2018-05-01

Perceptual attention as the locus of transfer to nonnative speech perception

Version	Accepted manuscript
Citation (published version):	Charles B. Chang. 2018. "Perceptual attention as the locus of transfer to nonnative speech perception." Journal of Phonetics, Volume 68, pp. 85 - 102 (18). doi: 10.1016/j.wocn.2018.03.003

https://hdl.handle.net/2144/29808

Boston University

Perceptual attention as the locus of transfer to nonnative speech perception

Charles B. Chang[a]

[a]Boston University, Linguistics Program, 621 Commonwealth Avenue, Boston, MA 02215, USA

Abstract

One's native language (L1) is known to influence the development of a nonnative language (L2) at multiple levels, but the nature of L1 transfer to L2 perception remains unclear. This study explored the hypothesis that transfer effects in perception come from L1-specific processing strategies, which direct attention to phonetic cues according to their estimated relative functional load (RFL). Using target languages that were either familiar (English) or unfamiliar (Korean), perception of unreleased final stops was tested in L1 English listeners and four groups of L2 English learners whose L1s differ in stop phonotactics and the estimated RFL of a crucial cue to unreleased stops (i.e., vowel-to-consonant formant transitions). Results were, overall, consistent with the hypothesis, with L1 Japanese listeners showing the poorest perception, followed by L1 Mandarin, Russian, English, and Korean listeners. Two exceptions occurred with Russian listeners, who underperformed Mandarin listeners in identification of English stops and outperformed English listeners in identification of Korean stops. Taken together, these findings support a cue-centric view of transfer based on perceptual attention over a direct phonotactic view based on structural conformity. However, transfer interacts with

---

Email address: `cc@bu.edu` (Charles B. Chang)
URL: `http://charleschang.net/` (Charles B. Chang)

Preprint submitted to Journal of Phonetics *March 4, 2018*

prior L2 knowledge, which may result in significantly different perceptual consequences for a familiar and an unfamiliar L2.

Keywords: selective perception routine, language transfer, unreleased stops, cue weighting, information value, functional load, coarticulation

1. Introduction

1.1. L1 influence on L2 perception

An enduring question in the study of second language (L2) acquisition has been the manner in which the phonological system of the native language (L1) constrains the development of an L2, especially an L2 to which a listener was not exposed until late in life. Although it is clear that adult L2 learners maintain access to at least some of the cognitive resources that contribute to successful L1 acquisition (see, e.g., Flege, 1995), they also tend to experience interference from their L1 knowledge, resulting in performance deficits vis-a-vis L1 speakers that are widely documented in the speech perception literature (Bradlow & Pisoni, 1999; Cutler, 2001; Cutler et al., 2008; Nábělek & Donahue, 1984). This phenomenon of crosslinguistic influence (in particular, of an L1 on an L2) is often referred to as TRANSFER (Altenberg, 2005; Bohn, 1995; Odlin, 1989).

The fact that different L1 backgrounds lead to disparate outcomes with the same L2 suggests that what gets transferred in L2 learning are specific aspects of L1 knowledge; however, the precise nature of transferred L1 knowledge is not well understood. In particular, there is no general consensus regarding the basis of transfer effects observed in L2 speech, although various bases have been described in the literature (e.g., Polka, 1991, 1992): phonetic (a mismatch between the fine-grained phonetic properties of a target L2 category or structure and those of its L1 correspondent[1]), phonemic (a mismatch between a target L2 segment and the

---

[1] A review of the issues involved in identifying this L1 correspondent is outside the scope of this paper. However, it should be noted that, at least for experienced L2 learners, the identification of crosslinguistic correspondents is probably not based solely on phonetic proximity, but rather heavily influenced by higher-level phonological considerations (Chang, 2015; Chang et al., 2011).

L1 inventory), and phonotactic (a mismatch between a target L2 structure and L1 distributional patterns). The relative importance of these factors was the subject of a study by Davidson (2011b) comparing L1 Catalan, English, and Russian listeners on perception of nonnative consonant clusters (generated by removing the first vowel in multisyllabic sequences of Catalan). Results showed that Russian listeners (familiar with the widest variety of clusters from their L1) were better at discriminating between the presence and absence of a cluster than Catalan listeners, who were in turn better than English listeners. Crucially, the Russian advantage occurred in spite of the fact that certain test consonants and all the phonetic implementations were nonnative (namely, those of Catalan), suggesting that "the presence of the relevant phonological structure in one's native language is perhaps the most important predictor of discrimination ability" (Davidson, 2011b, p. 280).

Another study that examined both phonetic and phonological influences of the L1 on L2 perception is Cho and McQueen's (2006) investigation of stop perception by L1 Korean and Dutch listeners. The goal of this study was to examine two accounts of L2 perception: a "phonological superiority" hypothesis linking L2 perception to (non)conformity of the L2 target with L1 constraints, and a "phonetic superiority" hypothesis linking L2 perception to the richness of the cohort of cues to the L2 target. To this end, listeners were tested on their ability to detect word-final voiceless stops in American English (a familiar L2 for both L1 groups) and Dutch (an unfamiliar L2 for the Korean group), both when the stops were released and when they were "dereleased" (i.e., unreleased because the release was spliced off). The results showed that, for both target languages, Korean listeners detected unreleased stops (which conform to the L1 pattern of final non-release, but are signaled by a weaker cohort of cues) more rapidly than released stops (signaled by a richer cohort of cues); however, their detection accuracy was higher for released stops, albeit only in English. In contrast, Dutch listeners detected released stops (which conform to the L1 pattern of final release) more rapidly and/or more accurately than unreleased stops. These findings were thus interpreted as supporting the "phonological superiority" hypothesis, while evincing an effect of cue richness given sufficient familiarity with the target L2.[2]

Although the above findings were given a phonological explanation, the question remains as to *how* phonological constraints such as phonotactic restrictions influence the perception of L2 speech. One approach to this question is to place speech perception squarely in the purview of phonology and, therefore, to account for perception using the same kinds of formal constraints used to account for phonological phenomena more generally (see, e.g., Escudero, 2009; Steriade, 2009). This type of account has, in fact, been used to explain transfer effects in perception, including L2 "perceptual illusions" (Berent et al., 2007; Dupoux et al., 1999; Parlato-Oliveira et al., 2010) and "perceptual assimilation" of sound sequences (Hallé & Best, 2007; Hallé et al., 1998). For example, the case of L1 Japanese speakers perceiving an illusory vowel within L2 consonant clusters was

---

[2] Note, however, that because the L1 pattern considered in this study can be interpreted as a fact about phonetic realization—i.e., the quality of final stops, as opposed to their (non)occurrence—the results may also reflect a phonetic kind of L1 transfer effect, even if the difference in phonetic realization between Korean and English arises through a categorical phonological process of laryngeal neutralization in Korean and a more variable type of process in English. What is crucial—because it could lead to difficulty in L1 Korean listeners' perception of English final stops—is the degree of perceived phonetic disparity between Korean and English final stops (cf. Park & de Jong, 2017, for perceptual mapping data suggesting that L1 Korean listeners perceive both released and unreleased English coda stops as unlike Korean stops).

attributed to a phonotactic ban against the clusters in the L1; similarly, L1 Mandarin speakers' tendency to misperceive English *can't* as *can* was interpreted as a "clear direct effect of their native language's ban on /nt/ clusters" (Ernestus et al., 2017, p. 60). As such, this view of L2 perceptual deficits is referred to here as the DIRECT PHONOTACTIC VIEW. The core of this view, crucially, is its linking of the difficulty of perceiving an L2-specific target x (where x may be a phoneme, a sequence of phonemes, or a subphonemic feature) *directly* to x's partial or total absence from the L1, exemplified in the proposal that "if a learner's L1 grammar lacks the phonological feature that differentiates a particular non-native contrast, he or she will be unable to perceive the contrast" (Brown, 1998, p. 136; see also Brown, 2000).[3] Thus, the logic of this view is that poor L2 perception of x arises because x *does not occur* in the L1, resulting in the listener either not expecting or failing to listen for x in the L2.

In contrast to the direct phonotactic view, there is an alternative, cue-based approach to explaining L2 perceptual patterns related to phonotactics. In fact, a cue-based explanation of the Russian advantage in Davidson (2011b) is alluded to by Davidson, who observed that "[i]f a contrast such as /#fət/∼/#ft/ exists in a language, listeners would have to closely attend the acoustic information corresponding to the schwa. However, if a language only allows one of these possibilities, then the production of the other sequence may be treated as a less optimal but potential variant of the phonotactics that do exist" (p. 279). In other words, perhaps the Russian advantage in discriminating clusters from non-clusters is not due to L1 phonotactics *per se*, but rather to the pattern of targeted perceptual at-

---

[3] This type of view was also reflected in the "contrastive analysis" approach to predicting L2 difficulties (Lado, 1957), which was based on the (non)occurrence in the L1 of an L2 target.

tention (PA) resulting from the L1 phonology. Russian listeners' L1 experience has tuned their perception to devote more PA to the properties of a vocalic interval between initial consonants because the presence of an intervening vowel has significant linguistic consequences (e.g., making a different word) with consonant sequences of all different types, whereas Catalan and English listeners' L1 experience has resulted in less PA to this vocalic interval because this is not as important in their respective L1s (which allow a comparatively limited set of clusters). Since this account links L2 perception to L1-specific attunement to phonetic cues (rather than directly to L1 phonotactics), it is referred to here as the CUE-CENTRIC VIEW. Note that this view does not reject the existence of phonotactics, which are understood to be part of what shapes PA to a cue in a given language. Rather, it does not base predictions for L2 perception on L1 phonotactics *in the first instance*. The predictions of this view come instead from a cue-based level of analysis, which thus subsumes certain "indirect effects" of L1 phonotactics such as (for a given L2 contrast) "difficulties interpreting the subsegmental cues because these cues do not occur or have different functions" in the L1 (Ernestus et al., 2017, p. 50).

In short, the L1 knowledge transferred to L2 perception can be conceptualized either in terms of categorical phonotactic constraints or in terms of gradient attunement to phonetic cues; however, although categorical phonotactics are part of the linguistic conditions that make a phonetic cue more or less important in a given language, the coarseness of categorical phonotactics limits the empirical power of the direct phonotactic view. In particular, the kind of contrastive analysis at the heart of the direct phonotactic view predicts only two types of transfer: "negative" transfer, which results in a perceptual decrement relative to L1 listeners (e.g., Goto, 1971; Sheldon & Strange, 1982), and "neutral" transfer, which results in performance comparable to L1 listeners' (e.g., Iverson et al., 2003). However, under certain conditions L1 influence may also manifest as "advantageous" transfer, which results in better-than-native perception (e.g., Bohn & Best, 2012; Chang & Mishler, 2012; Hallé et al., 1999). Such a NATIVE-LANGUAGE TRANSFER BENEFIT does not follow from phonotactic comparisons across languages (because, once a target is allowed to occur in a given context, it is not meaningful to talk of it being "more allowed" in that context in the L1 vs. L2), but is amenable to an explanation in terms of PA to phonetic cues.

