Place assimilation and phonetic grounding: A cross-linguistic perceptual study

Alexei Kochetov, University of Toronto
Connie K. So, University of Western Sydney

Available at: https://works.bepress.com/alexei_kochetov/10/
Place assimilation and phonetic grounding: a cross-linguistic perceptual study*

Alexei Kochetov
University of Toronto

Connie K. So
University of Western Sydney

This paper investigates predictions made by the ‘phonetic knowledge hypothesis’ (Jun 1995, 2004, Hayes & Steriade 2004) about the relation between perceptibility of stops and common patterns of major place assimilation. In two perceptual experiments, stimuli with Russian released and unreleased voiceless stops in clusters were presented for identification of 56 listeners, native speakers of Russian, Canadian English, Korean and Taiwanese Mandarin. Percentages of correct responses and reaction time data were used to determine scales of perceptual salience. Results reveal considerable perceptual differences between places of articulation, consistent across four language groups. Perceptual salience of place of articulation was strongly affected by presence or absence of stop releases. While the salience scale for released stops closely corresponded to cross-linguistic patterns of assimilation, the scale for unreleased stops did not. The results provide partial support for the hypothesis, while suggesting a less direct relation between scales of phonetic difficulty and phonological markedness.

1 Introduction

One of the fundamental questions in understanding human cognition is how the discrete knowledge of the world – symbolic concepts and categories – is related to continuous everyday perceptions and actions. This question, commonly known as the SYMBOL-GROUNDING PROBLEM (Harnad 1990), is relevant to all fields of cognitive science, including linguistics.

* We are grateful to Juliette Blevins, Steve Winters, the Phonology editors, an anonymous associate editor and three anonymous reviewers for insightful comments and suggestions on earlier versions of the manuscript. We would also like to thank Sonya Bird, Louis Goldstein, Hyung-Soo Kim, Keren Rice, Yue Wang and audiences of the Canadian Linguistic Association meeting and University of Victoria. This work was supported by the Social Sciences and Humanities Research Council of Canada. An earlier report based on partial results of Experiment 1 of the current paper appeared as Kochetov & So (2005).
However, the problem has been felt particularly acutely in generative phonology, with its traditionally almost exclusive focus on abstract symbolic representations and computations (Anderson 1985, Pierrehumbert et al. 2000, Port & Leary 2005). How do phonological symbols, and the cross-linguistic sound patterns they capture, relate to concrete facts of articulation and perception of speech sounds? This question has generally been considered to be beyond the scope of phonological theory proper. Some recent work within Optimality Theory (Prince & Smolensky 1993) has argued for a close connection between phonological symbols (i.e. grammatical representations and constraints), the cross-linguistic sound patterns they define (markedness laws), and phonetic facts (ease or difficulty of articulation and perception) (see e.g. many contributions to Hayes et al. 2004). According to Hayes & Steriade (2004), facts of articulatory and perceptual difficulty, mediated by speakers’ phonetic knowledge, are the ultimate source of phonological constraints and markedness laws:

The hypothesis ... is that phonological constraints can be rooted in phonetic knowledge ..., the speakers’ partial understanding of the physical conditions under which speech is produced and perceived. The source of markedness constraints as components of grammar is this knowledge. The effect phonetic knowledge has on the typology of the world’s sound systems stems from the fact that certain basic conditions governing speech perception and production are necessarily shared by all languages, experienced by all speakers, and implicitly known by all. This shared knowledge leads learners to postulate independently similar constraints. The activity of similar constraints is a source of systematic similarities among grammars and generates a structured phonological typology (Hayes & Steriade 2004: 1).

This hypothesis, further referred to as the PHONETIC KNOWLEDGE-BASED GROUNDING (PKG) approach, is summarised in (1) (from Hayes & Steriade 2004: 7).

(1) a. Facts of phonetic difficulty
   b. Speakers’ implicit knowledge of the facts in (a)
   c. Grammatical constraints induced from the knowledge in (b)
   d. Sound patterns reflecting the activity of the constraints in (c)

Universal facts of phonetic difficulty are internalised by speakers as phonetic knowledge; the speakers use this knowledge to make symbolic generalisations – i.e. grammatical constraints and their rankings; sound patterns observed across languages reflect the activity of constraint rankings in language-particular grammars. Note that explicit reference to phonetic knowledge distinguishes this hypothesis from other prominent phonetically based approaches, which can be referred to as ‘misperception-based grounding’ (Ohala 1981, 1990; cf. Evolutionary Phonology; Blevins 2004) and ‘gestural dynamics-based grounding’ (Articulatory Phonology; Browman & Goldstein 1989, 1992). The assumption that
grammatical constraints may be induced by speakers from acquired phonetic knowledge, rather than \textit{a priori} provided to them by Universal Grammar, sets the approach apart from Grounded Phonology (Archangeli & Pulleyblank 1994) and standard Optimality Theory (Prince & Smolensky 1993).

As Hayes & Steriade (2004: 7) note, speakers’ phonetic knowledge and grammatical constraints are not directly accessible to a student of phonology or phonetics and thus have to be inferred. Facts of phonetic difficulty and sound patterns, however, can be determined through phonetic experiments and typological research respectively. The relative accessibility of these two components makes the hypothesis empirically testable. The goal of the current study is to verify predictions made by the PKG approach about certain aspects of a common phonological process – place assimilation in consonant clusters. Specifically, we will examine the relation between facts of phonetic difficulty in perception of syllable-final stops and common sound patterns of place assimilation. Facts of phonetic difficulty will be investigated in cross-linguistic perceptual experiments.

2 Place assimilation in stop clusters

Some aspects of phonetic grounding of major place assimilation patterns have been previously examined Jun (1995, 2004; cf. Kohler 1990, Byrd 1992). Jun’s cross-linguistic survey of place assimilation reveals some interesting characteristics of the process, as shown in (2).

(2) \textit{Phonological asymmetries in place assimilation} (Jun 1995)

\begin{itemize}
  \item \textit{a.} position asymmetry preconsonantal (coda) > prevocalic (onset) \\
       (> = more likely target)
  \item \textit{b.} target place asymmetry coronal > labial > dorsal \\
       (> = more likely target)
  \item \textit{c.} trigger place asymmetry labial/dorsal > coronal \\
       (> = more likely trigger)
\end{itemize}

Regressive place assimilation is far more common across languages than progressive place assimilation (the position asymmetry; (2a)). Coronal consonants appear to be the most likely targets of assimilation, followed by labials, and then by dorsals (the target place asymmetry; (2b)).\footnote{See also Paradis & Prunet (1991) on the ‘special status’ of coronals, and Prince & Smolensky (1993), Rice (1996, 1999), de Lacy (2002), Hume (2003) and Blevins (2004: 125–129) on relative place markedness and various theoretical approaches to it.} Non-coronals are more likely triggers of place assimilation than coronals (the trigger place asymmetry; (2c)). Other asymmetries identified by Jun include target and trigger manner asymmetries.
An often cited example of the asymmetries in (2) is Korean (casual speech) place assimilation. Place assimilation in Korean is always regressive. Coronals assimilate to following dorsals and labials (3a); labials assimilate to following dorsals, but not to coronals (3b), and dorsals assimilate neither to labials nor to coronals (3c).


<table>
<thead>
<tr>
<th>input</th>
<th>output</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>/mit+ko/</td>
<td>[mikko]</td>
<td>‘believe and’</td>
</tr>
<tr>
<td>/kot+paloo/</td>
<td>[koppaloo]</td>
<td>‘straight’</td>
</tr>
<tr>
<td>/ip+ko/</td>
<td>[ikko]</td>
<td>‘wear and’</td>
</tr>
<tr>
<td>/ipta/</td>
<td>*[ipta]</td>
<td>‘wear + SENTENCE ENDER’</td>
</tr>
<tr>
<td>/ik+ta/</td>
<td>*[ikta]</td>
<td>‘ripe + SENTENCE ENDER’</td>
</tr>
<tr>
<td>/tjak+pʰa/</td>
<td>*[tjakpʰa]</td>
<td>‘destruction’</td>
</tr>
</tbody>
</table>

In the PKG approach, cross-linguistically common asymmetries found in patterns of assimilation indirectly reflect certain universal facts of phonetic difficulty. Thus, preconsonantal consonants should be more difficult to perceive or to produce (or both) than prevocalic consonants. Coronals should be more difficult to perceive/produce than labials and dorsals, while labials should be more difficult than dorsals. Stops followed by non-coronals should be more difficult than stops followed by coronals.

Jun’s (1995, 2004) survey of previous acoustic studies (of mainly English consonants) suggests that such phonetic differences do exist. Jun posits the scales of perceptual salience (‘acoustic strength’) in (4). (Note that perceptual salience is the reverse of perceptual difficulty.)

