated that the positional differences in RT—higher RT in coda than in onset—were significant in the vocalic (p < .001) and homorganic-hetero-organic (p < .05) contexts only. An examination of the second interaction revealed the following. For Russian listeners, RT was significantly higher for the onset /t/ in the hetero-organic-hetero-organic context than for the homorganic-homorganic context (i.e., /tt/ > /tp/; where “>” stands for ‘shorter RT’; p < .05); RT was also higher for single /pʲ/ in coda than in the homorganic-homorganic context (i.e., /pʲpʲ/ > /pʲ/; p < .05). For Japanese listeners, RT was significantly higher for the onset /t/ in the hetero-organic-hetero-organic context than in all the other contexts (i.e., t, tt, pt, tʲt > /pt/; all p < .01).

Recall that Language × Consonant × Position interaction revealed that Russian listeners showed higher RT for /t/ in onset than for /p/, /pʲ/, and /tʲ/ in onset (p < .001 – .01). These differences, however, were limited to /t/ versus /pʲ/ in the hetero-organic-homorganic context (i.e., /tp/ > /pt/; p < .01) and /t/ versus /p/ and /tʲ/ in the hetero-organic-hetero-organic context (i.e., /tp/, /tpʲ/, /ptʲ/ > /pʲt/; all p < .001). Japanese listeners also showed higher RT for the onset /t/ than for the onset /p/, /pʲ/, and /tʲ/; it was significant in the hetero-organic-hetero-organic context (i.e., /tp/, /tpʲ/, /ptʲ/ > /pʲt/; p < .01 – .05). In coda, no significant consonant differences were found for Russian listeners. Although the single /pʲ/ in coda showed a substantially higher RT than the other single consonants in coda (by on average 240ms), these differences were not significant, possibly due to the high range of variation in the /pʲ/ responses (SD of 610ms). The higher RT for the coda /t/ than for the coda /p/ in Japanese responses did not show interactions with Context. In sum, almost all RT differences in Context in both Japanese and Russian responses involved /t/, mainly next to consonants hetero-organic with respect to place (the hetero-organic-homorganic and hetero-organic-hetero-organic contexts).

Language and Speech
For Russian listeners, Position was significant only for /pj/ in the vocalic and hetero-organic-hetero-organic contexts (i.e., onset pj > coda pj, tpj > pj t; p < .01 – .05) and for /t/ in the hetero-organic-hetero-organic context (i.e., coda > onset: tpj > ptj; p < .01). For Japanese listeners, Position was significant for /p/ in the vocalic context (i.e., onset p > coda p; p < .01), for /t/ in the vocalic, homorganic-homorganic, and homorganic-hetero-organic contexts (i.e., onset t > coda t, tt > tt, tj > tj; p < .001 – .05), and for /tj/ in the vocalic and homorganic-hetero-organic contexts (i.e., onset tj > coda tj, ttj > tj t; p < .05). Thus, in both languages, RT for /t/ was sensitive to position, however, in different contexts and in the opposite direction (coda > onset for Russian and onset > coda in Japanese).

5 Discussion

The main results of the experiment, organized by position, context, consonant, and language, are discussed with respect to the predictions outlined in Section 2.3.

5.1 Differences between positions

Our main prediction was that both Russian and Japanese listeners would perform better overall with onset consonants than with coda consonants, but their performance in each position may vary depending on which contexts are phonotactically possible or impossible in their languages. The results indeed demonstrate that both Russian and Japanese listeners performed exceedingly well with syllable onset consonants. Both groups of listeners showed a dramatic drop in correct identification of both secondary articulation and place contrasts in syllable coda. These positional differences in correct identification involved almost all consonants in all contexts. Positional differences were also reflected in longer reaction time in coda than in onset in both Russian and Japanese responses (some consonants in some contexts). The relatively low performance of Russian listeners with coda consonants can hardly be attributed to their context-specific phonotactic knowledge, since four out of five coda contexts are phonotactically possible within words in Russian. Similarly, the higher than chance performance of Japanese listeners can hardly be attributed solely to their phonotactic knowledge, since all, except one of the coda contexts, are phonotactically impossible in Japanese. Of particular interest is the finding that the effect of position on correct identification was observed even in the contexts that have no phonotactic restrictions in the languages, such as the context homorganic with respect to place and secondary articulation (C_{hm-hm}). The results thus confirm the largely language-independent influence of syllable position on perception of consonant contrasts: the syllable-initial position is intrinsically more perceptually salient than the syllable-final position. Note that the somewhat longer vowels in onset syllables compared to coda syllables may have contributed to the better identification of onset consonants; however, this effect is unlikely to have been large (see Section 3.4).
5.2
Differences between contexts