A cue-centric view of transfer, however, has to account for the multidimensional nature of speech, which typically contains, for each contrast, multiple possible phonetic cues. So how do listeners sort out the multiple aspects of the speech signal to which they could attend? This is one of the main questions addressed in the automatic selective perception (ASP) framework for understanding crosslinguistic speech perception (Strange, 2011; cf. the overlapping PRIMIR framework of Werker & Curtin, 2005). According to ASP, L1 acquisition involves the development of "selective perception routines" (SPRs) that allow perception to be targeted, automatic, and robust in adverse conditions. SPRs are critical to becoming a skilled L1 listener; however, they are also the source of L1 interference in perception of an L2, which often requires the listener to attend to different properties of the speech signal than required by the L1 and/or to integrate them differently. Crucially, ASP posits that older learners maintain access to the language-general processing abilities evident in childhood. However, use of these abilities is affected by two factors: task demands (with high demands causing default to automatized, L1-specific SPRs) and L2 experience (with extensive experience leading to "phonologization" of L2 perception; see, e.g., Levy & Strange, 2008).

1.2. Relative functional load of a cue

In addition to properties of the perceiver (e.g., experience) and task (e.g., demands), properties of the stimulus are also likely to influence speech processing. In particular, two properties of a cue may affect the degree to which listeners attend to it: FUNCTIONAL LOAD and ACOUSTIC RICHNESS. The information-theoretic notion of functional load is usually applied to phonological contrasts (e.g., Martinet, 1933; Wedel et al., 2013), but may also be extended to the phonetic cues that distinguish them. If a contrast's functional load is the unique burden that it shoulders in distinguishing lexical items (measured in terms of minimal pairs differing in that contrast), then a cue's functional load can be thought of as its unique burden in distinguishing phonological contrasts; therefore, this goes up as the number of contrasts involving that cue increases, and down as the number of other cues helping to distinguish those contrasts increases. Note that this concept of a cue's functional load is inherently relative, because in order to estimate the unique burden of one cue given the multidimensional nature of speech, it is necessary to take into account other contributing cues; therefore, for clarity this concept is referred to here as RELATIVE FUNCTIONAL LOAD (RFL).

How does one estimate RFL of a given cue x? According to the above description, to increment RFL for each contrast that x distinguishes, one should divide by the number of other cues to that contrast; however, the load of each cue in the cohort probably depends on its availability, with a cue that is variably available shouldering less of a load than a cue that is always available. Therefore, it is reasonable to posit that the RFL for one cue accounts for the contributions of other cues according to their availability.[4] To illustrate what this means mathematically,

---

[4] The accuracy of RFL estimation will, therefore, be limited by our knowledge of what belongs

a sketch of a formula for RFL is provided in (1), where RFL of cue x is expressed as a function of a_x (availability of x as a proportion of time), c (number of contrasts distinguished by x), ω_y (number of other cues to current contrast y), and a_z (availability of the current other cue z).

(1) $$RFL_x = a_x \cdot \sum_{y=1}^{c}\left(1 - \frac{\omega_y}{1+\sum_{z=1}^{\omega_y} 1 \cdot a_z}\right)$$

RFL estimation, using (1) to predict a crosslinguistic hierarchy, is exemplified in Section 1.3. Note that, for one specific contrast, RFL is similar to the notion of "cue weighting"; however, RFL is a broader concept since it incorporates the linguistic work of cuing multiple contrasts across the language.

As for acoustic richness, this refers to a language-general notion of information density. For example, independent of RFL, a stop's release burst is an acoustically rich cue to place of articulation because it provides several clues to place: temporal, amplitudinal, and spectral (e.g., dorsal bursts tend to show longer duration, higher amplitude, and higher-frequency energy than labial ones). In contrast, formant transition cues to place provide mainly spectral information. This disparity in acoustic richness explains why, although burst cues have lower RFL than transition cues with respect to distinguishing final stops in English (due to variable availability of bursts in English), when the two are pitted against each other in cross-spliced stimuli, L1 English listeners tend to follow the burst cues (Wang, 1959). In other words, acoustic richness may override RFL with respect to directing attention to a cue. However, in the present study this will not be relevant, as the materials purposefully avoid setting up a conflict between different cues.

---

in the cohort of cues to any given contrast.

1.3. The present study

The study reported in this article endeavored to test a cue-centric view of L1 transfer based in RFL against a direct phonotactic view, focusing on the case of final stop perception. In regard to investigating transfer effects, final stop contrasts are useful to consider for three reasons. First, final stops are well-attested in the languages of the world, and the three cues—preceding vowel duration, vowel-to-consonant (VC) formant transitions, and release burst—occur, broadly, in any language that has stops (since they also occur in VCV sequences). Second, cues to place of articulation (transition and burst) are not temporally confounded like cues to many other L2 contrasts, so their respective perceptual effects can be separated more easily. Third, VC transitions, as an outcome of coarticulation, constitute a universal cue to final stops given that coarticulation is a universal phenomenon (Lindblom & MacNeilage, 2011). As previously mentioned, a release burst provides another, acoustically rich cue to stop identity, but may not always be available. In American English, for example, final stops are often unreleased (Byrd, 1993; Davidson, 2011a; Kang, 2003; Rositzke, 1943), while in Korean, final stops are consistently unreleased (Sohn, 1999).[5] Unreleased final stops thus provide an ideal testing ground for a study of transfer effects, since the perception of place in an unreleased stop relies on one highly available cue—VC transitions—to which

---

[5] Although Kim and Jongman (1996) describe Korean final stops as often having a (weak) release, note that they examined a specific utterance-medial context in which an alveolar stop was embedded before a velar stop, which is likely to cause the first stop to be incidentally released due to the articulatory coordination involved in the alveolar-to-velar transition. Others (including Cho and McQueen, 2006) have described Korean final stops as unreleased, and this was consistent with the Korean recordings for the present study (Section 2.3), which showed a 0% rate of release.

Table 1: Summary of L1 properties relevant to L2 perception of unreleased final stops. Phonemic and phonotactic properties are labeled in binary fashion (i.e., $-$ or $+$); cue-centric properties, in incremental fashion (where $-$ denotes the lowest degree). RFL = relative functional load.

Type	Property	Japanese	Mandarin	Russian	English	Korean
phonemic	vowel length contrast	$+$	$-$	$-$	$-$	$-$
phonotactic	stop contrast / ___#	$-$	$-$	$+$	$+$	$+$
phonotactic	nasal contrast / ___#	$-$	$+$	$+$	$+$	$+$
cue-centric	RFL of vowel duration	$++$	$+$	$+$	$+$	$+$
cue-centric	RFL of VC transition	$-$	$+$	$++$	$+++$	$++++$
cue-centric	RFL of final stop burst	$-$	$-$	$++$	$+$	$-$

any individual whose L1 contains VC(V) sequences would have been exposed.

Thus, the present study examined L2 perception of unreleased final voiceless stops to address two main research questions. First, is L2 perception of unreleased final stops influenced primarily by L1 transfer of categorical phonotactics or of perceptual attention to cues (**Q1**)? Second, how is L1 transfer in L2 perception of unreleased final stops influenced by prior knowledge of the target L2 (**Q2**)?

To address Q1, this study compared listeners from five different L1 backgrounds: Japanese, Mandarin, Russian, American English, and Korean. These languages were selected because of their diverse phonemic, phonotactic, and cue-centric properties (see Table 1), which lead to differences in predicted perceptual attention (PA) to the crucial cue to unreleased stop identity (i.e., VC transitions). Assuming that the role of VC transitions in cuing place contrasts in initial/prevocalic position is relatively small (because in this position place contrasts are cued by perceptually stronger CV transitions and, for stops, an acoustically rich release burst), the following discussion abstracts away from the RFL associated with initial place contrasts and focuses on the RFL of distinguishing final place contrasts. In Japanese, VC transitions draw the least PA because they carry the lowest RFL (namely, 0): the only consonant allowed word-finally is the "placeless" nasal, while the only consonants allowed syllable-finally are always homorganic to the following onset consonant (Iwasaki, 2013), which means that there are effectively no final place contrasts. In Mandarin, VC transitions draw more PA due to a slightly higher RFL, which follows from one place contrast between final nasals /n ŋ/ (a contrast that is also cued by covariation of the preceding vowel; Duanmu, 2007). Per (1), and assuming that the vowel quality cue is always available, this means that the RFL of VC transitions (RFL_{VC}) in Mandarin is approximately 0.5 ($= 1\cdot\sum_1^1(1-\frac{1}{1+\sum_1^1 1\cdot 1})$). In Russian, VC transitions draw yet more PA due to the higher RFL of distinguishing at least four place contrasts, among final nasals /m n/ and plosives /p t k/ (possibly also /mʲ nʲ pʲ tʲ/) (Timberlake, 2004). However, RFL_{VC} remains relatively low, because the VC transitions share the burden of cuing the plosive contrasts with a consistently available burst (Davidson & Roon, 2008; Jones & Ward, 1969; Zsiga, 2003). Counting the three primary points of articulation, RFL_{VC} comes to around 2 (0.5 from the nasals + 1.5 from the plosives; $1\cdot\sum_1^3(1-\frac{1}{1+\sum_1^1 1\cdot 1}) = 1.5$). In English, VC transitions draw more PA than in Russian due to a higher RFL, which follows from a higher number of final place contrasts (among /m n ŋ p t k b d g/) and the lower availability of the burst cue. Assuming an overall burst availability of approximately 0.5 (Davidson, 2011a; Kang, 2003), RFL_{VC} in English comes to around 3.5, including a contribution from nasal contrasts of 1.5 ($= 1\cdot\sum_1^3(1-\frac{1}{1+\sum_1^1 1\cdot 1})$) and a contribution from plosive contrasts of 2 ($= 1\cdot\sum_1^6(1-\frac{1}{1+\sum_1^1 1\cdot 0.5})$). Finally, in Korean, VC transitions draw the most PA because they have the highest RFL, cuing place contrasts among final /m n ŋ/ and /p t k/ (in the latter case, as the *sole* cue since a burst is not available; Sohn, 1999). RFL_{VC} in Korean thus comes to around 4.5, including a contribution from nasal contrasts of 1.5 ($= 1\cdot\sum_1^3(1-\frac{1}{1+\sum_1^1 1\cdot 1})$) and a contribution from plosive contrasts of 3 ($= 1\cdot\sum_1^3(1-\frac{0}{1+0})$)

Predictions in regard to Q1 diverge under the direct phonotactic and cue-centric views because of a difference in their underlying logic. On the one hand, the direct phonotactic view attributes L2 perceptual deficits to nonconformity with L1 phonotactics; therefore, how well L2 listeners can perceive unreleased final stops (of the unmarked, voiceless variety) should follow primarily from whether or not the natural class of L1 stops (i.e., [−sonorant, −continuant]) is allowed finally.[6] On the other hand, the cue-centric view attributes L2 perceptual deficits to the (lack of) motivation to attend to a crucial auditory cue, which is closely related to the cue's RFL in the L1; therefore, L2 listeners' ability to perceive unreleased final stops should follow primarily from RFL_{VC} in the L1. These two views thus predict different outcomes for Q1. Under the direct phonotactic view, all L2 listeners who speak an L1 disallowing final stops (e.g., Japanese, Mandarin) should be equally poor at perceiving unreleased final stops because the phonotactic handicap imposed by their L1s is the same. In contrast, under the cue-centric view, L2 listeners subject to the same L1 phonotactic constraint are still likely to show perceptual variation due to differences among L1s in RFL_{VC}. That is, L2 listeners should be poor at perceiving unreleased final stops only insofar as RFL_{VC} in their

---

[6] Variants of this view incorporating constraints on other features (e.g., place of articulation features) are discussed further in Section 4, where it is shown that these alternative formulations of the relevant phonotactic constraints do not significantly alter the empirical coverage of this view.

L1 is low (which would discourage attending to VC transitions). This predicts, for example, that L1 Japanese and Mandarin listeners will not be equally poor at perceiving unreleased final stops; rather, Mandarin listeners should be better because of the higher RFL_{VC} in Mandarin.