(4) Scales of perceptual salience (Jun 1995: 125–130)

<table>
<thead>
<tr>
<th>position</th>
<th>prevocalic (onset) &gt; preconsonantal (coda)</th>
</tr>
</thead>
<tbody>
<tr>
<td>target place</td>
<td>dorsal &gt; labial &gt; coronal</td>
</tr>
<tr>
<td>trigger place</td>
<td>labial/dorsal &gt; coronal</td>
</tr>
</tbody>
</table>

First, prevocalic consonants are perceptually more salient than preconsonantal consonants (4a), since the former are cued better (by ‘stronger’ CV transitions and bursts) than the latter (by ‘weaker’ VC transitions and optionally by bursts) (Malécot 1958, Bladon 1986, Ohala 1990, among others). Second, preconsonantal dorsals are the most salient, followed by labials and then coronals. The difference between coronals and non-coronals is due to ‘weaker’ coronal VC transitions (as a consequence of more rapid movement of the tongue tip; Byrd 1992). The difference between dorsals and non-dorsals is due to a salient convergence of the second (F2) and third (F3) formants during the VC transition (Stevens 1989). Third, stops followed by coronals are more salient than stops followed by
non-coronals (4c), an apparent effect of greater masking of C₁ cues by following non-coronals (Byrd 1992, Zsiga 1994).

Jun (1995) notes that the scales of perceptual salience in (4) refer to unreleased syllable-final stops, a point to which we will return later in this paper. In many languages, however, syllable-final stops can be audibly released – accompanied by a burst with possible additional noise (cf. optional coda stop releases in English: Henderson & Repp 1982, Lisker 1999, Zsiga 2000). Releases are known to provide additional information about place of articulation (e.g. Malécot 1958, Winitz et al. 1972, Wright 2004: 38), and therefore released stops are likely to be more perceptually salient than unreleased stops. This gives us another scale of perceptual salience: released stops > unreleased stops. Jun (1995) does not comment on the relative place perceptibility of released stops. We will therefore make a preliminary assumption (to be revised later) that the relative place perceptibility of released stops is similar to that for unreleased stops (4b).

The PKG approach views the perceptibility scales in (4) as part of speakers’ implicit phonetic knowledge. Importantly, this knowledge is shared by speakers of different languages, giving the assumption that basic conditions under which speakers produce and perceive stops are the same. The shared phonetic knowledge of the perceptibility scales leads speakers of different languages to posit similar rankings of constraints in language-specific phonological grammars. Jun (1995, 2004) proposes the fixed hierarchies of different Preserve faithfulness constraints in (5). These hierarchies reflect an assumption that speakers make use of their phonetic knowledge: they make more effort to ‘preserve’ sounds that are more perceptually salient and relax their articulation of sounds that are less perceptually salient (the Production Hypothesis: Jun 1995: 122; cf. Kohler 1990, Byrd 1992, Hura et al. 1992, Lindblom et al. 1995).

(5) HIERARCHIES OF GRAMMATICAL CONSTRAINTS (JUN 1995: 136)

a. \text{Pres}(\text{pl(dom)}) \gg \text{Pres}(\text{pl(lab)}) \gg \text{Pres}(\text{pl(cor)})

b. \text{Pres}(\text{pl(\_V)}) \gg \text{Pres}(\text{pl(\_ C)})

c. \text{Pres}(\text{pl(\_ cor)}) \gg \text{Pres}(\text{pl(\_ lab/dor)})

Different rankings of the markedness constraint WEAKENING (‘conserve articulatory effort’) against the hierarchies in (5) are capable of generating a range of attested patterns of place assimilation (Jun 1995, 2004). For example, the pattern attested in Korean is captured by the ranking of constraints shown in (6).

(6) A FRAGMENT OF KOREAN GRAMMAR (JUN 1995: 146)

\text{Pres}(\text{pl(\_V)}), \text{Pres}(\text{pl(dom)}), \text{Pres}(\text{pl(\_ cor)}) \gg
\text{WEAKENING} \gg
\text{Pres}(\text{pl(lab)}), \text{Pres}(\text{pl(cor)}), \text{Pres}(\text{pl(\_ lab/dor)}), \text{Pres}(\text{pl(\_ C)})

Ranking Pres(pl(\_V) above Pres (pl(\_ C) ensures the regressive direction of place assimilation. The higher ranking of Pres(pl(dom)) and
PRES(pl(_cor)) ensures that assimilation does not affect syllable-final dorsals and consonants followed by coronals (e.g. /kp/, /kt/ and /pt/), while the lower ranking of the other place-referring constraints enforces assimilation in the other consonant clusters (e.g. /pk/, /tp/ and /tk/).

Jun’s phonetically based account of place assimilation undoubtedly contributes to a better understanding of phonetic grounding of phonological processes. One limitation of this account, however, is that the universal scales of perceptual salience in (4) are mainly deduced from previous acoustic studies, rather than being established in perceptual experiments (cf. Hume et al. 1999, Wright 2001, Winters 2003). It is well known that listeners attend to certain acoustic properties more than others (Bregman 1990); therefore, some acoustic differences may or may not be relevant perceptually. Accordingly, statements about the universal salience of positions and places may not be fully valid, unless verified in cross-linguistic perceptual experiments.

While a number of perceptual studies have confirmed the position perceptibility scale (4a) (Fujimura et al. 1978, Ohala 1990, Wright 2001, Kingston & Shinya 2003), relatively little perceptual work has been done to directly verify the place salience scales and their universality, as also noted by Hume et al. (1999) and Winters (2003). The current study is intended to contribute to the investigation of the target place salience scale (4b), its universal status and its possible relation to grammatical constraints and patterns of place assimilation. This is done by conducting a perceptual experiment where Russian syllable-final stops /p t k/ are presented to listeners, who are native speakers of four languages: Russian, Canadian English, Korean and Taiwanese Mandarin. The PKG approach is examined by testing the specific hypotheses about phonetic differences in stop perception in (7).

(7) a. **Place target salience hypothesis**
   
   Syllable-final dorsals are more perceptually salient than labials and coronals; labials are more salient than coronals
   
   (dorsal > labial > coronal).

b. **Release salience hypothesis**

   Syllable-final released stops are more perceptually salient than unreleased stops (released > unreleased).

---

2 Jun (2004) cites one perceptual study (Winitz et al. 1972) as supporting his place target salience scale in (4b). The study, however, does not investigate the perceptibility of unreleased stops. Rather, it examines identification of place in voiceless English stops based in (i) burst only and (ii) burst and transitions (100 ms of the adjacent vowel). The authors note that ‘final /k/ depends strongly for its perception on transitional segments’ (1972: 1313), given a considerably better identification of the segment in the second condition. The relative place perceptibility of final stops in two conditions, (i) /t/ > /p/ > /k/, (ii) /p/ > /k/ > /t/ (1972: 1311–1312), however, does not fully agree with Jun’s place salience scale.
c. *Universal salience hypothesis*

The relative salience of places of articulation and releases is language-independent.

Note that the universal salience hypothesis emphasises cross-linguistic similarity in relative salience. Given language-particular phonological and phonetic differences involving stop place of articulation, some absolute perceptual differences among languages are not unexpected (Hume *et al.* 1999, among others).

While our primary focus is on the place target salience scale, the results of the experiments will also allow us to evaluate Jun’s predictions about the place trigger scale (labial/dorsal > coronal). Further, the results are expected to provide evidence for or against two alternative views of assimilatory sound changes. Previous studies have produced conflicting results as to what types of errors in place identification are more common (Hura *et al.* 1992, Kingston & Shinya 2003). Prevalence of non-assimilatory errors has been interpreted as evidence for the role of phonetic knowledge (Hura *et al.* 1992), and therefore as evidence for the PKG approach (Steriade 2001, Hayes & Steriade 2004). That is, if perceptual errors are predominantly non-assimilatory, they cannot, on their own, trigger an assimilatory sound change. Therefore, there must be another mechanism that converts these misperceptions to an assimilatory sound change. This mechanism is presumably speakers’ knowledge of relative perceptibility, of the type assumed by the production hypothesis (Kohler 1990, among others). In contrast, if perceptual errors are predominantly assimilatory, an assimilatory sound change can arise as grammaticalisation of these errors, without reference to speakers’ phonetic knowledge. This will support the ‘innocent misapprehension’ view of assimilation (Ohala 1981, 1990). Finally, the results of our experiments will allow us to determine whether the perceptibility scales deduced from the perceptual experiments do correspond to the constraint hierarchies and patterns of place assimilation discussed in Jun (1995, 2004).

3 **Experiment 1**

The primary goal of Experiment 1 is to test the hypotheses with respect to place target salience, release salience and universal salience in (7), and to examine patterns of perceptual errors as possibly indicative of the role of phonetic knowledge in place assimilation.

3.1 **Method**

3.1.1 *Materials.* The experiment was based on Russian speech stimuli, carefully controlled for the lack of categorical or gradient place assimilation. The stimuli included sequences of the type [taC₁#C₂ap], as shown
in (8), excised from nonsense utterances from a Russian corpus collected using a magnetic articulometer (EMMA: Perkell et al. 1992).

(8) Target stop

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>labial</td>
<td>tap, tap</td>
</tr>
<tr>
<td>coronal</td>
<td>tat, tap</td>
</tr>
<tr>
<td>dorsal</td>
<td>tak, tap</td>
</tr>
</tbody>
</table>

The utterances were produced by two female speakers of Standard Russian (one from Moscow and one from Perm). In each utterance, the target stop C₁ (syllable-final /p t k/) was followed by a heterorganic stop C₂ (syllable-initial /p t k/). Utterances where the target consonants were followed by homorganic stops – [tap pap], [tat tap] and [tak kap] – were also included. These were used in the articulatory analysis as controls for gestural reduction (see below). In the perceptual experiment, stimuli with homorganic sequences were used as fillers, to ensure that the stimulus set was balanced, with all possible combinations of C₁ and C₂. In all utterances, the vowel context was the same, the low central vowel /a/. The speakers were instructed to stress both syllables. All utterances were read five times in the carrier phrase ['Et V_j V`p]at`]. Two tokens of each utterance (the third and fourth repetitions) were selected for the experiment, giving a total of 24 stimuli with heterorganic sequences (6 stimuli X 2 tokens X 2 speakers) and 12 stimuli with homorganic sequences (3 stimuli X 2 tokens X 2 speakers).