Our prediction with respect to contexts was that in both positions, the performance of Russian listeners would decrease from context to context in the following order: V, C_{hm-hm} > C_{ht-hm}, C_{ht-ht} > C_{hm-ht}; while the performance of Japanese listeners would decrease in the following order: V (onset), C_{hm-hm} > V (coda), C_{ht-hm}, C_{ht-ht}, C_{hm-ht}. This prediction was confirmed, at least for some consonants in some positions and contexts. In onset, relatively low identification rates for both languages were found in the hetero-organic-hetero-organic context (C_{ht-ht}); however, this context effect was limited to the cluster pjt only (0.82 for Russian and 0.56 for Japanese). Although this particular cluster is phonotactically impossible (within words) in either language, the result may as well be explained by the gradient progressive palatalization effect (see Section 3.4): an overlap of the tongue tip gesture of /t/ by the tongue body gesture of the preceding /p/ results in higher F2 at the release of /t/. (Recall that the tongue body at the release of /t/ in this context was 15% higher than expected.) This would make /t/ in pjt more perceptually ambiguous than, for example, /t/ in pt. The results showed that most identification errors for pjt in Russian and Japanese responses involved confusions with /tj/. Note also that both groups showed relatively high reaction time for /t/ in pjt.

The context effects appeared to correlate better with language phonotactics in coda position. As expected, consonants in the phonotactically possible vocalic and homorganic-homorganic contexts (V and C_{hm-hm}) showed relatively high correct identification for this position in both Russian and Japanese responses. Russian listeners performed the worst with the homorganic-hetero-organic context (C_{hm-ht}), which is the only phonotactically impossible word-internal context in Russian. Unsurprisingly, most identification errors in this context involved secondary articulation: plain consonants in ppj and ttj were perceived as palatalized, while palatalized consonants in pjp and tjt were perceived as plain (more so for /p/ and /t/ than for /p/ and /t/). Japanese listeners performed the worst in the same context, as well as in the hetero-organic-hetero-organic context (C_{ht-hl}); both contexts are phonotactically impossible in Japanese. As expected, listeners’ responses showed a high percentage of errors, both in secondary articulation and place: coronals in tpj and tjp were perceived as labials, while labials in ptj and ppj were perceived as coronals (also, more so for /p/ and /t/ than for /t/ and, especially for /p/). The results for the C_{ht-hm} context were not significantly different from all the other contexts; however, this could be due to the surprisingly high identification rate for /p/ in C_{ht-hm} (pt, 0.80; compared 0.18–0.38 for the other consonants in the context).

It is important that the relatively high percentage of secondary articulation errors in Russian and Japanese responses in the homorganic-hetero-organic and hetero-organic-hetero-organic contexts (C_{hm-hl} and C_{ht-hl}) can be as well attributed to the gradient regressive palatalization and depalatalization effects in these sequences in the stimuli (see Section 3.4). This is particularly likely in the case of the plain-palatalized clusters (CCj), where the tongue body measured at the onset of the C1 closure was half as high as for a palatalized consonant (in CiCi). Note also that coda stops in the C_{hm-hl} clusters were not released (unlike in the coronal-labial C_{ht-hm} and C_{ht-hl} clusters in the stimuli) making the identification of the coda consonants more difficult.
5.3 Differences between consonants