Given the linguistic differences outlined in Table 1, there were three specific predictions that followed from the cue-centric view. **P1**, in regard to a familiar L2 (English), was that perceptual success with L2 unreleased final stops would be correlated with the PA devoted to VC transitions in listeners' L1; therefore, the following cline of success was predicted (from lowest to highest): Japanese < Mandarin < Russian < Korean. Note that one part of this cline (Japanese < Korean) is supported by data in Tsukada et al. (2007), where unreleased stops from Thai and released and unreleased stops from Australian English were better discriminated by Korean than Japanese listeners. As for the complementary case of perceiving the *absence* of a final stop, a useful cue to (non)occurrence of a coda other than VC transitions is vowel duration, which tends to be shorter in closed than in open syllables crosslinguistically (Katz, 2012; Maddieson, 1985). Since vowel duration also marks a phonemic length contrast in Japanese (Tajima et al., 2008) but not in Mandarin, the RFL of vowel duration is higher in Japanese (Table 1); this should result in Japanese listeners attending to vowel duration more than Mandarin listeners, which could compensate for, or even overcome, their lack of PA to VC transitions with respect to detecting final stop occurrence. Consequently, **P2** was that Japanese listeners would be no worse (and possibly better) than Mandarin listeners at telling that a speech stimulus did not end in /p t k/.

In regard to Q2, following from ASP's notion of SPRs and a positive relationship between L2 experience and phonologization of L2 perception, it was hypothesized that negative transfer would be more evident in the perception of an unfamiliar, as opposed to familiar, L2, as an unfamiliar L2 would not yet be associated with any L2-specific SPRs. Consequently, listeners were tested on perception of unreleased final stops in two L2s: English (familiar) and Korean (unfamiliar). Since greater transfer of L1 SPRs was expected in perception of Korean, it followed that a relative lack of PA to VC transitions in the L1 should particularly disadvantage listeners in perception of Korean. Thus, **P3** was that group differences between L1s where the RFL of VC transitions is lower (i.e., Japanese, Mandarin) vs. higher (i.e., Russian, English, Korean) would be larger in the perception of Korean than in the perception of English.

2. Methods

2.1. Participants

Participants in the perception experiments were five groups of listeners with different L1s: American English (NEng), Japanese (NJpn), Korean (NKor), Mandarin Chinese (NMnCh), and Russian (NRus). The NEng and NKor groups were those from Chang (2016). All listeners were recruited from the Greater Washington, DC and New York metropolitan areas, gave informed consent, and were paid for their participation. Due to a lack of the proper equipment, participants were not able to undergo formal audiometric evaluation; however, their background questionnaires indicated no history of hearing, speech, or language impairments.[7] The five groups consisted of an equal number of participants, who were gender-matched and comparable in mean age (early to late 20s; see Table 2).

---

[7] The full list of items on this questionnaire is publicly accessible via the Open Science Framework at https://osf.io/pb26g/.

Table 2: Summary of characteristics of the five L1 groups: total number of participants, number of females, mean age at the time of study, mean age upon first arrival in the U.S. (AoAr), and mean years of speaking L2 English (standard deviations in parentheses). NA = not applicable.

L1 group	n total	n female	Age (yr)	AoAr (yr)	Years speaking English
English (NEng)	28	16	21.3 (5.3)	NA	NA
Japanese (NJpn)	28	16	29.6 (8.5)	26.2 (8.2)	12.3 (5.6)
Korean (NKor)	28	16	26.1 (6.5)	19.7 (6.5)	11.9 (5.7)
Mandarin (NMnCh)	28	17	23.3 (2.2)	21.4 (2.2)	10.0 (4.6)
Russian (NRus)	28	16	29.8 (8.9)	25.0 (6.7)	11.6 (7.6)

The L2 English (NJpn, NKor, NMnCh, NRus) groups consisted of late learners of English (age of onset of 7 or later) who had come to the U.S. as young adults, with similarly advanced mean ages of arrival. These groups reported having spoken English for similar lengths of time (10+ years on average), which did not differ significantly [Kruskal-Wallis $\chi^2(3) = 1.362$, n.s.]. The NJpn, NMnCh, and NRus groups consisted of, respectively, native Japanese speakers raised primarily in Japan, native Mandarin speakers born and raised in mainland China or Taiwan, and native Russian speakers born and raised in Russia, Ukraine, or another republic of the former Soviet Union. These groups had no experience with languages containing obligatorily unreleased stops (including Korean and varieties of Chinese with final glottal stops). The NKor group consisted of native Korean speakers who were born and raised primarily in South Korea and had no experience with languages containing unreleased final stops other than Korean and English.

The L1 English (NEng) group consisted of native English speakers who were born and raised in the U.S. in English-speaking households and reported limited knowledge and use of other languages. Eleven NEng participants reported speaking only English, while the other 17 reported being able to speak at least one other language (Farsi, French, Japanese, Mandarin, Russian, and/or Spanish); the latter participants, however, had learned these other languages formally after childhood (mean length of study 5.0 yr) and tended to report low current proficiency, using descriptors such as “not fluent” and “only slight knowledge”. No NEng participants reported fluency in or regular use of another language for communicative purposes. Crucially, like the NJpn, NMnCh, and NRus groups, the NEng group had no experience with languages containing obligatorily unreleased stops.

The NEng and NKor groups each played the role of a control group in the experiment(s) targeting their respective native language. In the English perception experiments, there were four L2 groups familiar with the target language and the NEng group served as an L1 control group, while in the Korean perception experiment, there were four L2 groups unfamiliar with the target language and the NKor group served as an L1 control group. Thus, it should be noted that, unlike NEng listeners, NKor listeners were not “functionally monolingual” L1 listeners (since they knew and used English on a regular basis) and were not tested on an unfamiliar language. This difference between the NEng and NKor groups is unimportant for the current study, however, because the goal is to examine patterns of *relative* performance in each of the two language conditions, as opposed to absolute performance over both language conditions.

2.2. Stimuli

The auditory stimuli were those used in Chang and Mishler (2012) and Chang (2016) and are summarized in Table 3. The stimuli for Experiment 1 consisted of 48 minimal pairs of monosyllabic English words differing in the presence and

Table 3: Korean and English stimuli used in Experiments 1–3. Real words in Experiment 1 are given in English orthography; nonce words in Experiments 2–3 are given in IPA transcription.

Experiment	Stimulus items
1 (English discrimination)	*weep-wheat*, *whip-wit*, *rape-rate*, *cap-cat*, *hoop-hoot*, *taupe-tote*, *pop-pot*, *pup-putt*, *tripe-trite*, *tarp-tart*, *warp-wart*, *kelp-Celt*; *seat-seek*, *sit-sick*, *bait-bake*, *net-neck*, *rat-rack*, *loot-Luke*, *oat-oak*, *cot-cock*, *mutt-muck*, *bite-bike*, *Bart-bark*, *port-pork*; *chic-sheep*, *lick-lip*, *peck-pep*, *wreck-rep*, *tack-tap*, *slack-slap*, *coke-cope*, *soak-soap*, *shock-shop*, *pike-pipe*, *hike-hype*, *hark-harp*; *keep-key*, *type-tie*, *ripe-rye*, *gulp-gull*; *beet-bee*, *suit-sue*, *mart-mar*, *silt-sill*; *peek-pee*, *make-may*, *lake-lay*, *spike-spy*; *ape*, *dupe*, *hop*, *cup*, *quit*, *great*, *tot*, *curt*, *cheek*, *slick*, *lock*, *cork*, *new*, *row*, *four*, *hell*
2 (English identification)	ˈɹʌzip̚, ˈɹʌzit̚, ˈɹʌzik̚, ˈɹʌzi, ˈɹʌzup̚, ˈɹʌzut̚, ˈɹʌzuk̚, ˈɹʌzu, ˈɹʌzɑp̚, ˈɹʌzɑt̚, ˈɹʌzɑk̚, ˈɹʌzɑ, ˈɹʌzeɪp̚, ˈɹʌzeɪt̚, ˈɹʌzeɪk̚, ˈɹʌzeɪ, ˈɹʌzoʊp̚, ˈɹæzoʊt̚, ˈɹʌzoʊk̚, ˈɹʌzoʊ, ˈɹʌzɑɪp̚, ˈɹʌzɑɪt̚, ˈɹʌzɑɪk̚, ˈɹʌzɑɪ, ˈɹʌzɑɹp̚, ˈɹʌzɑɹt̚, ˈɹʌzɑɹk̚, ˈɹʌzɑɹ, ɹəˈzip̚, ɹəˈzit̚, ɹəˈzik̚, ɹəˈzi, ɹəˈzup̚, ɹəˈzut̚, ɹəˈzuk̚, ɹəˈzu, ɹəˈzɑp̚, ɹəˈzɑt̚, ɹəˈzɑk̚, ɹəˈzɑ, ɹəˈzeɪp̚, ɹəˈzeɪt̚, ɹəˈzeɪk̚, ɹəˈzeɪ, ɹəˈzoʊp̚, ɹəˈzoʊt̚, ɹəˈzoʊk̚, ɹəˈzoʊ, ɹəˈzɑɪp̚, ɹəˈzɑɪt̚, ɹəˈzɑɪk̚, ɹəˈzɑɪ, ɹəˈzɑɹp̚, ɹəˈzɑɹt̚, ɹəˈzɑɹk̚, ɹəˈzɑɹ
3 (Korean identification)	mjuɾip̚, mjuɾit̚, mjuɾik̚, mjuɾi, mjuɾup̚, mjuɾut̚, mjuɾuk̚, mjuɾu, mjuɾap̚, mjuɾat̚, mjuɾak̚, mjuɾa, mjuɾɛp̚, mjuɾɛt̚, mjuɾɛk̚, mjuɾɛ, mjuɾop̚, mjuɾot̚, mjuɾok̚, mjuɾo, mjuɾʌp̚, mjuɾʌt̚, mjuɾʌk̚, mjuɾʌ, mjuɾɨp̚, mjuɾɨt̚, mjuɾɨk̚, mjuɾɨ

place of articulation of a final voiceless stop (e.g., *beet*, *bee*; *weep*, *wheat*). The set of words was selected such that most of the English vowels were represented and the two phonological forms in each pair had comparable spoken frequencies (differing by less than an order of magnitude). Spoken frequency estimates were calculated using data from the Corpus of Contemporary American English (Davies, 2008) and took into account all words with the same phonological form (e.g., spoken frequency of the form /bit/ was taken to be the sum total of those for *beet* and *beat*). The /p/-, /t/-, /k/-, and non-stop-final words selected had, respectively, mean spoken frequencies of 23.8, 20.2, 50.5, and 82.0 words per million (wpm) and were distributed roughly equally among low-frequency (< 1 wpm), mid-frequency (1–10 wpm), and high-frequency (> 10 wpm) items.

The stimuli for Experiment 2 consisted of 56 disyllabic English nonce words that varied by final consonant, vowel, and stress. Items followed a $C_1V_1C_2V_2(C_3)$ template (C = consonant, V = vowel) and were made to be identifiably English-like by filling the first two consonant slots with English consonants absent from L2 listeners' L1 inventories: the voiced alveolar approximant /ɹ/ (C_1), which is absent from all of the non-English L1s, and the voiced alveolar fricative /z/ (C_2), which is absent from Mandarin and Korean. The first vowel (V_1) was a mid central unrounded vowel (stressed /ʌ/ or unstressed [ə]), while the second vowel (V_2) ranged over the rhymes /i u ɑ eɪ oʊ ɑɪ ɑɹ/. Point vowels /i u ɑ/ were included because they each have a parallel in the inventories of the other languages, while /eɪ oʊ ɑɪ ɑɹ/ were included because one or more of these is absent from Japanese, Korean, and Russian (which contain either no or a limited set of diphthongs). Finally, the third consonant slot (C_3) varied among /p/, /t/, /k/, and zero (i.e., absence of a final stop). With the alternation of primary stress between initial and final syllables, this resulted in 56 nonce words (7 possible final rhymes x 4 possible codas x 2 possible stress patterns), such as [ɹəˈzit̚] and [ˈɹʌzit̚].