The EMMA data was used in the experiment in order to ensure that all syllable-final oral gestures of the stop consonants were fully produced. A number of studies have shown that articulatory gestures tend to reduce spatially in syllable- and word-final positions (Browman & Goldstein 1995, Krakow 1999). Such reduction can be either gradient or categorical (the latter corresponding to deletion or assimilation), and there are considerable differences in reduction patterns (Ellis & Hardcastle 2002 on English; Son et al. 2007 on Korean). Gradient and especially categorical reduction is known to have considerable perceptual consequences (Nolan 1992, Jun 1996, Surprenant & Goldstein 1998, Son et al. 2007). Therefore, a lack of a proper control over gestural reduction in a study investigating perceptual sources of place assimilation would not be desirable.

An articulatory analysis of the current stimuli showed that all syllable-final stops were fully produced – they exhibited neither gradient nor categorical reduction in the magnitude of the relevant gestures. Some sample tokens used in the current study are shown in Fig. 1, which displays the spatial positions of EMMA receivers indicating the positions

---

3 The use of EMMA data in perceptual research is not uncommon, particularly in studies investigating the link between production and perception, for example, Surprenant & Goldstein (1998), Ellis & Hardcastle (2002), Pouplier & Goldstein (2005) and Perkell et al. (2006).
of the following articulators (from left to right): the upper and lower lips, the front of the lower teeth and the tongue, approximately from the tip to the dorsum. Positions of receivers during the production of stops as $C_1$ in heterorganic sequences (left) are compared to positions of the same stops as $C_1$ in homorganic sequences (right). (The snapshots were taken at the midpoint of the $C_1$ constriction in heterorganic sequences and at 25% of the $C_1C_2$ constriction in homorganic sequences.) The relevant articulators are circled. Notice that the lip constriction during [p] in [tap kap] has similar spatial values to that during [p] in [tap pap] (top row). The tongue-tip constriction during [t] in [tat pap] is similar to that during [t] in [tat tap] (middle row) (the tongue shape is somewhat different). The

**Figure 1**

Positions of EMMA receivers, indicating relative positions of articulators during the production of [p], [t] and [k] in heterorganic (left) and homorganic (right) sequences (sample tokens).
tongue-dorsum constriction during [k] in [tak tap] is similar to that during
[k] in [tak kap] (bottom row).

Overall, the oral articulatory gestures of C₁ and C₂ in the stimuli ex-
hibited relatively little overlap. Figure 2 shows a sample token of [tak tap].
Notice that the tongue-dorsum constriction for [k] in (c)) is released about
60 ms prior to the formation of the tongue-tip constriction for [t] (b). The
delay in achieving the C₂ target results in a fairly long audible release of C₁
prior to the closure of C₂ in the acoustic signal (a).

An acoustic analysis of the stimuli showed that audible C₁ releases as in
Fig. 2 were found in most of the heterorganic tokens used in the exper-
iment (as determined on the basis of waveforms and spectrograms). There
were four tokens without clear releases (two tokens of [tap tap] and one
each of [tap kap] and [tat pap]). In the homorganic sequences, C₁ stops
were always unreleased. Releases of [t] and [k] were longer in duration
than releases of [p] (on average 36 ms and 27 ms for [t] and [k] vs. 18 ms
for [p]). Labial releases were also lower in amplitude than coronal and

---

4 Spatial values for the stimuli were within two standard deviations (on average
1 mm) from the respective control values, which were based on means for five
tokens of each homorganic sequence.

5 Similar findings for stop releases in Russian are reported by Zsiga (2000) in an
acoustic study of word-boundary stop–stop sequences.
dorsal releases. Overall, [t], [p] and [k]-initial sequences had approximately the same closure duration: 156 ms, 160 ms and 158 ms respectively (measured from the offset of the preceding vowel to the release of C₂). F₂ transitions towards the [p] closure involved lowering (to 1220 Hz, averaged for both speakers), while they stayed level or rose slightly before the [t] and [k] closures (to 1516 Hz and 1562 Hz respectively). F₃ values prior to the closure of [k] were lower than those prior to the closures of [p] and [t] (2702 Hz vs. 2938 Hz and 3029 Hz). Note that these differences in VC transitions are overall similar to those reported for English stops by Stevens (1998).

The stimuli were used in two conditions: ‘complete’ and ‘edited’. Stimuli for the ‘complete’ condition were the same as described above. In these stimuli, the place of C₁ was cued by both VC transitions and the release of C₁ (if present). Stimuli for the ‘edited’ condition were created from the ‘complete’ stimuli by removing a C₁ release (if present) from the acoustic signal by setting the amplitude during the C₁C₂ closure to zero (cf. Wright 2001, Winters 2003). The beginning and end points of the edited interval were made at audio waveform zero-crossings in order to avoid introducing any transients. The intensity levels of sound files were adjusted in order to maintain the same average intensity (at the midpoint of the preceding vowel). This procedure left the ‘edited’ stimuli with VC transitions as the only cue to the place of C₁, eliminating the other potential sources of place information – stop releases, and any noise occurring during stop closures. In both ‘complete’ and ‘edited’ files, 50 ms of silence were appended before and after the signal. The inclusion of the ‘edited’ set of stimuli was necessary to test the ‘net’ contribution of VC transitions to place salience. Recall that the originally proposed place salience scale was based on the acoustic properties of vowel transitions to unreleased stops (Jun 1995).

The analysis of perceptual results was based on the heterorganic stimuli, which were most relevant to an investigation of assimilation. However, means obtained from the homorganic stimuli were used to provide a comparison with those from the heterorganic condition.

3.1.2 Participants. The participants in Experiment 1 were 56 undergraduate students at Simon Fraser University (Vancouver), who were native speakers of Russian, Canadian English, Korean or Taiwanese Mandarin (14 listeners per language group). The Russian listeners (8 females, 6 males) were born in Russia or the former Soviet Union, and had arrived in Canada at an average age of 18. The Canadian English listeners

---

6 Based on figures of VC transitions to stops in the context of [a] produced by a male English speaker (Stevens 1998: 343, 357, 367), F₂ values descend to approximately 1000 Hz before [b], and ascend to approximately 1550 Hz and 1400 Hz before [d] and [g] respectively. F₃ values are lower before [g] (approximately 2450 Hz) than before [b] and [d] (approximately 2600 Hz for both).

7 Recall that releases were absent in four heterorganic tokens and all homorganic tokens.
(9 females, 5 males) were born in British Columbia, Canada. The Korean listeners (7 females, 7 males) were born in South Korea (mostly in Seoul), and had arrived in Canada at an average age of 16. The Taiwanese Mandarin listeners (12 females, 2 males) were born in Taiwan (mostly in Taipei), and had arrived in Canada at an average age of 13. The average age of the participants was 23. All non-English native speakers were English bilinguals. None of the listeners reported hearing problems; all were naive as to the purpose of the experiment.

The reason for employing listeners/native speakers of these languages was to provide some control for differences in language-particular phonological or phonetic knowledge. Russian is a ‘non-assimilating’ language, while English is known to exhibit coronal place assimilation, which is mainly gradient (Browman & Goldstein 1992, Nolan 1992, Ellis & Hardcastle 2002). Korean is a language exhibiting coronal and labial place assimilation (Kim-Renuad 1991). (Both English and Korean place assimilation processes tend to occur in casual speech.) In Standard Mandarin, on the other hand, the place contrast in stops is absent syllable-finally, since stops are not permitted in codas in general (Norman 1988, Duanmu 2002).

Assuming Jun’s (1996, 2004) account of place assimilation, the grammars of the four languages differ in the relative ranking of the WEAKENING constraint against the fixed hierarchy of PRESERVE constraints referring to place features, as shown in (9). 8

(9) Assumed rankings of the WEAKENING constraint in the grammars of Mandarin, Korean, English and Russian (stops)

\[
\begin{array}{c|c|c|c|c}
& \text{Mandarin} & \text{Korean} & \text{English} & \text{Russian} \\
\hline
\text{WEAKENING} & \text{WEAKENING} & \text{WEAKENING} & \text{WEAKENING} \\
\end{array}
\]

Some interlanguage differences in the phonetic implementation of syllable-final stops are also worth noting. While stops in (heterorganic) sequences in Russian are usually released, they are optionally released in English (Henderson & Repp 1982, Lisker 1999, Zsiga 2000) and are apparently hardly ever released in Korean (Kim-Renuad 1991). Overall, significant similarities between the three groups in their perceptual performance can be attributed to factors independent of listeners’ native language backgrounds, such as general facts of perceptual difficulty (and possibly English bilingualism, a factor discussed below). Differences

---

8 The rankings for English and Korean are characteristic of the casual speech register. For all languages, the constraint $\text{Pres(pl(V))}$ is ranked above the other faithfulness constraints in (9), while $\text{Pres(pl(C))}$ is ranked below them. See Jun (2004) on a possible OT account of gradient place assimilation, such as coronal place assimilation in English. It should be noted that the ranking for Mandarin may not fully reflect the knowledge of some of our Mandarin listeners, who were familiar with Taiwanese (a Min dialect of Chinese) (as discussed in §3.2.1).
between the four groups may be explained by listeners’ language-particular phonological/phonetic knowledge.