The results also revealed substantial differences between individual consonants: overall, both Russian and Japanese listeners identified /p/ and /t/ better than /pʲ/ and /tʲ/. The poorer identification of /pʲ/ was found in coda (in some contexts), while poorer identification of /t/ was found in both onset (in some contexts) and coda (in most contexts). The reaction time results showed that Russian listeners had higher RT for the coda /pʲ/ (in some contexts) and for the onset /t/ (in some contexts), compared to other consonants; the Japanese had higher RT for /t/ in coda, compared to other consonants. Below I will examine the two most consistent consonant differences — the contrast between the palatalized /p/ and /pʲ/ and the contrast between the plain /p/ and /t/.

The lower correct identification and longer reaction time for the coda /pʲ/, compared to the coda /tʲ/, by Russian listeners can be in part attributed to the phonotactic restrictions on the consonant in Russian. The relatively high inter-speaker variation found for with the intervocalic /p/ in coda may suggest that at least some speakers did not find the nonsense item tapʲ, or its production (possibly syllabification), natural in this context. It is important, however, that the identification of /pʲ/ by Japanese listeners was rather similar to the Russian identification, and therefore the primary explanation for the asymmetry has to be sought elsewhere. In part this asymmetry can be explained by somewhat different acoustic and articulatory properties of the palatalized /pʲ/ and /tʲ/ in the context of the low vowel /a/ ([apʲ] versus [atʲ]). For both consonants, raising/fronting of the tongue body results in the F2 transitions rising throughout the vowel (from mid to high frequencies). For /pʲ/, however, the closing of the lips has the opposite effect of lowering F2 at the offset of the preceding vowel (cf., Stevens, 1998, pp.340 – 344). Given this, F2 values at the offset of the vowel in [apʲ] would be somewhat lower than the values at the offset of the vowel in [atʲ]. Thus, the coda /pʲ/, at least in this vowel context, can be considered as less perceptually salient than the coda /tʲ/. The acoustic and perceptual differences between the two are likely to be greater, given the language-particular details of the production of these consonants (see Section 3.4): the overall lower tongue body gesture for /pʲ/ (than for /tʲ/), susceptible to gradient reduction in coda. The lack of the audible release for the coda /pʲ/ in hetero-organic clusters in the stimuli, compared to the audibly released clusters with the coda /tʲ/ could have also contributed to the perceptual asymmetry. The timing of the tongue body gesture at the onset of the lip constriction found in the coda /pʲ/ (as opposed to the timing at the release of the lip constriction in the onset /pʲ/ could have been expected to enhance the perceptibility of the consonant in this position. This production strategy, however, was not sufficient to improve the correct identification of /pʲ/. In sum, the asymmetry between /pʲ/ and /tʲ/ found in the results can be largely attributed to the acoustic and articulatory properties of these consonants. These factors, in turn, can be seen as a source of cross-linguistically common avoidance of palatalized labials syllable- or word-finally, compared to palatalized coronals (Kochetov, 2002).

The better correct identification for the coda /p/, compared to the coda /t/, can also be explained by the overall spectral differences between the two consonants in the context of the low central vowel /a/ ([ap] versus [at]). The closing of the lips for /p/,
accompanied by the retraction of the tongue body (velarization), results in the lowering of F2 from mid to low frequencies. The fronting/raising of the tongue tip/lamina to the teeth/alveolar ridge for /t/ (apico-laminal denti-alveolar: Bolla, 1981) does not have a robust acoustic effect in this vowel context: F2 values at the offset of the vowel would be only slightly higher than F2 values during the steady part of the vowel (cf., Stevens, 1998, pp.355–357). Importantly, the F2 values for /t/ would be almost intermediate between the values for the plain /p/ and the palatalized /t̠/. This would not only make /t/ less perceptually salient than /p/ (at least in this vowel context), but also make it more perceptually similar to palatalized consonants compared to /p/. This explains the rather frequent confusion of /t/ with /t̠/ in a number of contexts (see Table 4), particularly when preceded or followed by palatalized consonants: in pjt, tjt, tpj, and ttj. The fact that /t/ was released in clusters hetero-organic with respect to place, did not seem have had a noticeable effect on listeners’ responses. Unlike the plain coronal /t/, the palatalized /t̠/, distinguished in this vowel context by the high F2, was more perceptible than /t/ in most of the coda contexts, as well as in some onset contexts.