The stimuli for Experiment 3 consisted of 28 disyllabic Korean nonce words that varied by final consonant and vowel. As in Experiment 2, all items followed a $C_1V_1C_2V_2(C_3)$ template. These stimuli were originally constructed with the NEng listeners in mind (to make the perceptual task as easy as possible for them; see Chang, 2016), so they were made to depart as little as possible from English phonology (while remaining consistent with Korean phonology) by filling the first two consonant slots with Korean consonants that also occur in English: the voiced bilabial nasal /m/ (C_1), which occurs in all the other languages, and the voiced alveolar flap [ɾ] (C_2), an allophone of /l/ which also occurs in Japanese and Russian. The first vowel (V_1) was a high back rounded vowel with a palatal on-glide ([ju]), while the second vowel (V_2) ranged over the seven-vowel inventory of modern Korean: /i u a ɛ o ʌ ɨ/ (Chang, 2012). As in Experiment 2, the final consonant slot (C_3) varied among /p/, /t/, /k/, and zero. This resulted in 28 nonce words (7 possible final vowels x 4 possible codas), such as [mjuɾat̚].

Creation of the auditory stimuli was performed in two steps. In the first step, the target items were produced by native speakers and audio-recorded. The English stimuli were recorded by two male native speakers of American English (age 19 and 25 yr), who were raised in Maryland and had no experience with a language containing obligatorily unreleased stops. The Korean stimuli were recorded by a male native speaker of Korean (age 32 yr) born and raised in Seoul. All recordings were made in the U.S. in a sound-attenuated booth at 44.1 kHz with 24-bit resolution, using a Zoom H4N mobile audio recorder and an Audix HT5 head-mounted condenser microphone positioned approximately 2 cm to the left of the talker's mouth. Items for Experiments 1–2 were presented via English spelling (with the stressed syllable underlined for the nonce items), and items for Experiment 3 via Korean spelling, on randomized individual index cards three times. To regulate the rate of presentation, a Qwik Time QT-3 metronome was used to present items at a rate of approximately one every two seconds.

In the second step, speech tokens were selected containing the coarticulatory transitions of interest from among the three repetitions of each stimulus. Although both released and unreleased blocks of tokens were collected of the English items, released tokens ultimately provided the basis for the English stimuli (in both English perception experiments) because the presence of a release burst made it clear that the oral closure of the final stop consonant was realized (whereas unreleased tokens were sometimes realized with just a glottal stop). Additionally, previous research comparing the perception of unreleased stops and "dereleased" stops (i.e., released stops with the release burst removed) in English found the two to be very similar (Lisker, 1999; Malécot, 1958). Thus, to approximate unreleased stops in the English stimuli while ensuring the presence of VC formant transitions, released tokens were used and edited in Praat (Boersma & Weenink, 2011) to remove the final release burst. The Korean tokens were produced as unreleased, so they did not undergo editing to remove a release burst. Both English and Korean stimuli were furthermore normalized in Praat to a peak intensity of 0.99.

To check that the nonce word stimuli actually contained the variation in vowel duration that serves as a cue to the presence of a coda consonant, the duration of the final vowel in each of the 140 stimuli for Experiments 2–3 was measured in Praat via visual inspection of a wide-band spectrogram, by marking vowel onset and offset, respectively, at the first point and last point where all of the first three formants (F_1, F_2, F_3) were clearly visible. These acoustic data showed that the nonce word stimuli did in fact contain the expected durational variation. Final vowels in English stop-final stimuli were significantly shorter than those in English non-stop-final stimuli, both for the first talker [$M_{stop.final} = 162$ ms, $M_{non.stop.final} = 238$ ms; Welch-corrected two-sample $t(15.4) = -5.533, p <$ .0001] and for the second talker [$M_{stop.final} = 148$ ms, $M_{non.stop.final} = 272$ ms; Welch-corrected two-sample $t(14.2) = -6.187, p <$ .0001]. The same pattern held for the Korean stimuli [$M_{stop.final} = 117$ ms, $M_{non.stop.final} = 193$ ms; Welch-corrected two-sample $t(8.1) = -6.355, p <$ .001].

2.3. Procedure

All listeners were tested in a quiet room at an American university. In all, they completed three experiments in a single session, in numerical order with intervening breaks. The tasks were first explained (in listeners' L1, with the exception of the NMnCh and NRus groups due to the lack of Mandarin- and Russian-speaking experimenters), and listeners were then specifically instructed to listen carefully to the stimuli and to respond as quickly and accurately as possible. Stimuli were presented on a computer running E-Prime 2.0 using high-quality binaural headphones, and listeners entered their responses on a Psychology Software Tools Model 200A serial response box connected to the computer.

Since the goal of all three experiments was to examine language transfer in speech perception while abstracting away from effects of semantic context, most of the design features were meant to encourage listeners to process the stimuli at a phonological (i.e., not merely psychoacoustic) level, with minimal top-down influence. The English experiments were focused on listeners' phonologically informed perception as L2 users, either with (Experiment 1) or without (Experiment 2) the aid of long-term phonological representations associated with lexical items. Thus, the default experimental paradigm used was sound identification in non-words, a metalinguistic task that forces listeners to think about phonological categories, and this was the task used in Experiment 2 and the Korean experiment (Experiment 3). On the other hand, because lexical frequency was not relevant for the research questions (and, in fact, presented a potential source of interference which could obscure between-group differences in L1 transfer), the English experiment with lexical stimuli (Experiment 1) used the discrimination paradigm with frequency-balanced word pairs to avoid unintended effects of lexical frequencies; however, a long inter-stimulus interval (ISI) as well as talker variability were used to encourage discrimination at a phonological level (see, e.g., Flege, 2003).

In Experiment 1, listeners completed a speeded AX categorial discrimination task (Flege, 2003) with English words ("speeded" refers to the instructions to listeners to respond both accurately and as quickly as possible). Words in each pair were uttered by different talkers, each trial consisting of the presentation of a trial counter on screen for 1 sec, the playing of the first word (A), a 1-sec ISI, and then the playing of the second word (X). A listener's response indicated whether X was the same word as A or a different word. The experiment began with 12 practice trials and moved on to 192 test trials (96 "same" trials and 96 "different" trials), which were divided into two randomized blocks with an even distribution of "same" and "different" trials spanning both possible talker orders.

In Experiment 2, listeners completed a speeded one-interval, four-alternative forced choice (4AFC) identification task with English nonce words. To increase the difficulty of this task (since all listeners were familiar with English) and thereby lower the likelihood of ceiling performance (which would have the undesirable effect of obscuring between-group differences), the task incorporated sentence embedding as well as alternation between different talkers. On each trial, a trial counter was presented on screen for 1 sec and then a randomly selected precursor was played (either *This word is...*, *Now the word is...*, or *The next word is...*), followed by one of the 56 nonce words. A listener's response indicated whether the final sound of the last word was /p/, /t/, /k/, or something else ("other"). The experiment began with eight practice trials and moved on to three randomized blocks of 56 test trials. In the first block, trials were spoken by the first talker; in the second block, by the second talker; and in the final block, by either talker.

In Experiment 3, listeners completed a similar 4AFC identification task with Korean nonce words. Since all listeners except the NKor listeners were unfamiliar with Korean, these stimuli were presented in isolation and uttered by one talker only (i.e., features increasing difficulty in Experiment 2 were not incorporated here). Thus, absolute levels of performance in Experiment 3 are not directly comparable to those in Experiment 2; however, this is not a problem because the crucial variable in all experiments is not absolute performance, but *relative* performance (compared to other groups). The structure of each trial in Experiment 3 was similar to that of trials in Experiment 2, consisting of the presentation of a trial counter on screen for 1 sec and then the playing of one of the 28 nonce words. As in Experiment 2, a listener's response indicated whether the final sound of the word was /p/, /t/, /k/, or something else ("other"). The experiment began with eight practice trials and moved on to three randomized blocks of 28 test trials.

3. Results

3.1. Experiment 1: Stop discrimination in English

The data from Experiment 1 were analyzed in terms of d', a unitless measure of perceptual sensitivity to stimulus changes (i.e., discrimination ability) that accounts for response bias (Macmillan & Creelman, 2005).[8] A higher d' is interpreted as reflecting more successful perception. For each participant, two d' scores were calculated: one for discrimination of "stop/stop" contrasts (i.e., word pairs differing in the place of a final stop, such as *weep* vs. *wheat*), and one for discrimination of "stop/zero" contrasts (i.e., word pairs differing in the presence of a final stop, such as *beet* vs. *bee*). For the first d' score, "hits" and "false alarms" were, respectively, correct responses on "different" stop/stop trials (e.g., *weep*/*wheat*) and incorrect responses on "same" stop/stop trials (e.g., *weep*/*weep*). For the second d' score, "hits" and "false alarms" were, respectively, correct responses on "different" stop/zero trials (e.g., *beet*/*bee*) and incorrect responses on "same" stop/stop trials (e.g., *beet*/*beet*) and zero/zero trials (e.g., *bee*/*bee*).

Inspection of the d' scores using the Shapiro-Wilk test of normality (Shapiro & Wilk, 1965) suggested that although nine out of the ten sets of scores (from 5 listener groups x 2 contrast types) were normally distributed [$W > 0.956, p > .290$], the NJpn group's scores on stop/stop contrasts were not [$W = 0.922, p = .039$]; therefore, non-parametric statistics (namely, the Kruskal-Wallis one-way analysis of variance; Kruskal & Wallis, 1952) were used in R (R Development Core Team, 2015) to test for between-group differences in discrimination performance.

---

[8] All data from Experiments 1–3 (in trial-by-trial format) are publicly accessible via the Open Science Framework at https://osf.io/e5qsj/.

There were two factors: Group (NEng, NJpn, NKor, NMnCh, NRus), a between-participants factor, and Contrast (stop/stop, stop/zero), a within-participants factor. Additional pairwise tests comprised only the four planned comparisons between adjacent groups on the predicted cline of perceptual success for each contrast (as opposed to all 20 comparisons); therefore, multiple-comparisons correction of p-values was not performed to avoid increasing the chance of type II error.

Given P1 as well as the L1 status of the NEng group, the predicted cline of success for stop/stop discrimination was NJpn < NMnCh < NRus < NKor < NEng, while that for stop/zero discrimination was {NJpn, NMnCh} < NRus < NKor < NEng. Figure 1 shows the marked differences in d' scores that emerged among the four L2 English groups in comparison to the NEng group for both contrast types, which resulted in a main effect of Group [Kruskal-Wallis $\chi^2(4) = 66.267, p < .0001$]. A main effect of Contrast [Kruskal-Wallis $\chi^2(1) = 87.565, p < .0001$] arose due to the fact that stop/zero contrasts (mean $d' = 1.57$) were discriminated better than stop/stop contrasts (mean $d' = 0.84$) by all groups. When the data were further examined by contrast type, a significant effect of Group was found both for stop/stop contrasts [Kruskal-Wallis $\chi^2(4) = 71.409, p < .0001$] and for stop/zero contrasts [Kruskal-Wallis $\chi^2(4) = 49.459, p < .0001$].