3.1.3 Data collection and analysis. The experiment used a forced-choice phoneme-identification task. The stimuli were presented individually; the participants were instructed to listen to speech tokens and to identify the last consonant of the first syllable (e.g. [tap tap]) by pressing buttons labelled ‘p’, ‘t’ and ‘k’ on a response box. The participants were asked to respond as quickly as possible and to guess when they were in doubt. Instructions were given in English. All stimuli were presented in random order with a four-second interstimulus interval. Each token was presented twice. Trials were blocked by condition (‘complete’ and ‘edited’); these blocks were presented in alternating order in a single session.

A total of 5376 responses to stimuli with heterorganic sequences were collected (6 utterances × 2 speakers × 2 tokens × 2 repetitions × 2 conditions × 56 listeners). Data were averaged across 8 responses to the same stimulus type (2 speakers × 2 tokens × 2 repetitions) for each listener. Two measures were used in the analysis of data: percentage of correct responses (correct identification responses divided by all responses) and the reaction time for correct responses (RT), which was measured from the onset of the sound file. Higher percentages of correct and faster RT values were taken to indicate greater perceptual salience of place of articulation.

In addition to percent correct and RT data, the types of listeners’ errors (incorrect responses) were examined. In particular, it was of interest to determine whether the proportions of ‘assimilatory’ errors (C₁ perceived as having the same place as C₂) were similar to or higher than the proportions of ‘non-assimilatory’ errors (C₁ perceived as having a different place from C₂).

3.2 Results and discussion

The results of the experiment are summarised in Table I as a confusion matrix of all identification responses of C₁, combined for four language groups (a total of 56 participants). The results are shown by the condition (‘complete’ and ‘edited’), the target consonant (C₁: /p/, /t/ or /k/) and the context (C₂: /p/, /t/ or /k/). Mean percentages of correct and RT values are provided in the rightmost columns. Given the substantial qualitative and quantitative differences between the two conditions, we will examine the percentage of correct and RT results separately for the ‘complete’ and the ‘edited’ stops (§§3.2.1 and 3.2.2). In addition, we will examine types of errors in both conditions (§3.2.3).

Due to a programming error, fewer tokens of the released condition stimulus [tap tap] were presented. This resulted in 350 responses for this particular stimulus, compared to 450 responses for the other stimuli.
3.2.1 Percent correct and reaction time: the ‘complete’ condition. Two individual mixed-design ANOVAs were carried out to examine the effects of $C_1$ Place (/p t k/) and Language Group (Russian, Canadian English, Korean, Taiwanese Mandarin) on listeners’ mean percent correct and RT values respectively. With respect to percent correct, there were significant effects of $C_1$ Place and Language Group [$F(2,104)=30.401$, $p<0.001$ and $F(3,52)=5.434$, $p<0.01$ respectively]. The interaction of $C_1$ Place and Language Group was not significant. Post hoc LSD tests for $C_1$ Place revealed that /k/ was significantly different from /p/ and /t/ ($p<0.001$ for both) and that /p/ was significantly different from /t/ ($p<0.05$). Post hoc LSD tests for Language Group indicated that Russian listeners showed results significantly different from Korean and Mandarin listeners ($p<0.01$ for both). Similarly, English listeners performed differently from Korean listeners ($p<0.01$). The difference between English and Mandarin listeners was marginally significant ($p=0.061$). With respect to RT (for correct responses), there was a significant effect of $C_1$
Place [F(2,104) = 7.521, p < 0.001]. Post hoc LSD tests for C₁ Place revealed that /k/ was different from /p/ and /t/ (p < 0.001 for both). The effect of Language Group was marginally significant [F(3,52) = 2.526, p = 0.068]. The interaction of the two factors was not significant.

Figure 3 presents mean percent correct (a) and RT (b) results for the ‘complete’ stimuli according to place of articulation of C₁ (/p t k/) and the language group (Russian, Canadian English, Korean, Taiwanese Mandarin). We can see that the /k/ has the highest mean percentage of correct values, followed by /p/, and then /t/. This pattern is quite consistent

(a) Mean percent correct and (b) mean reaction time values by language group for ‘complete’ syllable-final /p t k/ (C₁).

A re-examination of individual RT values for English listeners revealed that three out of fourteen listeners showed relatively slow RT values for /t/ (1700–2100 ms, compared to RTs lower than 1400 ms for the other English listeners). Excluding the RT data for these three listeners from an ANOVA analysis produced significant effects for both Place and Language Group [F(2,98) = 5.332, p < 0.001 and F(3,49) = 2.877, p < 0.05 respectively]. Post hoc LSD tests for Place revealed that /k/ was different from /p/ and /t/ (p < 0.001 for both). Post hoc LSD tests for Language Group revealed that the RT for Russian listeners was significantly different from the RT for Korean and Mandarin listeners (p ≤ 0.05 for both). Similarly, the RT for English listeners was significantly different from that for Korean listeners. The difference between English and Mandarin listener groups was marginally significant (p = 0.076). The interaction of Place and Language Group was not significant.
across the language groups. Russian and English listeners show higher mean percent correct values than Korean and Mandarin listeners. In terms of RT, /k/ shows shorter mean RT than the other stops. Russian and English listeners show a tendency for shorter RT than Korean and Mandarin listeners.

The identification of non-coronals was not necessarily better or worse when they were followed by a coronal. An examination of Table I reveals that the identification of /p/ was only 3% better before /t/ than before /k/, and that of /k/ was 2% worse before /t/ than before /p/. RT differences were also minor, but showed somewhat longer values when C₂ was a coronal than a non-coronal.

Overall, the results fully support the place target salience hypothesis in (7a): /k/ is more perceptually salient than /p/ and /t/ (based on both percent correct and RT); /p/ is more salient than /t/ (based on percent correct). The results also largely support the universal salience hypothesis in (7c): relative salience of places of articulation was the same for all four language groups (based on percent correct and in part on RT). The differences between the language groups in the magnitude of salience were in the expected direction: listeners whose languages exhibit no or limited place assimilation (Russian and English) performed better than listeners whose languages exhibit more extensive assimilation (Korean) or a lack of place contrast in coda (Mandarin). However, this difference in performance can be also attributed to language-particular phonetic knowledge. In the current condition, C₁ stops were predominantly released (see §3.1.1). Russian and English listeners are accustomed to released syllable-final stops in their native languages (although to differing extents), while Korean and Mandarin listeners are not used to such sounds in their native languages. The familiarity of the Korean and Mandarin participants with English released stops may have been a factor, yet apparently not a strong enough one to override their native perceptual biases. The lack of significant differences between the Korean and Mandarin groups is less expected, given the phonotactic differences between the two languages (see §3.1.2; cf. Flege & Wang 1989). The lack of the effect can be attributed to two factors. First, Mandarin listeners could have had greater exposure to English than Korean listeners (judging by their average age of arrival in Canada). Second, as mentioned earlier, our Mandarin listeners were originally from Taiwan, and some of them were familiar with Taiwanese (a Min dialect of Chinese), which allows (unreleased) stops in word-final position (Norman 1988).

Finally, the results did not confirm the trigger place scale in (4c). However, the lack of this effect can be due to the relatively small overlap in the Russian stop clusters in the stimuli (see §3.1.1). The substantial masking effect of following non-coronals was found for English, where stop gestures show considerable overlap (Zsiga 1994, Byrd 1996).

3.2.2 Percent correct and reaction time: the ‘edited’ condition. Two mixed-design ANOVAs were carried out to examine the effects of C₁ Place
and Language Group on listeners’ mean percent correct and RT respectively. With respect to percent correct, only the effect of C₁ Place was significant [F(2,104) = 27.493, p < 0.001]. Post hoc LSD tests further indicated that /p/ was significantly different from /t/ and /k/ (p < 0.001 for both); /t/ was significantly different from /k/ (p < 0.05). With respect to RT, the C₁ Place effect was significant [F(2,98) = 3.585, p < 0.05].¹¹ The effect of Language Group and the interaction of C₁ Place and Language Group were not significant. Post hoc LSD tests revealed that /p/ was significantly different from /t/ and /k/ (p < 0.05 for both).

Figure 4 presents percent correct (a) and RT (b) results for the ‘edited’ stops according to place and language group. The labial /p/ has the highest mean percent correct values, followed by /t/ and then /k/. This pattern is consistent for all four language groups. The language groups show similar overall mean percent correct values. Note that for Mandarin listeners, the mean percent correct value for /k/ is below chance level (32.1% compared to chance level 33.3%). In terms of RT, /p/ was identified faster than /t/ and /k/. Again, overall mean RT values are similar across the groups.

Consonants followed by coronals were not identified any better than consonants followed by non-coronals. Thus correct identification of /p/ and /k/ was on average 52% when followed by /t/ and 55% when followed by a non-coronal (/k/ and /p/ respectively) (see Table I). RT was slower before coronals (1532 ms) than before non-coronals (1407 ms).