Further, the patterns of place confusions revealed a robust asymmetry between /t/ and /p/ in coda: the plain coronal /t/ was confused with the following labials /p/ and /p̠/ more often than the plain labial /p/ was confused with the following coronals /t/ and /t̠/. The high rate of place confusions by the Japanese can be safely attributed to the language phonotactic restrictions on clusters hetero-organic with respect to place and to the phonological process of place assimilation (Ito & Mester, 2003); yet, this language knowledge does not seem to be sufficient to account for the particularly high susceptibility of /t/ to the “perceptual place assimilation.” The fact that Russian listeners showed the same tendency to perceptually assimilate /t/ rather than /p/, suggests that this asymmetry is to some extent language-independent. Note that the perceptual asymmetry between coda labials and alveolars has been extensively documented for English (e.g., Byrd, 1992; Surprenant & Goldstein, 1998; Wright, 2001) and often interpreted as a factor in the relative markedness of the two places of articulation (Jun, 1995; Kingston, to appear).

5.4 Differences between languages

With respect to language differences, we expected a somewhat better performance for Russian than Japanese listeners, with certain similarities and differences depending on the phonotactically possible and impossible word-internal contexts in each language. The results did show the overall better performance for Russian listeners in terms of correct identification, both in onset and coda position (in some contexts). Their better performance in onset was rather small (by 0.02) yet significant. Russian listeners’ better performance in coda position was more robust (by 0.21). Both effects were limited to certain contexts and consonants. The reaction time results pointed to the better performance of Japanese listeners in onset (some consonants) and to the better performance of Russian listeners in coda (some consonants).

In terms of contexts in onset position, Russians did better than Japanese in the vocalic and homorganic-homorganic contexts. Interestingly both contexts are phonotactically possible in both languages. They also did better with the hetero-organic-hetero-organic context (for some consonants), which is possible in Russian...
(unless C₁ is /p̊/), but not in Japanese. (Note, however, that the differences in the first two contexts were rather small, not exceeding 0.03, and the differences in the last context were largely due to the poor identification of /t/ in /p̊t/). Both groups were not different in terms of the identification of onset consonants in homorganic-hetero-organic and hetero-organic-homorganic contexts, one of which is impossible in both languages and the other one is possible in Russian only (also, unless C₁ is /p̊/).

In terms of contexts in coda position, Russians did better in the vocalic (for some consonants), homorganic-homorganic, hetero-organic-hetero-organic, hetero-organic-homorganic (for some consonants) context. Only the second context is possible in Japanese, while the other contexts are possible in Russian (with some restrictions on /p̊/ in clusters). Both groups did poorly with the homorganic-hetero-organic context, impossible in both languages (word-internally in Russian). Thus, language differences in the identification of consonants showed some correlation with listeners' phonotactic knowledge; this was more the case in coda than in onset contexts, pointing to the phonotactic relevance of following, rather than preceding segments.

Note that the poor performance of both groups with the homorganic-hetero-organic context can be also attributed to the perceptual effect of the gradient palatalization and depalatalization and the lack of audible C₁ releases in this context. The overall better performance of Russian listeners could have also been enhanced by their experience with word-final coda consonants in general (which would likely result in their higher perceptual attunement to VC transitions) and with all consonant sequences of interest at word boundaries.