Since there were significant effects of Group on d' scores for both contrast types, between-group comparisons were conducted for both contrast types to identify the source of these effects. On stop/stop contrasts, NJpn listeners had the lowest d' scores (mean of 0.44), followed by NMnCh listeners (mean of 0.56), NRus listeners (mean of 0.81), NKor listeners (mean of 1.17), and NEng listeners (mean of 1.24). This hierarchy was as predicted, although pairwise comparisons revealed that the NJpn-NMnCh and NKor-NEng differences were not sig-

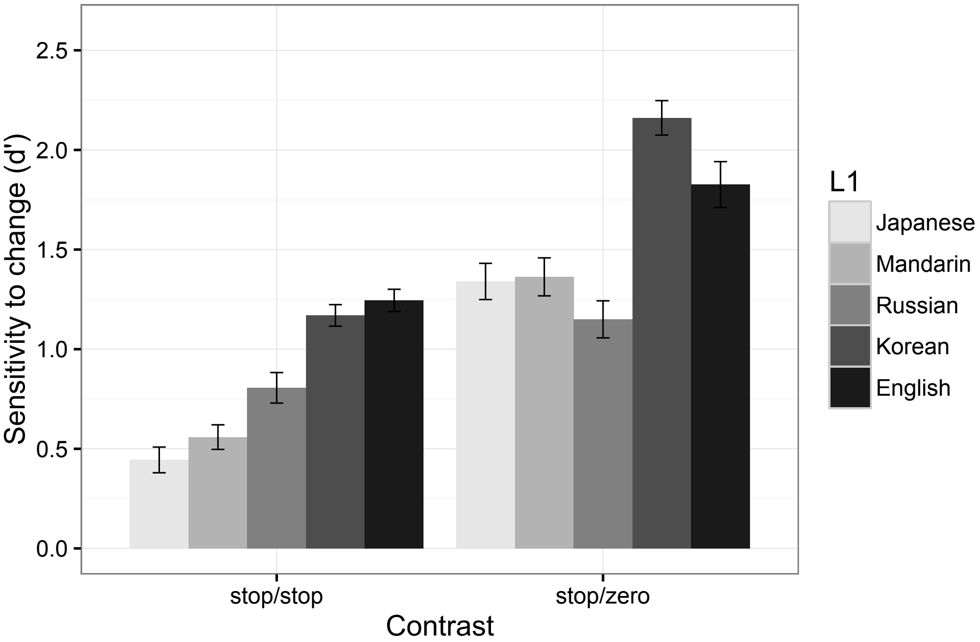

Figure 1: Perceptual sensitivity (d') in Experiment 1 (English discrimination), by contrast type and L1 group. "Stop/stop" and "stop/zero" refer to minimal pairs differing in final stop (e.g., *weep*, *wheat*) or presence of a final stop (e.g., *beet*, *bee*), respectively. Chance performance (50% correct overall) corresponds to a d' of 0. Error bars mark ±1 standard error of the mean over participants.

nificant [Kruskal-Wallis $\chi^2(1) < 1.170$, n.s.]. However, the NMnCh-NRus difference [Kruskal-Wallis $\chi^2(1) = 5.339, p < .05$] and the NRus-NKor difference [Kruskal-Wallis $\chi^2(1) = 11.304, p < .001$] were both significant. In short, d' scores on stop/stop contrasts showed the following hierarchy of perceptual sensitivity: {NJpn, NMnCh} < NRus < {NKor, NEng}. Overall, these results are more consistent with the cue-centric view (which predicts the difference between NRus and NKor/NEng) than the direct phonotactic view (which predicts only the difference between NJpn/NMnCh and NRus/NKor/NEng).

On stop/zero contrasts, the five groups showed a different relative ordering of d' scores. The group with the lowest d' scores here was the NRus group (mean of 1.15), followed by the NJpn group (mean of 1.34), the NMnCh group (mean of 1.36), the NEng group (mean of 1.83), and the NKor group (mean of 2.16). The fact that NRus listeners' d' scores here were lower, instead of higher, than NMnCh listeners' was unexpected, although pairwise comparisons revealed that neither the NMnCh-NRus difference nor the NJpn-NMnCh difference was significant [Kruskal-Wallis $\chi^2(1) < 2.274$, n.s.]. The NRus-Kor difference [Kruskal-Wallis $\chi^2(1) = 32.148, p < .001$] was significant and in the expected direction, whereas the NKor-NEng difference [Kruskal-Wallis $\chi^2(1) = 4.401, p < .05$] was significant and in the opposite direction of the prediction. Thus, d' scores on stop/zero contrasts showed the following hierarchy of perceptual sensitivity: {NJpn, NMnCh, NRus} < NEng < NKor. Overall, these results are also more consistent with the cue-centric view than the direct phonotactic view: the cue-centric view both predicts the failure of NMnCh listeners to outperform NJpn listeners and is able to account for the better-than-native perception of NKor listeners, whereas the direct phonotactic view incorrectly predicts a NMnCh advantage over the NJpn group

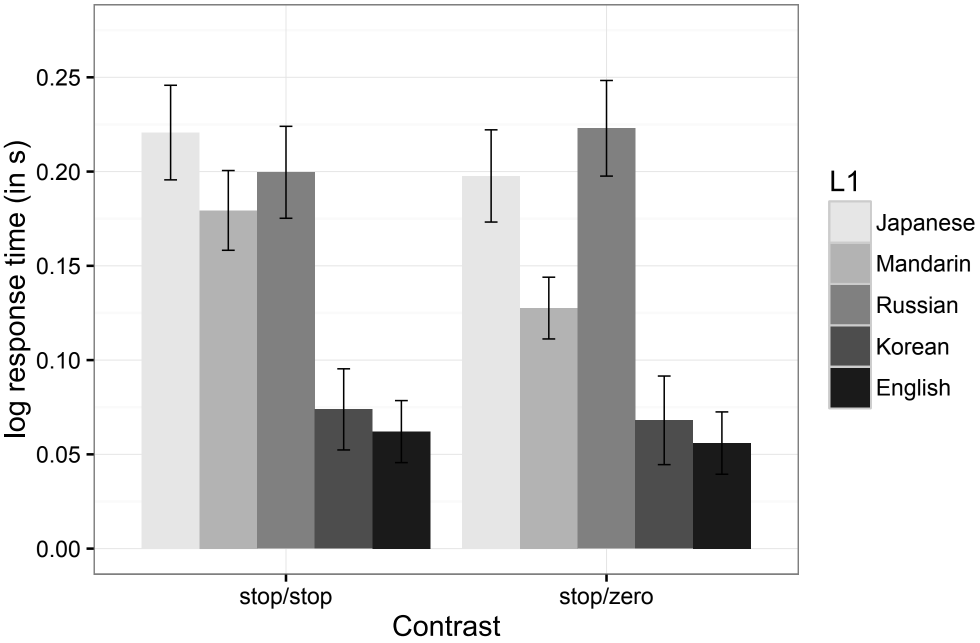

Figure 2: Log response time for correct "different" responses in Experiment 1 (English discrimination), by contrast type and L1 group. "Stop/stop" and "stop/zero" refer to minimal pairs differing in final stop (e.g., *weep*, *wheat*) or presence of a final stop (e.g., *beet*, *bee*), respectively. Error bars mark ±1 standard error of the mean over participants.

and is unable to explain the NKor advantage over the NEng group.[9]

To check whether the group differences in d' scores could be accounted for in terms of a speed-accuracy trade-off (e.g., d' scores in one group being low because of a higher error rate arising from faster responses), response times (RTs) were also examined, following exclusion of extreme RTs greater than 2.5 standard deviations from each participant's mean (6% of the data; see, e.g., Sumner & Samuel, 2009) and log transformation to correct for positive skew (Newell &

---

[9] Note that the lack of difference between NMnCh and NRus is not predicted under either view.

Rosenbloom, 1981). Figure 2 shows the average log RTs for correct discrimination judgments across groups and contrast types. There was no effect of Contrast on RTs [Kruskal-Wallis $\chi^2(1) = 0.640$, n.s.], but a significant effect of Group [Kruskal-Wallis $\chi^2(4) = 70.893, p < .0001$], reflecting the overall similarity of RTs across the two contrast types and the substantial variation of RTs across groups. Crucially, however, the pattern of RT differences provided no indication that differences in d' were attributable to differences in RTs. On the contrary, groups that achieved higher d' scores consistently did so with RTs that were either not significantly different from, or in fact faster than, RTs of groups with lower d' scores (e.g., NKor/NEng vs. NJpn/NMnCh/NRus, on both contrast types).

3.2. Experiment 2: Stop identification in English

The data from Experiment 2 were analyzed by building a logistic mixed-effects regression model of the log odds of correct identification (Dixon, 2008; Jaeger, 2008) in R (R Development Core Team, 2015). Higher odds of correct identification are interpreted as reflecting more successful perception. Starting with random-effect terms for Participant and Item, the model was augmented incrementally by fixed-effect terms for Final (stop, non-stop; reference level = stop), Group (NEng, NJpn, NKor, NMnCh, NRus; reference level = NEng), and a Final x Group interaction. All variables were treatment-coded, and the reference level of the Group variable was set to contrast each of the L2 English groups with the L1 English (i.e., NEng) group. The basic model with only random intercepts by Participant and by Item was improved by adding the Final term [$\chi^2(1) = 132.130, p < .0001$], the Group term [$\chi^2(4) = 56.310, p < .0001$], and the Final x Group interaction [$\chi^2(4) = 176.940, p < .0001$]. Thus, the final model of English identification performance [$n = 23520$, log-likelihood $= -10892$] in-

Table 4: Fixed-effect terms in the logistic mixed-effects model of the likelihood of accuracy in Experiment 2 (English identification). Significance codes: * $p < .05$, ** $p < .01$, *** $p < .001$.

Predictor	β	SE	z	p	
(Intercept)	0.364	0.177	2.061	.039	*
Final: non-stop	3.769	0.300	12.581	$<.001$	***
Group: NJpn	-1.148	0.201	-5.700	$<.001$	***
Group: NMnCh	-0.398	0.201	-1.983	.047	*
Group: NRus	-1.009	0.202	-5.006	$<.001$	***
Group: NKor	0.247	0.201	1.226	.220	
Final: non-stop x Group: NJpn	-0.097	0.247	-0.394	.694	
Final: non-stop x Group: NMnCh	-1.490	0.236	-6.320	$<.001$	***
Final: non-stop x Group: NRus	0.954	0.284	3.355	$<.001$	***
Final: non-stop x Group: NKor	-0.741	0.275	-2.693	.007	**

cluded all three fixed effects, summarized in Table 4.[10]

As in Experiment 1, the predicted cline of success for stop identification was NJpn < NMnCh < NRus < NKor < NEng, while that for non-stop identification was {NJpn, NMnCh} < NRus < NKor < NEng. Figure 3 shows the considerable

---

[10] Note that all of the final models for Experiments 2–3 contained a parsimonious random-effects structure including only random intercepts (as opposed to the maximal random-effects structure with all possible random intercepts and slopes) because attempts to build models with more complex random-effects structures either failed to converge or yielded models that showed signs of overparameterization and/or less stable fit, consistent with concerns in the literature regarding maximal models for actual psycholinguistic data (e.g., Bates et al., 2015). More complex models, moreover, did not generate results for the fixed effects that were substantially different from those of parsimonious models. Therefore, the results reported below are from the parsimonious models.

cross-group variation that was found in this experiment. Model results (Table 4) revealed that NEng listeners accurately identified English final stops with higher than 50-50 odds [$\beta = 0.364, z = 2.061, p < .05$]; however, they were much more likely to identify final non-stops (as "other" sounds) accurately [$\beta = 3.769, z = 12.581, p < .0001$], and this was the case for all groups. Consistent with the results of Experiment 1, NJpn listeners had the lowest accuracy of all groups on final stops, and the NJpn group, as well as the MnCh and NRus groups, were all significantly less likely than NEng listeners to identify final stops accurately [βs < -0.397, zs < -1.982, ps $< .05$]; NKor listeners, by contrast, did not differ significantly from NEng listeners [$\beta = 0.247, z = 1.226$, n.s.].