Comparing the current results to those obtained for the ‘complete’ condition, it is obvious that the ‘edited’ stops were overall identified considerably less accurately and more slowly. Interestingly, this effect was accompanied by place and language group differences. With respect to place, the decrease in percent correct values in the ‘edited’ condition was almost double for /k/ (on average 48%), while it was moderate to small for /t/ and /p/ (17% and 9% respectively). The increase in RT values was the greatest for /k/ (on average 289 ms), followed by /t/ (145 ms) and then /p/ (74 ms). With respect to the language groups, the overall decrease in percent correct values was the greatest for Russian listeners (32%), followed by English listeners (26%) and then Korean and Mandarin listeners (20% and 19% respectively). The increase in RT values was the greatest for Russian and English listeners (240 ms and 232 ms respectively), followed by Mandarin listeners (143 ms) and then Korean listeners (63 ms).

Recall that the ‘edited’ stimuli were employed in order to determine the ‘net’ contribution of VC transitions on the place perceptibility. Thus the VC transition results only partially support the place target salience hypothesis (7a). Labial /p/ is more salient than coronal /t/ (based on percent correct and RT). Dorsal /k/, however, is perceptually less salient than /p/ (based on percent correct and RT). Moreover, /k/ is perceptually less salient than /t/ (based on percent correct). The perceptibility scale

¹¹ Reaction time results for the unreleased condition are based on data from 53 of the 56 participants (due to no responses to some of the stimuli from three participants).
obtained – labial > coronal > dorsal – is at odds with the scale proposed by Jun (1996, 2004) in (4a). The result is particularly surprising, since the scale was originally proposed for unreleased stops – presumably based on language-independent, acoustic properties of their VC transitions. Clearly, the listeners in the current study did not take full advantage of the place information provided by the convergence of F2 and F3 during the VC transition to /k/. The poor identification of /k/ is also surprising, given the finding of the articulatory analysis that the tongue-dorsum raising gestures in the stimuli had a considerable extent in time and were not spatially reduced (see Figs 1 and 2). It may be the case that unreleased /k/ is inherently less salient in this particular vowel context (after the low central /a/), and the results may be different in other vowel contexts.\footnote{In a front vowel context, F2 before a velar stop is considerably higher, resulting in a clearer convergence of F2 and F3 (Stevens 1998: 367). This may suggest a somewhat better identification of velar stops in front vowel contexts.}

This question obviously requires more study. It should be noted, however, that similarly poor identification of unreleased dorsals has been found in perceptual studies investigating final stops in a variety of vowel contexts in a number of languages (English: Householder 1956, Malécot 1958, Lisker 1999, Wright 2001, Winters 2003; Cantonese: Gao 2005; Dutch: Winters 2003; Korean: Oh 2002; Thai: Abramson & Tingsabadh 1999).
The results strongly support the release salience hypothesis in (7b): released stops are considerably more perceptually salient than unreleased stops (based on percent correct and RT). This confirms some previous findings (Malécot 1958, among others). Interestingly, however, the results show that releases are most important for dorsals, followed by coronals and then labials. This does seem to correlate with the relative acoustic salience of stop releases. Recall that in the stimuli, both dorsals and coronals had louder and longer releases than labials (see §3.1.1; cf. Cho & Ladefoged 1999 on VOT differences among places). In the absence of such releases, the lingual stops—coronals, and especially dorsals—seem more perceptually vulnerable than labials. The finding that the scales of perceptual salience are not necessarily the same for released and unreleased stops was also obtained by Winters (2003).

It should be noted that releases may not be the only source of the obtained differences between the two conditions. Recall that the editing procedure employed in the experiment removed all acoustic noise during the C_1C_2 interval (see §3.1.3). Some of this noise, resulting from incomplete closure, may provide additional place information (Repp 1983a, Chitoran et al. 2002). The use of homorganic stimuli, where C_1 stops were naturally unreleased, allowed us to determine whether the closure noise was perceptually relevant here. In this case, the difference between the ‘complete’ and ‘edited’ conditions is the presence or absence of any noise during the closure. The comparison of the means for two sets revealed that the ‘edited’ homorganic sequences had a somewhat lower percent correct and slower reaction time than the ‘complete’ homorganic sequences (61% vs. 68% and 1490 ms vs. 1401 ms). This indicates a relatively small effect of closure noise in the results (comparing the magnitude of differences between the conditions in heterorganic sequences; Table I). The difference, however, can be also attributed to the listeners’ overall lower confidence in identifying less informative stimuli in the ‘edited’ condition. In general, the relative place salience for homorganic stops in both conditions (based on percent correct) was in the same direction as for heterorganic stops in the ‘edited’ condition—labial > coronal > dorsal. This shows that the VC transitions of naturally and artificially unreleased stops in the stimuli were not substantially different (cf. Lisker 1999 on English stops).

The results largely support the universal salience hypothesis in (7c). The relative salience of places of articulation was the same for all four language groups, based on percent correct and in part on RT. The lack of significant differences between the language groups in the identification of unreleased stops was due to the considerably lower performance of Russian and English listeners. This suggests that familiarity with released stops in native languages was the main factor that contributed to the language group differences in the ‘complete’ condition (see §3.2.1).

The means for ‘complete’ homorganic sequences were 78% for /p/, 67% for /t/ and 62% for /k/; the means for ‘edited’ homorganic sequences were 66% for /p/, 61% for /t/ and 54% for /k/.

---

13 The means for ‘complete’ homorganic sequences were 78% for /p/, 67% for /t/ and 62% for /k/; the means for ‘edited’ homorganic sequences were 66% for /p/, 61% for /t/ and 54% for /k/.
3.2.3 Types of errors: the ‘complete’ and ‘edited’ conditions. Figure 5 presents proportions of assimilatory errors in both conditions by language group (a) and place of articulation (b). Recall that in assimilatory errors, $C_1$ is perceived as being homorganic with $C_2$ (e.g. /p#t/ → /t#t/); in non-assimilatory errors $C_1$ and $C_2$ are perceived as being heterorganic (e.g. /p#t/ → /k#t/). The proportions of assimilatory errors here were based on the total numbers of errors. Values higher than 0.5 indicate that there were more assimilatory than non-assimilatory errors.

As seen in Fig. 5a, assimilatory errors substantially outnumbered non-assimilatory errors across all language groups. Proportions, however, were greater in the ‘complete’ condition than in the ‘edited’ condition: on average 72% vs. 59%. As seen in Fig. 5b, all three places showed a predominance of assimilatory errors, with a rather higher value for the coronal in the complete condition. The lower proportion of assimilatory errors for unreleased /k/ was due to its frequent confusion with /t/, apparently independently of the place of the following consonant (see Table I).

Considering all errors (assimilatory and non-assimilatory), the ‘k’ response was overall more frequent than the ‘p’ and ‘t’ responses in the ‘complete’ condition (on average 41% vs. 31% and 28% of all errors).
This was found for all the language groups except Korean. In the ‘edited’ condition, the ‘k’ response was consistently disfavoured across the groups, in comparison with the ‘p’ and ‘t’ responses (23% vs. 41% and 36%). This seems to reflect a higher number of assimilatory errors before /k/ in the one condition, and a higher number of confusions of the edited /k/ with either /p/ or /t/ in the other condition (see Table I).

In sum, the analysis of errors shows that there is a substantial assimilatory bias in both conditions. Strikingly, although the overall number of errors is greater when stops are unreleased (compared to the released stops in the ‘complete’ condition), the proportion of assimilatory errors for the unreleased condition is lower. This suggests that there are several acoustic sources of assimilatory bias. The first source is stop releases. Releases are likely to contain not only the information about C₁ but also, due to co-articulation, some information about C₂. Recall that the releases in the stimuli were fairly long (on average 28 ms); they were longest for /t/ and shortest for /p/ (see §3.1.1). The information about C₂ contained in some cases in the release of C₁ was interpreted by listeners as information about C₁. Another source of the assimilatory bias is possibly related to C₂ proper – its release and CV transitions to the following vowel. CV transitions are known to be perceptually more salient than VC transitions (Fujimura et al. 1978, Repp 1983b, Bladon 1986, Ohala 1990). Yet another possible source of assimilatory errors are the VC transitions, which, due to C₁C₂ overlap, may contain information about both C₁ and C₂ (cf. Byrd 1992, Žsiga 1994, Surprenant & Goldstein 1998, Son et al. 2007). The prevalence of assimilatory errors in the results is in contrast with some previous studies (Hura et al. 1992),¹⁴ but is compatible with others (Kingston & Shinya 2003). The findings showing such an assimilatory bias in place perception somewhat weaken the claim of the PKG approach to the role of speakers’ phonetic knowledge in place assimilation: if listeners’ errors are predominantly assimilatory, the reference to phonetic knowledge of perceptual salience may not be necessary (see §2).

3.3 Conclusion

The results of Experiment 1 fully confirm the place target salience hypothesis (released dorsal > labial > coronal) for the ‘complete’ condition, but only partly confirm it for the ‘edited’ condition (unreleased labial > coronal, not dorsal > labial, coronal). While in the first condition place of articulation of C₁ is primarily cued by both VC transitions and releases, in the second condition it is cued by the VC transitions only.