In terms of individual consonants in onset, Russian listeners were better at the identification of /t/ and /t̊/ (both in certain contexts only); while Japanese listeners were better at the identification of /p/; and both groups did equally well at the identification of /p̊/. Note that the Japanese showed shorter reaction time than the Russians for the plain /p/ and /t/ consonants in onset. In coda, Russians were better at the identification of /p̊/, /t̊/, and /t̊̊/ (the last two occurring in some contexts only); the groups did not show differences in the identification of /p/. In addition, Russians showed shorter reaction time for the coda /t/ and /t̊̊/ than the Japanese. Some of these differences, particularly those related to coronals, could be in part explained by the phonetic difference between the Russian /t̊̊/ and the Japanese /tɕ/ (see Section 2.3), as well as by the difference between the Russian /t/ in coda and the corresponding Japanese syllable /to/. The better performance of Japanese listeners with /p/ in onset (both correct identification and RT), and the lack of difference for the coda /p/ may be due to the general bias in Japanese responses towards /p/ (which was particularly high in coda).

6 Conclusion

The findings of the experiment point to a complex interaction of language-specific and universal perceptual factors. The differences between the language groups are well motivated by the phonological and phonetic differences between the two languages, particularly those in the phonotactics of place and secondary articulation contrasts. There are, however, striking similarities between the two groups in the level of performance and the types of errors made when processing the same consonants in different syllable positions. These similarities are strongly indicative of universal,

—Language and Speech
language-independent perceptual asymmetries between positions (onset vs. coda) and between consonants (plain vs. palatalized, labial vs. dental/alveolar) (see Section 1). The asymmetries between positions likely result from a number of factors: (a) the articulatory properties of gestural reduction in coda and overlap in consonant clusters (Kochetov, 2002; Surprenant & Goldstein, 1998); (b) the way acoustic information is distributed throughout the signal (more information during CV transitions, bursts of released stops; Fujimura et al., 1978; Ohala, 1990); and (c) properties of human perception such as the attunement to CV transitions rather than to VC transitions (Kingston, to appear; Ohala, 1990; Redford & Diehl, 1999). As we saw, the overall effect of position was rather similar for two different types of contrasts—place and secondary articulation, suggesting that the onset/coda (or CV/VC transition) asymmetry may be rather general, and not specific to major place articulation (as, e.g., has been claimed in Steriade, 2001).

The asymmetries between consonants likely follow from the physical properties of the consonants in question: (a) the magnitude and coordination of the gestures involved in their production (Byrd, 1992; Kochetov, 2002; Surprenant & Goldstein, 1998); (b) their acoustic spectral and durational characteristics; and (c) their relative perceptibility (Jun, 1995; Kavitskaya, to appear; Kingston, to appear; Ohala, 1990; Winters, 2001; Wright, 2001).

Although the focus of the current study is the perception of specific phonological contrasts—Russian plain and palatalized labial and coronal voiceless stops—by listeners of two languages, its general implications are much wider. As an investigation of language-specific and universal factors affecting the perception of a phonological contrast, this study contributes to the recent body of work on the role of perception in phonology (see, e.g., papers in Hume & Johnson, 2001b; see also Flemming, 1995; Jun, 1995; Silverman, 1997; Steriade 2001). The findings of the study also suggest that the interactions between perception and phonology are rather complex (cf., Hume & Johnson 2001a). Phonology influences perception synchronically, resulting in language-specific properties of perceptual performance. At the same time, perception affects phonology, although rather indirectly, resulting over time in universal markedness tendencies. As we saw, the differences between Russian and Japanese in the phonology of place and secondary articulation contrasts have important consequences for the perception of the contrasts by Russian and Japanese listeners. Further, we witnessed the role of universal physical limitations and biases in the perception of this phonological contrast in different positions.

These universal factors likely play an important role in the emergence of cross-linguistic patterns of place and palatalization, namely, the common maintenance of the contrasts in syllable onsets, their common neutralization in syllable codas and the greater susceptibility to neutralization of some consonants than others, in particular, the susceptibility of plain coronals to place assimilation and of palatalized labials to coda neutralization (see Jun, 1995, on major place of articulation and Kochetov, 2002, on palatalization).
References


Downloaded from las.sagepub.com at UNIV TORONTO on July 15, 2011