To test additional group comparisons that were not evident in the main model in Table 4, alternative models were built with the same overall structure but with different reference levels for Group and/or Contrast. A model with NJpn set as the reference level of Group showed that NMnCh listeners were significantly more likely to be accurate on final stops than NJpn listeners [$\beta = 0.772, z = 3.187, p < .01$], whereas NRus listeners were not [$\beta = 0.143, z = 0.589$, n.s.]. A second model with NRus set as the reference level of Group showed that NMnCh listeners were more likely to be accurate on final stops than NRus listeners as well [$\beta = 0.661, z = 2.511, p < .05$]. In short, results on final stops showed the following cline of perceptual success: {NJpn, NRus} < NMnCh < {NKor, NEng}.

As for final non-stops (i.e., sonorants), all groups found these relatively easy to identify accurately as "other" sounds and showed near-ceiling performance on these stimuli. Nevertheless, a model with 'non-stop' set as the reference level of Contrast revealed that the NJpn and NMnCh groups were both significantly less likely to be accurate on final non-stops than the NEng group [βs < -1.820, zs

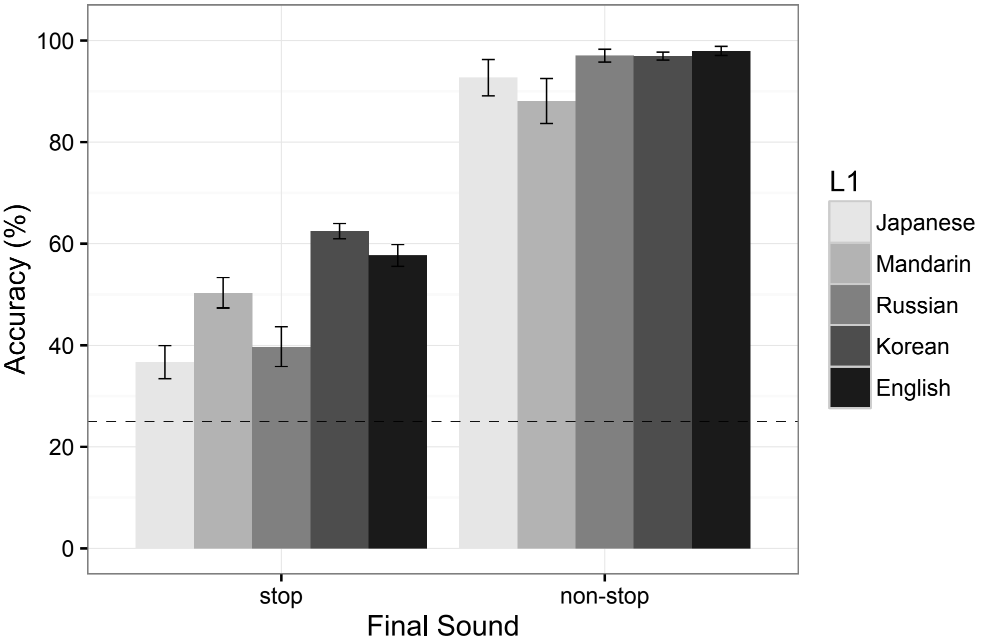

Figure 3: Percent accuracy in Experiment 2 (English identification), by final sound type and L1 group. “Stop” and “non-stop” refer, respectively, to final unreleased stops and to final non-stops (correctly identified as the “other” category, i.e. not /p t k/). Error bars mark ±1 standard error of the mean over participants. The dotted line marks the level of chance performance.

$< -2.083, p\text{s} < .05$]. However, the NJpn and MnCh groups were not significantly different from each other on final non-stops, as shown in a second model with 'non-stop' as the reference level of Contrast and NJpn as the reference level of Group [$\beta = 0.573$, $z = 0.628$, n.s.]. In short, results on final non-stops showed the following cline of perceptual success: {NJpn, NMnCh} < {NRus, NKor, NEng}.

Notably, the observed differences between groups on final stops were relatively consistent across vowel contexts. When the analysis considered only those items where the second (final) vowel was one of the point vowels, the overall pattern of results was found to remain the same. In other words, reducing the crosslinguistic disparity between the vowels in the L2 target items and the vowels of the various L2 listeners' L1s did not significantly change the results, suggesting that the overall pattern of between-group differences (on final stops especially) was not due to differences in crosslinguistic similarity of vowels.

Accuracy on final stops, however, showed considerable variation according to place, largely attributable to the diverse response biases evident in listeners' errors (Figure 4). Although NKor listeners showed relatively little bias, NEng listeners, as described in prior work (Chang, 2016; Chang & Mishler, 2012), were biased to respond "t" for stop-final stimuli. This bias was consistent with the fact that /t/ is the stop most likely to occur without release in American English, and was also found in all groups' errors on non-stop-final stimuli (although less so for the NMnCh group). Unlike NEng listeners, NJpn and NRus listeners were both heavily biased to respond "other" for stop-final stimuli, which may indicate that to their ears these stimuli did not sound like they ended in a stop; this would be consistent with the strong implication of release for stops in Japanese and Russian. The bias toward "other" was evident in NMnCh listeners, too, but less strongly, as

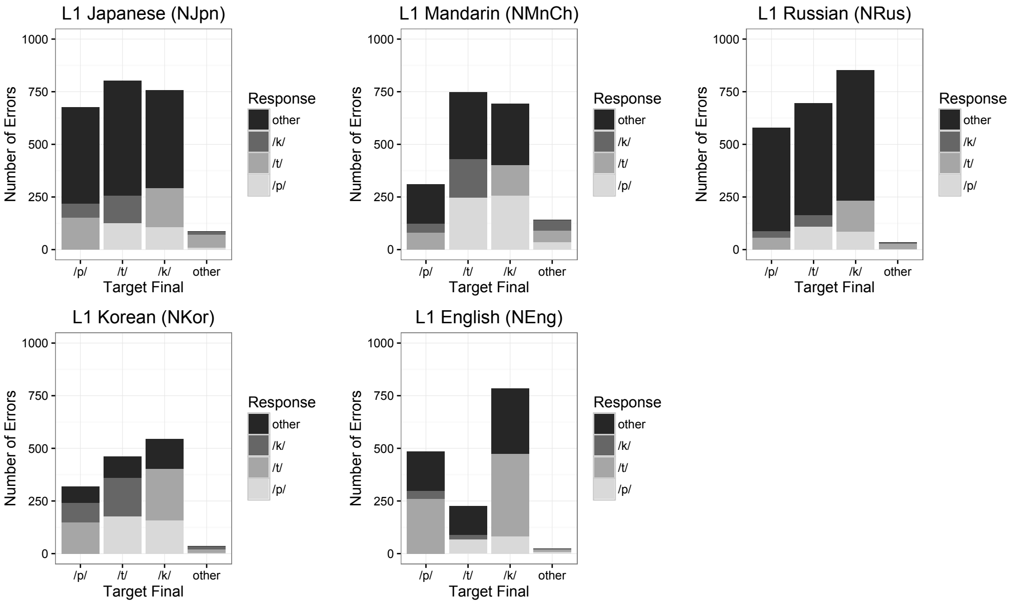

Figure 4: Total error counts in Experiment 2 (English identification), by group, target, and response. The groups are the five L1 groups; the targets and responses correspond to the four answer choices (/p/, /t/, /k/, “other”). For each target, error types are presented in order from bottom to top, shaded progressively darker according to response (with incorrect /p/ responses at the bottom in the lightest gray and incorrect “other” responses at the top in black).

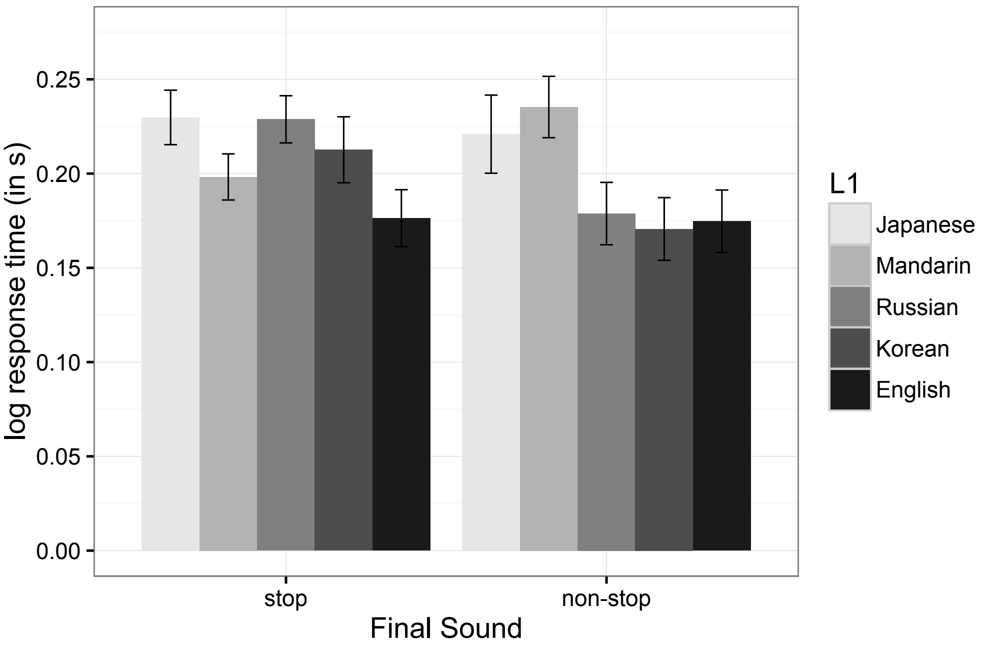

Figure 5: Log response time for correct responses in Experiment 2 (English identification), by contrast type and L1 group. “Stop” and “non-stop” refer, respectively, to final unreleased stops and to final non-stops (correctly identified as the “other” category, i.e. not /p t k/). Error bars mark ±1 standard error of the mean over participants.

they were also inclined to respond “p” for stop-final stimuli.

As in Experiment 1, RTs in Experiment 2 were examined to check whether group differences in identification accuracy could be attributed to differences in response speed. The average log-transformed RTs for correct identification judgments are shown in Figure 5 (excluding extreme data points greater than 2.5 standard deviations from each participant’s mean, which comprised 9% of the data). There was no effect of Contrast on RTs [Kruskal-Wallis $\chi^2(1) = 2.348$, n.s.], but a significant effect of Group [Kruskal-Wallis $\chi^2(4) = 12.659, p < .05$]. Again, however, the specific pattern of group differences in RTs only supported the accuracy results: groups that achieved higher accuracy showed RTs that were either not significantly different from, or faster than, the RTs of groups that achieved lower accuracy (e.g., NEng vs. NJpn/NMnCh on non-stop-final stimuli).