The results also confirm the release salience hypothesis (released > unreleased) and the universal salience hypothesis. They also show

¹⁴ Blevins (forthcoming) notes that the prevalence of non-assimilatory errors involving syllable-final nasals /m n n/ found by Hura et al. (1992) is probably a lexical frequency effect. Specifically, the high number of ‘n’ responses in the study corresponds to the high frequency of this consonant word-finally in English proper names, similar to the words used in the experiment in Hura et al. (1992).
that some place information was possibly provided by noise during the closure. No support is provided for the place trigger scale (labial/dorsal > coronal). Finally, assimilatory errors are more common than non-assimilatory errors, more so when stops are released.

The results can be attributed to the relative salience of releases and VC transitions for different places of articulation. Releases in our data appear to be most informative for dorsals, followed by coronals and then labials. The VC transitions are most informative for labials, followed by coronals and then dorsals. The high proportions of assimilatory errors in responses in both ‘complete’ and ‘edited’ conditions, however, suggest that the results could have been to some extent affected by the releases and CV transitions of the following stop (C2). In sum, the place target and place release salience scales determined in Experiment 1 appear to reflect contributions from several cues: the VC transitions and releases of C1, as well as the CV transitions and releases of C2. Only by isolating the effects of the VC transitions and releases of C1 can we determine the relative place salience of C1 proper. Excluding the C2 information would also help us determine whether the CV transitions of C2 were partly responsible for assimilatory errors in Experiment 1.

4 Experiment 2

The goal of Experiment 2 is to determine whether the perceptual salience scales established in Experiment 1 reflect C1 salience proper (VC transitions and releases). Another goal is to clarify the factors causing assimilatory errors in the identification of C1.

4.1 Method

The stimuli for the experiment were created from the ‘complete’ and ‘edited’ CVC1#C2VC stimuli used in Experiment 1 by removing the C2VC portions from individual signals, while leaving C1 releases intact (TRUNCATED STIMULI; cf. Byrd 1992, Surprenant & Goldstein 1998). This resulted in six new heterorganic stimuli for each of the two conditions, as shown in (6). The same was done for the homorganic utterances, which were used as fillers (as in Experiment 1). The design, the participants, the task and the analyses were identical to Experiment 1. (Experiment 2 was conducted in the same session, following Experiment 1.)

(10) Target stop

- labial: [tap] from [tap tap], [tap kap]
- coronal: [tat] from [tat pap], [tat kap]
- dorsal: [tak] from [tak pap], [tak tap]
4.2 Results and discussion

The results of the experiment are summarised in Table II as a confusion matrix of all identification responses of $C_1$, combined for four language groups. As in Table I, the results are shown by the condition (‘complete’ and ‘edited’), the target consonant ($C_1$: /p t k/) and the context ($C_2$: /p t k/). Mean percent correct and RT values are provided in the rightmost columns.

As in Experiment 1, percent correct and RT results will be examined separately for the ‘complete’ and the ‘edited’ stops (§§4.2.1 and 4.2.2), followed by an examination of types of errors for both conditions (§4.2.3). In each case, differences across the language groups will be also examined.

4.2.1 Percent correct and reaction time: the ‘complete’ condition. Two individual mixed-design ANOVAs were carried out to examine the effects

<table>
<thead>
<tr>
<th>condition</th>
<th>target ($C_1$)</th>
<th>context ($C_2$)</th>
<th>‘p’</th>
<th>‘t’</th>
<th>‘k’</th>
<th>no response</th>
<th>% correct</th>
<th>reaction time (correct)</th>
</tr>
</thead>
<tbody>
<tr>
<td>complete</td>
<td>p</td>
<td>t</td>
<td>339</td>
<td>70</td>
<td>37</td>
<td>2</td>
<td>75.67</td>
<td>1219</td>
</tr>
<tr>
<td></td>
<td></td>
<td>k</td>
<td>346</td>
<td>77</td>
<td>25</td>
<td>0</td>
<td>77.23</td>
<td>1183</td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>p</td>
<td>131</td>
<td>288</td>
<td>29</td>
<td>0</td>
<td>64.29</td>
<td>1237</td>
</tr>
<tr>
<td></td>
<td></td>
<td>k</td>
<td>53</td>
<td>314</td>
<td>80</td>
<td>0</td>
<td>70.25</td>
<td>1201</td>
</tr>
<tr>
<td></td>
<td>k</td>
<td>p</td>
<td>9</td>
<td>30</td>
<td>409</td>
<td>0</td>
<td>91.29</td>
<td>992</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t</td>
<td>10</td>
<td>25</td>
<td>412</td>
<td>1</td>
<td>91.96</td>
<td>987</td>
</tr>
<tr>
<td>edited</td>
<td>p</td>
<td>t</td>
<td>319</td>
<td>94</td>
<td>34</td>
<td>1</td>
<td>71.21</td>
<td>1239</td>
</tr>
<tr>
<td></td>
<td></td>
<td>k</td>
<td>311</td>
<td>99</td>
<td>37</td>
<td>1</td>
<td>69.42</td>
<td>1182</td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>p</td>
<td>113</td>
<td>284</td>
<td>51</td>
<td>0</td>
<td>63.39</td>
<td>1303</td>
</tr>
<tr>
<td></td>
<td></td>
<td>k</td>
<td>91</td>
<td>317</td>
<td>39</td>
<td>1</td>
<td>70.76</td>
<td>1268</td>
</tr>
<tr>
<td></td>
<td>k</td>
<td>p</td>
<td>82</td>
<td>158</td>
<td>206</td>
<td>2</td>
<td>45.98</td>
<td>1262</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t</td>
<td>123</td>
<td>153</td>
<td>172</td>
<td>0</td>
<td>38.39</td>
<td>1356</td>
</tr>
<tr>
<td>means</td>
<td></td>
<td></td>
<td>78.45</td>
<td>1136</td>
<td>71.21</td>
<td>69.42</td>
<td>1239</td>
<td>1182</td>
</tr>
</tbody>
</table>

Table II

A confusion matrix of all responses to truncated stimuli (4 language groups combined). Mean percent correct (%) and reaction time (ms) values are given separately by condition, target consonant ($C_1$) and context ($C_2$). Cells with correct responses are shaded.
of \(C_1\) Place (/p t k/) and Language Group (Russian, Canadian English, Korean, Taiwanese Mandarin) on participants’ mean percent correct and RT. With respect to percent correct, there were significant effects of \(C_1\) Place and Language Group \([F(2,104)=49.136, p<0.001\) and \(F(3,52)=5.311, p<0.01\) respectively]. Post hoc LSD tests for \(C_1\) Place revealed that /k/ was different from the other two stops \((p<0.001\) for both); /p/ was significantly different from /t/ \((p<0.001\) for both). Post hoc LSD tests for Language Group indicated that results for Russian and English groups were significantly different from those for the Korean group \((p<0.01\) for both). Other group differences were not significant. There was, however, a significant interaction effect of the two factors \([F(6,104)=2.853, p<0.05\] indicating the lack of the \(C_1\) Place effect for the Russian group. With respect to RT (for correct responses), there were significant effects of \(C_1\) Place and Language Group \([F(2,104)=24.562, p<0.001\) and \(F(3,52)=5.754, p<0.01\) respectively]. The interaction of the two factors was not significant. Post hoc LSD tests for \(C_1\) Place revealed that /k/ was significantly different from /p/ and /t/ \((p<0.001\) for both). Post hoc LSD tests for Language Group revealed that Korean listeners performed significantly differently from the other three groups \((p<0.01\) for all three).

Figure 6 presents percent correct (a) and RT (b) results for the ‘complete’ stops by place and language group. Note that /k/ has the highest percent correct values, followed by /p/ and then /t/. This pattern is consistent across all groups, except for Russian listeners (who do not show the /p/ vs. /t/ difference). Overall percent correct values are higher for Russian and English listeners than for Korean listeners, and are not different from Mandarin listeners. In terms of RT, /k/ shows a faster RT than the other stops. Russian listeners exhibit a slower RT than the other groups.

Overall, the current results are very similar to those obtained in Experiment 1 (see §3.3.1). Removing the \(C_2\)VC portion of the signal did not affect percent correct values for \(C_1\) qualitatively (except for the /p/ vs. /t/ difference for Russian listeners). A notable difference was an increase in the Mandarin listeners’ correct responses and faster RTs in the truncated stimuli. Overall, the listeners spent less time identifying consonants in truncated stimuli than in full stimuli, most likely because the former were shorter than the latter. Another difference worth noting was the poorer percent correct and RT results for /p/ in the Russian listeners’ responses.

The results thus support the place target salience hypothesis and, in part, the universal salience hypothesis (based on both percent correct and RT). The overall better performance of Mandarin listeners on truncated (monosyllabic) stimuli than on full (disyllabic) stimuli can possibly be explained by their greater experience with processing monosyllabic words in their native languages. Another possibility, noted by a reviewer, is that the poorer performance of Mandarin listeners in Experiment 1 could be due to the presence of onset consonants \((C_2)\). Given the listeners’ greater experience with onsets than codas, the presence of onset consonants made it more difficult for them to concentrate on codas. The task was easier...
when this distracting information was removed in Experiment 2. Other factors, noted in §3.2.1, could also have affected the performance of both Mandarin and Korean groups of listeners. The relatively poor identification of /p/ by Russian listeners can be attributed to the fact that /p/ in the stimuli was weakly released or sometimes unreleased (see §3.1.1). Russian utterance-final stops, however, are characterised by relatively strong (long and loud) releases (Bondarko 1977).

4.2.2 Percent correct and reaction time: the ‘edited’ condition. Two mixed-design ANOVAs were carried out to examine the effects of C₁ Place and Language Group on participants’ mean percent correct and RT. With respect to percent correct, there was a significant effect of C₁ Place [$F(2,104) = 38.616, p < 0.001$], but not of Language Group. The interaction of the two factors was not significant. Post hoc LSD tests further indicated that /k/ was significantly different from /t/ and /p/ ($p < 0.001$ for both). Differences between the other places were not significant. With respect to RT, there was also a significant effect of C₁ Place [$F(2,100) = 5.812, p < 0.01$]. The effect of Language Group was not significant, nor

The reaction time results are based on data from 54 of the 56 participants (due to no responses to some of the stimuli from two participants, one Russian and one English listener).
was the interaction of $C_1$ Place and Language Group. Post hoc LSD tests indicated that /k/ was significantly different from /p/ and /t/ ($p < 0.05$ for both).

Figure 7 presents percent correct (a) and RT (b) results for the ‘edited’ stops according to place of articulation and language group. Notice that /k/ has substantially lower percent correct values than /p/ and /t/ (yet values for /k/ are above the chance level, which is 33.3%). This pattern is shown by all language groups. Overall percent correct values are also similar across the groups. In terms of RT, identification of /k/ was slower than /p/ and /t/. There are no consistent differences among the language groups.

A comparison of the current results with those found in Experiment 1 (the ‘complete’ condition; see §3.2.2) reveals that removing the $C_2$VC portion from the signal resulted in markedly improved percent correct values for the coronal /t/ – on average by 17%. The other consonants were hardly affected. Again, there was an increase in percent correct values of all three consonants in the VC stimuli for Mandarin listeners. Although it took listeners less time overall to identify stops in the truncated stimuli than in the full stimuli, this difference can be attributed to the relative duration of the two sets of stimuli. A notable difference is decrease of almost half a second (451 ms) in the Russian listeners’ RT values for unreleased /t/.

The results thus do not support the place target salience hypothesis (7a). The ‘edited’ (unreleased) dorsal is no more salient than the labial and the coronal; /p/ is no more salient than /t/. The difference between the scale of perceptual salience established on the basis of the ‘edited’ stimuli when truncated (labial, coronal > dorsal) or not truncated (labial > coronal > dorsal) lies in the relation between the labial and the coronal. This suggests that coronal VC transitions proper are not ‘weaker’ than non-coronal VC transitions (Byrd 1992, Jun 1995, 2004).

Rather, the coronal VC transitions are less robust contextually, and are easily ‘overridden’ perceptually by following non-coronal CV transitions. At least this is the case for Russian stops in this particular vowel context; whether this effect is context-specific or general is left for further research. As in Experiment 1, the results strongly support the release salience hypothesis (based on percent correct and RT): released stops used in the ‘complete’ condition are substantially more salient than unreleased stops in the ‘edited’ condition. The results also largely support the universal salience hypothesis (based on percent correct and partly on RT). The reason for the better RT results for /t/ (compared to the other stops) in the Russian listeners’ responses is not clear.

To control for the possible effect of closure noise, the results for the ‘complete’ and ‘edited’ stimuli were compared for the homorganic sequences. Interestingly, identification of the ‘edited’ homorganic stimuli

Note that although Jun’s (1995, 2004) conclusions about the coronal–non-coronal asymmetry were based mainly on previous work on English, the asymmetry was assumed to be language-independent (see e.g. Jun 1995; 125–127).
was no worse or slower than the ‘complete’ homorganic sequences (56% vs. 50% and 1225 ms vs. 1268 ms). This shows that, whatever closure noise was present in ‘complete’ stimuli, it was not helpful in the identification of C₁. As with the heterorganic stimuli, dorsals in homorganic stimuli showed the lowest percent correct and the slowest RT.

4.2.3 Types of errors: the ‘complete’ and ‘edited’ conditions. Figure 8 presents proportions of assimilatory errors in the VC stimuli. The results are presented separately by language group (a) and place of articulation (b). As seen in Fig. 8a, assimilatory errors were more common than non-assimilatory errors in the ‘edited’ condition, while the two types of errors were almost equally common in the unreleased condition. This pattern was observed for all language groups. The average difference between the two conditions in proportions of assimilatory errors was 11% (60% vs. 49%). As seen in Fig. 8b, released /t/ shows a predominance of assimilatory errors; for the other stops, non-assimilatory errors were as common as or even more common than assimilatory errors.

Taking both types of errors into account, there were no major response biases in the ‘complete’ condition (on average 36%, 35% and 30% for the ‘p’, ‘t’ and ‘k’ responses respectively). In the ‘edited’ condition, all language groups strongly disfavoured the ‘k’ response over the ‘p’ and ‘t’
responses (15% vs. 38% and 47%). Also, all groups except for English showed some preference for the ‘t’ response over the ‘p’ response. This reflects the high number of confusions of the edited /k/ with the other two stops, particularly with /t/ (see Table II).

The results for the truncated ‘complete’ stimuli are similar to the results for the ‘complete’ stimuli used in Experiment 1 (see §3.2.3): both show consistent dominance of assimilatory errors. However, the proportion of assimilatory errors is lower in the truncated stimuli by an average of 12% (and is almost absent for Mandarin listeners). The lack of the assimilatory effect in the truncated ‘edited’ stimuli is in contrast with the predominance (albeit relatively small) of assimilatory errors in the ‘edited’ stimuli used in Experiment 1.

To examine whether the presence of an additional syllable-final /p/ in the stimuli (t[pap], t[pap], etc.) had a biasing effect on listeners’ responses in Experiment 1, we compared the overall numbers of ‘p’ responses to both the full and the truncated stimuli. If such an effect were present, we would expect a considerably higher percentage of ‘p’ responses to the full stimuli than to the truncated stimuli. Overall, there were 1898 ‘p’ responses to the full stimuli, (36.09% of all responses; see Table I); there were 1927 ‘p’ responses to the truncated stimuli (35.90% of all responses; see Table II). Even if the biasing effect were present in the full responses, its magnitude was negligible (0.19%).
The analysis of errors helps us explain the results obtained in Experiment 1. The assimilatory bias found there apparently had two sources: (i) the CV transitions and the release of C₂, and (ii) the release of C₁. The role of CV transitions and releases of C₂ as a source of perceptual place assimilation has been previously noted (Fujimura et al. 1978, Repp 1983b, Bladon 1986, Ohala 1990, Byrd 1992, Kingston & Shinya 2003). A similar role for releases of C₁, however, has not received as much attention (see Repp 1983a). Interestingly, in our results this factor primarily involved /t/, the stop with the longest release in the stimuli (see §3.1.1). The VC transitions to C₁ in our results did not appear to contain C₂ information (unlike in English and Korean: Byrd 1992, Zsiga 1994, Son et al. 2007). This can be explained by the relative lack of overlap in Russian stop clusters in general, and in the stimuli in particular (see §3.1.1).

4.3 Conclusion

In Experiment 2, the relative salience of syllable-final stops (C₁) was examined in the absence of the following consonant (C₂). The results of the experiment re-confirmed the place target salience hypothesis (dorsal > labial > coronal) for the ‘complete’ stops, which were cued by both the VC transitions and (in most cases) the releases. However, the results did not confirm this hypothesis for the ‘edited’ stops, which were cued only by the VC transitions. The scale established for this condition (labial, coronal > dorsal) was in part similar to that obtained for the unreleased stops in Experiment 1 (labial > coronal > dorsal). The results confirm the release salience hypothesis (released > unreleased) and, to a large extent, the universal salience hypothesis. Assimilatory errors were more common than non-assimilatory errors only for /t/ in the ‘complete’ condition. This allowed us to determine the main sources of the assimilatory bias: C₁ releases (for coronals) and C₂ CV transitions and releases.

5 General discussion and conclusion

The goal of the study was to investigate the relation between facts of phonetic difficulty in the perception of syllable-final stops in clusters and common sound patterns of place assimilation. Experiments 1 and 2 were intended to test three specific hypothesis of the PKG approach in (7): the place target salience hypothesis (dorsal > labial > coronal), the release salience hypothesis (released > unreleased) and the universal salience hypothesis (relative salience is language-independent). The results of the experiments were also expected to provide evidence for the place trigger hypothesis (labial/dorsal > coronal) and to clarify the role of phonetic knowledge in shaping certain properties of place assimilation processes.

The methodological choice of the language of the stimuli (Russian, a language without place assimilation) and the native languages of the listeners (Russian, English, Korean and Mandarin) was expected to
provide some control for both production and perception of place distinctions. The results of both experiments do in fact show that the relative scales of salience (place target salience and release salience) are largely language-independent, thus supporting the universal salience hypothesis. Thus, the assumption taken by the PKG approach that ‘certain basic conditions governing speech perception...are necessarily shared by all languages, experienced by all speakers’ (Hayes & Steriade 2004: 1) appears to be valid. Still, the universal status of the scales determined by the current experiments should be taken with caution. It reflects the relative salience of stops in the stimuli used in the experiment – Russian voiceless stops in a single vocalic context. While we propose that the results can be extended to similar syllable-final stops in general, future work is required to determine the full extent of the universality of the current findings. Specifically, more work is needed to examine the relative contributions of different degrees of gestural overlap, gradient reduction, voicing and vowel contexts to scales of perceptual salience. Further, the current results obtained with Korean and Taiwanese Mandarin listeners who are English bilinguals should be taken as possibly reflecting some influence of L2, particularly some familiarity with English released syllable-final stops. Ideally, the current results need to be replicated with monolingual speakers of these languages.