3.3. Experiment 3: Stop identification in Korean

The data from Experiment 3 were subjected to the same analysis as the data from Experiment 2: logistic mixed-effects regression on the log odds of correct identification. As in Experiment 2, higher odds of accuracy are interpreted as reflecting more successful perception. The model-building procedure was the same, starting with random-effect terms for Participant and Item and augmenting the model incrementally with fixed-effect terms for Final (stop, non-stop; reference level = stop), Group (NEng, NJpn, NKor, NMnCh, NRus; reference level = NKor), and a Final x Group interaction. As in Experiment 2, all variables were treatment-coded, and the reference level of the Group variable was set to contrast each of the groups unfamiliar with the target language (Korean) with the L1 Korean (i.e., NKor) group. The basic model with only random intercepts by Participant and by Item was improved by adding the Final term [$\chi^2(1) = 24.080, p < .0001$], the Group term [$\chi^2(4) = 81.941, p < .0001$], and the Final x Group interaction [$\chi^2(4) = 118.200, p < .0001$], so the final model [$n = 11760$, log-likelihood $= -5114$] included all of these fixed effects, summarized in Table 5.

Given the L1 status of the NKor group in Experiment 3, the predicted cline of success for stop identification was NJpn < NMnCh < NRus < NEng < NKor, while that for non-stop identification was {NJpn, NMnCh} < NRus < NEng < NKor. Figure 6 shows that there was variation among the groups in their identification performance in Korean, too. Model results (Table 5) revealed that NKor listeners were highly likely to identify Korean final stops accurately [$\beta = 2.122, z =$

Table 5: Fixed-effect terms in the logistic mixed-effects model of the likelihood of accuracy in Experiment 3 (Korean identification). Significance codes: * $p < .05$, ** $p < .01$, *** $p < .001$.

Predictor	β	SE	z	p	
(Intercept)	2.122	0.247	8.580	$< .001$	***
Final: non-stop	2.081	0.450	4.629	$< .001$	***
Group: NJpn	-1.992	0.255	-7.825	$< .001$	***
Group: NMnCh	-1.995	0.254	-7.845	$< .001$	***
Group: NRus	-0.552	0.257	-2.147	.032	*
Group: NEng	-1.318	0.255	-5.176	$< .001$	***
Final: non-stop x Group: NJpn	1.301	0.371	3.511	$< .001$	***
Final: non-stop x Group: NMnCh	-0.886	0.325	-2.722	.006	**
Final: non-stop x Group: NRus	1.330	0.484	2.749	.006	**
Final: non-stop x Group: NEng	-0.076	0.348	-0.217	.828	

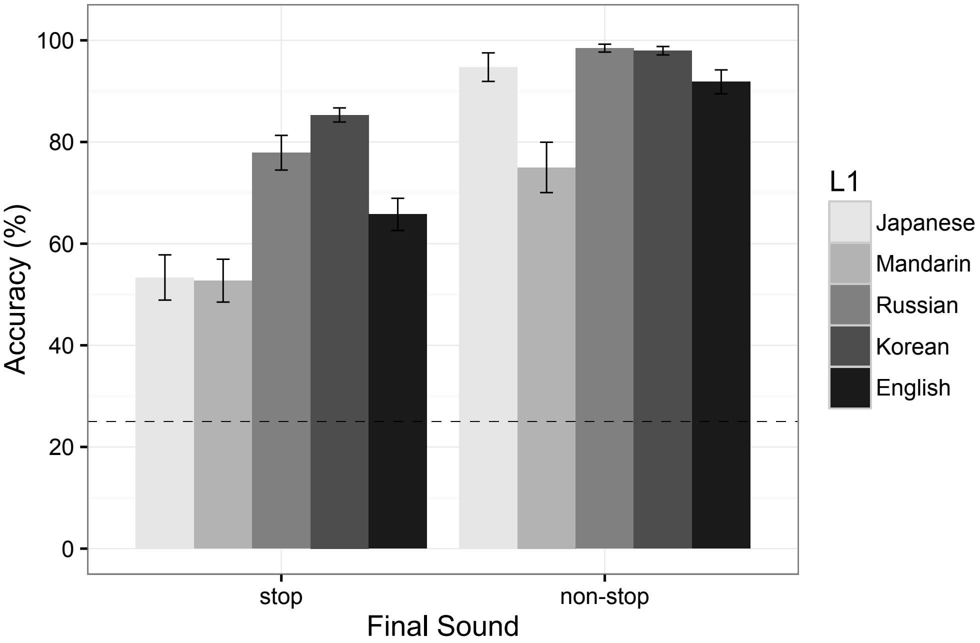

Figure 6: Percent accuracy in Experiment 3 (Korean identification), by final sound type and L1 group. "Stop" and "non-stop" refer, respectively, to final unreleased stops and to final non-stops (correctly identified as the "other" category, i.e. not /p t k/). Error bars mark ±1 standard error of the mean over participants. The dotted line marks the level of chance performance.

$8.580, p < .0001$]; however, they were still more likely to identify final non-stops (as "other" sounds) accurately [$\beta = 2.081, z = 4.629, p < .0001$], and this was true of all groups. In comparison to the NKor group, all other groups were significantly less likely to identify final stops accurately [βs < 0.551, zs < 2.146, ps $< .05$], but there were further differences among them.

To test additional group comparisons that were not evident in the main model in Table 5, alternative models were built with the same structure but different reference levels for Group and/or Contrast. NJpn and NMnCh listeners were the least likely to be accurate on final stops (showing nearly identical levels of accuracy, contrary to P1), and a model with NJpn set as the reference level of Group showed that NRus and NEng listeners were both significantly more likely to be accurate on final stops than the NJpn group [βs > 0.674, zs > 2.573, ps $< .05$]. Alternative models with NMnCh or NEng set as the reference level of Group further showed that NEng and NRus listeners were both significantly more likely to be accurate on final stops than NMnCh listeners [$\beta\text{s} > 0.675, z\text{s} > 2.585, p\text{s} < .01$], and that NRus listeners were more likely to be accurate on final stops than NEng listeners [$\beta = 0.768, z = 2.900, p < .01$]. Thus, results on final stops showed the following cline of perceptual success: {NJpn, NMnCh} < NEng < NRus < NKor.

Similar to Experiment 2, most groups found final non-stops (i.e., sonorants) in Korean relatively easy to identify accurately as "other" sounds. The exception was the NMnCh group, which failed to reach 80% accuracy on non-stop-final stimuli. A model with 'non-stop' set as the reference level of Contrast showed that the NMnCh group was significantly less likely to be accurate on final non-stops than the NKor group [$\beta = -3.743, z = 5.164, p < .0001$], as was the NEng group [$\beta = -2.078, z = 2.802, p < .01$]; however, the NJpn and NRus groups were not significantly different from the NKor group [$|\beta| < 0.785, |z| < 0.986$, n.s.]. A second model with 'non-stop' as the reference level of Contrast and NMnCh as the reference level of Group showed that the NEng group was still more likely to be accurate on final non-stops than the NMnCh group [$\beta = 1.667, z = 2.812, p < .01$], while a third model with 'non-stop' as the reference level of Contrast and NJpn as the reference level of Group showed that the NEng group was marginally less likely to be accurate on final non-stops than the NJpn group [$\beta = -1.283, z = -1.884, p = .059$]. In short, results on final non-stops showed the following cline of perceptual success: NMnCh < NEng < {NJpn, NRus, NKor}.

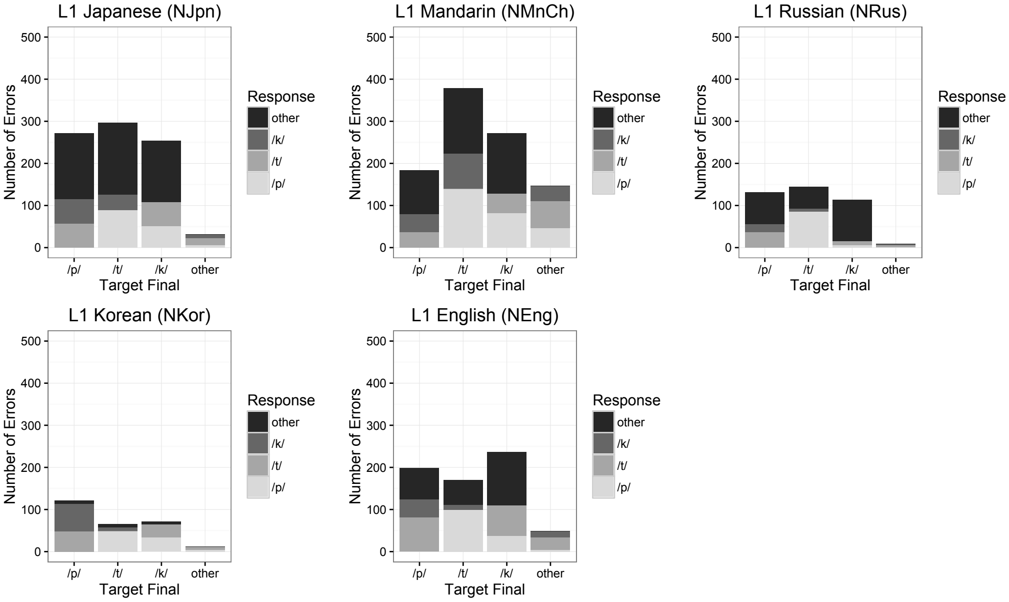

Figure 7: Total error counts in Experiment 3 (Korean identification), by group, target, and response. The groups are the five L1 groups; the targets and responses correspond to the four answer choices (/p/, /t/, /k/, "other"). For each target, error types are presented in order from bottom to top, colored progressively darker according to response (with incorrect /p/ responses at the bottom in the lightest gray and incorrect "other" responses at the top in black).

As in Experiment 2, the overall patterns in Experiment 3 remained the same when results were limited to items with a final point vowel; however, again there was considerable variation on final stops according to place due to different response biases across groups. Error analyses (Figure 7) showed no clear bias for NKor listeners other than toward "p" errors on final /t/. NEng listeners again showed a bias to respond "t", but less strongly here, as their most common error on final /k/ was instead to respond "other". NJpn and NMnCh listeners were again biased to respond "other" for stop-final stimuli (although somewhat less strongly than in Experiment 2). NRus listeners, too, were biased to err by responding "other", except in the case of final /t/, where their most common error was to respond "p". What was most striking about NRus listeners' errors here, however, was their rarity, which resulted in the NRus group being 38% more accurate on the Korean stops in Experiment 3 than on the English stops in Experiment 2.

In Experiment 3 as well, RTs were examined to check whether group differences in accuracy could be attributed to differences in response speed. The average log-transformed RTs for correct identification judgments are shown in Figure 8 (excluding extreme data points greater than 2.5 standard deviations from each participant's mean, which comprised 7% of the data). There was no effect of Contrast on RTs [Kruskal-Wallis $\chi^2(1) = 0.312$, n.s.], but a significant effect of Group [Kruskal-Wallis $\chi^2(4) = 35.626, p < .0001$]. As in Experiment 2, however, the specific pattern of group differences in RTs strengthened the findings on accuracy: groups that achieved higher accuracy showed RTs that were either not significantly different from, or faster than, the RTs of groups that achieved lower accuracy (e.g., NRus/NKor/NEng vs. NJpn/NMnCh on stop-final stimuli).