The results of the study demonstrate that there are robust phonetic differences – degrees of perceptual difficulty or salience – between released and unreleased stops, as well as between stops of different places of articulation. The finding that released syllable-final stops are considerably more salient than unreleased stops fully confirms the release salience hypothesis (cf. Male´cot 1958, Winitz et al. 1972, Wright 2004). Releases thus provide important information about the place of articulation of C₁. Interestingly, this information is beneficial (to different degrees) both to speakers who are accustomed to syllable-final stop releases in their native languages (Russian and Canadian English vs. Korean and Taiwanese Mandarin) and to those who are not. Another important finding is that releases are of different importance for different places of articulation, partly corresponding to the relative ‘strength’ (amplitude and duration) of place releases (dorsal > coronal > labial). The finding shows that more acoustic information in speech signal can correspond to greater perceptual salience, thus supporting the reliance of some previous PKG work on acoustic differences (Jun 1995, 2004, Steriade 2001, among others). While releases are clearly perceptually important for the identification of C₁, they may also contain information about C₃ (cf. Repp 1983a). Such conflicting cues can be potentially confusing to listeners, resulting in assimilatory errors (particularly for coronals). The results thus demonstrate the importance of acoustic and perceptual properties of different place releases (as well as closure noise), which are in part related to different degrees of overlap of articulatory gestures (Zsiga 2000, Chitoran et al. 2002).

While the results provide clear support for the universal salience and release salience hypotheses, support for the place salience scale is partial
and less clear. The perception of stops in the ‘complete’ condition (cued by VC transitions and releases) exhibited the expected relative place salience relations: dorsals are the most salient, followed by labials and then coronals (dorsal > labial > coronal). The perception of stops in the ‘edited’ condition (cued by VC transitions only) exhibited the expected relations between labials and coronals (labial > coronal), but not those involving dorsals (the expected dorsal > labial, coronal). The results for unreleased dorsals clearly contradict Jun’s (1995, 2004) perceptual salience scale, where dorsals are considered to be the most salient. Recall that that scale was deduced mainly from the acoustics of unreleased stops. This shows that more acoustic information (such as F3 lowering for dorsals) does not always correspond to greater perceptual salience, suggesting that any generalisations about perceptual salience should be based primarily on perception rather than acoustics (cf. Hume et al. 1999, Winters 2003).

Another unexpected finding was that the relative salience of ‘edited’ labials and coronals (labial > coronal) was not observed in truncated stimuli. This suggests that unreleased coronals are not less salient than labials *per se*, but rather when followed by a consonant of a different place. VC transitions to coronals appear to be easily ‘masked’ by the CV transitions of following consonants. Non-coronals are less susceptible to such masking. Note that another possible source of the coronal/non-coronal perceptual asymmetry – high degrees of overlap (Byrd 1992, Zsiga 1994, Surprenant & Goldstein 1998, Son et al. 2007) – could not be observed in our results, due to the fact that there was relatively little overlap in clusters in the stimuli. This factor may also explain the lack of support for the scale of place trigger salience.

Another question to be addressed is a possible influence of lexical frequency on perceptual responses (see e.g. Blevins 2007). Do the place salience scales established in the current experiments represent general facts of phonetic difficulty or are they mere artefacts of listeners’ frequency biases? More research is needed to provide a fuller answer to this question. As far as the current results are concerned, however, little can be said about the role of lexical frequency in listeners’ responses.18

What do the results of the study tell us about the relation between the physics of speech and phonological symbols? In other words, what do the results tell us about the relation between facts of phonetic difficulty (our

18 Although Russian, English and Korean allow syllable-final stops, relative frequencies of these consonants are not exactly the same. For example, in both Russian and English, /t/ is found in syllable-final position more often than /k/ or /p/, while /k/ is more frequent in this position than /p/. The magnitude of these differences, however, is greater for Russian than for English (Kucéra & Monroe 1968: 57, Kessler & Treiman 1997: 302). If frequencies of the stops in Russian and English were relevant to the listeners’ performance, one would expect them to be reflected in perception errors. Specifically, ‘t’ responses would have dominated ‘k’ and ‘p’ responses in errors as well as ‘k’ responses would have dominated ‘p’ responses (t > k > p). Also, place differences in responses would have been greater for Russian than for English listeners. Further, such frequency-based effects would have held in both experimental conditions (with ‘edited’ and ‘complete’ stimuli). None of these effects, however, were found in our results.
scales of place salience) on the one hand and grammatical constraints and sound patterns of place assimilation they define on the other? Clearly, there is a partial correspondence between some facts of phonetic difficulty and patterns of place assimilation. The lower perceptual salience of coronals compared to released or unreleased labials and to released dorsals corresponds to their greater susceptibility to place assimilation (Jun 1995, 2004), and possibly to their relative unmarkedness in general (Paradis & Prunet 1991, among others). Note that the lower salience status of coronals is a result of several factors, among which is their vulnerability to masking by following consonants and confusability of their releases. The greater salience of released dorsals, compared to released labials and coronals, corresponds to their lesser susceptibility to place assimilation. Yet the poor salience of unreleased dorsals (at least in some vowel contexts) would predict their greater susceptibility to place assimilation, compared to labials and coronals. This is apparently in contrast to what has been observed cross-linguistically (Jun 1995, 2004; however, see Rice 1996, 1999 on the relative unmarkedness of velars).

The sensitivity of place salience scales to phonetic detail such as presence or absence of releases (cf. Winters 2003) appears to present a problem for the PKG approach. Assuming that the perceptibility scales obtained in this study are truly universal, we would expect two different types of assimilation patterns to emerge. Speakers of a language where stops are commonly released, such as Russian, would posit a constraint hierarchy based on their acquired knowledge of the relative salience of released stops (dorsal > labial > coronal), shown in (11a). In contrast, speakers of a language where stops are commonly unreleased, such as Korean, would posit a different constraint hierarchy, based on their knowledge of the salience of unreleased stops (labial > coronal > dorsal), shown in (11b). This is, however, contrary to what the PKG approach predicts: speakers of different languages are expected to arrive at similar constraint rankings, thus resulting in cross-linguistically recurrent patterns of assimilation.

\[
\begin{align*}
(11) \quad a. \quad & \mathrm{Pres}(\mathrm{pl}(\mathrm{dor})) \gg \mathrm{Pres}(\mathrm{pl}(\mathrm{lab})) \gg \mathrm{Pres}(\mathrm{pl}(\mathrm{cor})) \\
& \quad b. \quad \mathrm{Pres}(\mathrm{pl}(\mathrm{lab})) \gg \mathrm{Pres}(\mathrm{pl}(\mathrm{cor})) \gg \mathrm{Pres}(\mathrm{pl}(\mathrm{dor}))
\end{align*}
\]

Note that the constraint hierarchy in (11a) is empirically adequate, and can generate attested patterns of place assimilation, such as greater assimilatory susceptibility of coronals, followed by labials and then dorsals (cf. (5)). The hierarchy in (11b), however, is empirically problematic; it predicts patterns of place assimilation that apparently have not been attested – those involving greater assimilatory susceptibility of dorsals, followed by coronals and then labials. Rather puzzlingly, such a pattern of place assimilation is the exact opposite of the pattern found in Korean (see (3)). In sum, there is a partial mismatch between scales of perceptual salience deduced from phonetics and phonological markedness hierarchies. This suggests that, if speakers do indeed use perceptual salience
scales in constructing grammatical constraint hierarchies, such scales have to be relatively abstract, context-independent and ‘blind’ to minor phonetic detail. With respect to place of articulation, the scales can possibly refer to ‘prototypical’ realisations of such segments – fully released (syllable-initial or syllable-final) stops, or to more abstract phonetic/phonological representations of the consonants generalised across a range of phonetic and lexical contexts. Another possibility is that the perceptibility scales internalised by speakers are only partially responsible for shaping markedness hierarchies. The mapping between the two may be obscured by other synchronic and diachronic factors yet to be identified. In either case, the correspondence between the facts of phonetic difficulty and recurrent sound patterns appears to be less direct than predicted by the PKG approach.

The ‘indirect mapping scenario’ is also supported by the current finding of a strong perceptual bias towards assimilatory errors. The presence of this bias in experimental conditions suggests that regressive assimilation in consonant clusters can arise, at least in part, due to listeners’ misperceptions, without the necessarily direct reference to listeners’ phonetic knowledge (Ohala 1981, 1990; cf. Hume & Johnson 2001a, Blevins 2004). Overall, misperceptions are more likely when stops are unreleased (and stop gestures are more overlapped; Byrd 1992, Son et al. 2007). This suggests that the lack of stop releases is a factor eroding place distinctions and thus a likely historical step towards phonological place assimilation. The diachronic source of place assimilation and patterns of markedness in general, therefore, cannot be discounted. To conclude, a better understanding of phonetic grounding of phonological symbols – grammatical constraints and patterns they capture – is likely to be achieved by a phonetically based approach that integrates synchrony and diachrony in a principled way.

REFERENCES


Berkeley: Department of Linguistics, University of California, Berkeley. 2069–2072.