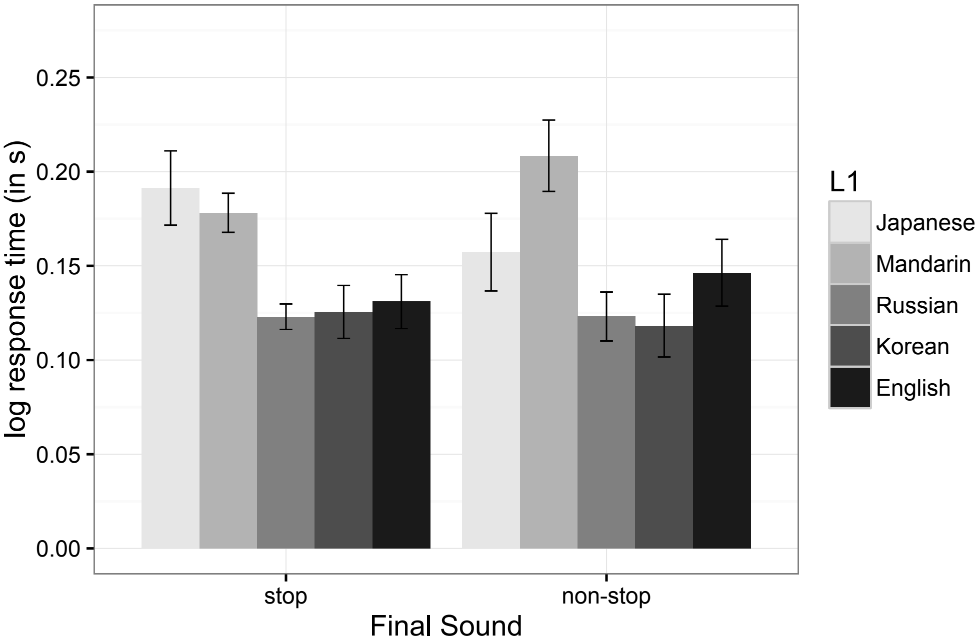

Figure 8: Log response time for correct responses in Experiment 3 (Korean identification), by contrast type and L1 group. “Stop” and “non-stop” refer, respectively, to final unreleased stops and to final non-stops (correctly identified as the “other” category, i.e. not /p t k/). Error bars mark ±1 standard error of the mean over participants.

4. Discussion

In regard to Q1 in Section 1.3, the results of Experiments 1–3 (summarized in Table 6) provided more support for the cue-centric view than the direct phonotactic view of L1 transfer in L2 speech perception. In the discrimination of unreleased stop contrasts in English (Experiment 1), the patterning of L2 listener groups was consistent with the predicted cline of perceptual success (i.e., P1: NJpn < NMnCh < NRus < NKor), although the difference between NJpn and NMnCh listeners did not reach significance. In the discrimination of the presence vs. absence of an unreleased stop, the NKor group displayed greater sensitivity than the L1 listener (NEng) group as well. In the identification of unreleased stops in English (Experiment 2), there was a similar cline of perceptual success, except that the NRus group underperformed the NMnCh group. The NRus group also showed an unexpected pattern of performance in the identification of unreleased stops in Korean (Experiment 3), where they diverged from the predicted cline of perceptual success by outperforming the NEng group.

Although one aspect of the results, the failure of NMnCh listeners to outperform NJpn listeners on final stops in Experiment 3, contradicts P1 from the cue-centric view, there are three aspects of the results that cannot be explained under the direct phonotactic view: (1) the advantageous transfer (i.e., native-language transfer benefit) evident in NKor listeners' better-than-native sensitivity to stop/zero contrasts in English, (2) NMnCh listeners' advantage over NJpn listeners in identification of unreleased stops in English, which supports P1, and (3) NJpn listeners' advantage over NMnCh listeners in identification of the absence of a final voiceless stop in Korean, which supports P2. This is because the direct phonotactic view provides no way of deriving native-language transfer benefits

Table 6: Summary of results in Experiments 1–3. NJpn = L1 Japanese; NMnCh = L1 Mandarin Chinese; NRus = L1 Russian; NKor = L1 Korean; NEng = L1 American English.

Experiment	Condition	Observed cline of perceptual success
1 (English discrimination)	stop/stop contrasts	{NJpn, NMnCh} < NRus < {NKor, NEng}
	stop/zero contrasts	{NJpn, NMnCh, NRus} < NEng < NKor
2 (English identification)	final stops	{NJpn, NRus} < NMnCh < {NKor, NEng}
	final non-stops	{NJpn, NMnCh} < {NRus, NKor, NEng}
3 (Korean identification)	final stops	{NJpn, NMnCh} < NEng < NRus < NKor
	final non-stops	NMnCh < NEng < {NJpn, NRus, NKor}

or, at a more basic level, differences between nonnative listeners whose L1s have the same high-ranked constraint against the target L2 configuration.[11] The cue-

---

[11] Although it is possible for the same target L2 configuration to be ruled out in various ways in the L1, a different formulation of the relevant L1 phonotactic constraints does not change the core limitation of the direct phonotactic view: namely, a level of analysis that is too coarse for fine-grained predictions about L2 perception. For example, Mandarin stop phonotactics can be formulated in three ways, in view of the ban on final /m/: (1) a manner-specific place constraint *[+sonorant, labial]#, complementing a general manner constraint *[−sonorant, −continuant]#, (2) a general place constraint *[labial]#, complementing a manner-specific place constraint *[−sonorant, coronal/dorsal]#, and (3) a general place constraint *[labial]#, overlapping a general manner constraint *[−sonorant, −continuant]#. All three formulations reflect the fact that certain place features are more free to occur in final position in Mandarin compared to others (e.g., *[labial]# is maximally restrictive, whereas *[−sonorant, coronal]# is less restrictive), but none speaks to the cohort of cues that need to be attended to in order to perceive those features. In other words, phonotactic constraints, with their focus on linguistic targets rather than the perceptual cues that are necessary to recover those targets, are fundamentally underinformative when it comes to perception.

centric view, by contrast, is able to account for these effects straightforwardly as the product of listeners' gradient L1 attunement to a crucial auditory cue.

Each of these three findings merits further comment. The first finding is discussed in greater detail in Chang (2016), which reports data from heritage Korean listeners that supports the interpretation of the NKor group's relatively weak advantage over the NEng group in English perception as indeed the result of L1 transfer from Korean. In short, heritage Korean listeners of the same age and education level as the NEng group show a much stronger advantage, outperforming NEng listeners by a significantly greater margin on both stop/zero discrimination and stop identification with response speeds that tend to be faster. These results suggest that, despite the inherent opportunity cost of exposure to Korean (which necessarily reduces the amount of exposure to the target language, English), heritage Korean listeners, as well as NKor listeners, extract a generalizable perceptual benefit from their experience attending to VC transitions in Korean. Again, this kind of native-language transfer benefit does not follow from the direct phonotactic view, but is easily explained under the cue-centric view.

As for the second and third findings involving the differences in performance between the NJpn and NMnCh groups, note that these findings cannot be an artifact of differences in L2 proficiency, education level, or other variables that might be related broadly to improved perception because the directionality of the group difference is inconsistent across conditions for the same target language (English) and, moreover, within the same experiment (Experiment 2). That is to say, if a hypothetically higher English proficiency level is what led to the NMnCh group outperforming the NJpn group in English stop identification, this should have led to better performance on non-stop-final stimuli, too. Therefore, the observed pattern on non-stop-final stimuli, where it is the NJpn group outperforming the NMnCh group, rules out an explanation of the NJpn-NMnCh disparities in terms of differences in uncontrolled factors that would globally affect English perception.

Instead, it is argued that the NJpn-NMnCh disparities are due to the different ways in which L1-specialized perception routines bias NJpn and NMnCh listeners' processing of L2 speech. In the case of Japanese, little perceptual attention (PA) is devoted to VC transitions because these carry a low relative functional load (RFL). Before a word-final consonant, they do not cue a contrast because none exists; before a word-medial consonant or consonant sequence (which is always followed by a vowel), they cooccur with CV transitions and/or the release burst of a prevocalic stop, both arguably stronger cues. On the other hand, Japanese SPRs involve high PA to vowel duration, due to its high RFL as the marker of a length contrast (cf. /kado/ 'corner' vs. /kaːdo/ 'card', /kaze/ 'wind' vs. /kazeː/ 'taxation'; Tajima et al., 2008). In the case of Mandarin, more PA is devoted to VC transitions due to their higher RFL in Mandarin. Unlike Japanese, Mandarin does contrast consonants in word-final position, even if the contrast is limited to sonorants (/n ŋ/, which have some weak internal cues, /ɻ/, depending on dialect, and /j w/; Duanmu, 2007, 2014) and there is some covariation of vowel quality with the coda. However, Mandarin SPRs do not include much PA to vowel duration due to a low RFL; Mandarin has no length contrast, and other contrasts involving duration, such as tone contrasts (see, e.g., Chang & Yao, 2007), are signaled by strong primary cues (e.g., voice pitch, voice quality). Thus, the picture that emerges from considering the RFL of, and resulting PA to, VC transitions and vowel duration in Japanese and Mandarin is one that predicts exactly the complementary disparities between NJpn and NMnCh listeners in English perception:

more PA to VC transitions for NMnCh listeners is reflected in better identification of final stops, while more PA to vowel duration for NJpn listeners is reflected in better identification of the absence of a final stop (i.e., an "open" syllable quality).

In regard to Q2 about the interaction of L1 transfer with L2 familiarity, those listeners whose L1s did not provide much motivation to attend to VC transition cues were indeed relatively more disadvantaged when the target language was unfamiliar (Experiment 3). This was evident in the larger decrements in accuracy on final stops for NJpn and NMnCh listeners (compared to NRus, NEng, and NKor listeners) in Experiment 3 than in Experiment 2, supporting P3. With the exception of the NJpn-NEng difference (which was actually larger in Experiment 2), all other group differences between the NJpn and NMnCh groups on the one hand and the NRus, NEng, and NKor groups on the other hand were larger in Experiment 3 (mean difference of 23% in Experiment 3 vs. 10% in Experiment 2), a pattern that could not be explained in terms of speed-accuracy tradeoffs. These findings are thus consistent with the view (of several theoretical frameworks, such as ASP and the Ontogeny Phylogeny Model; see Major, 2001) that L1 transfer decreases over the course of L2 acquisition with the development of an L2 system. For listeners whose L1 provides good reason (i.e., high RFL) to attend to VC transitions (e.g., NKor), transfer of L1 SPRs to L2 perception is less detrimental, and can even be advantageous, since these SPRs devote substantial PA to VC transitions. However, for listeners whose L1 provides little reason to attend to VC transitions, transfer of L1 SPRs to L2 perception is especially negative because in these SPRs VC transitions are largely ignored. Consequently, acquiring knowledge of the target L2 (including appropriate perceptual attunement to VC transitions) stands to particularly benefit listeners who are most at risk for negative transfer.

Although Experiments 2–3 differed in design in a few ways, comparing the results from these two experiments by group reveals two patterns. First, accuracy on final stops was higher in Experiment 3 than in Experiment 2 for all groups (as expected from the isolated presentation format, strictly monophthongal vowel contexts, and unitary talker used in Experiment 3). Second, the increase in accuracy from Experiment 2 to 3 differed considerably across groups. Whereas the NMnCh and NEng groups showed small increases in accuracy (2% and 8%, respectively), the NJpn, NRus, and NKor groups showed significantly larger increases (17–38%). By comparison, the absence of a final voiceless stop was identified with similarly high accuracy in Experiment 3 relative to Experiment 2 by the NJpn, NRus, and NKor groups (increases of 1–2%), but with lower accuracy by the NEng and NMnCh groups (decreases of 6–13%). However, the most salient disparity in performance between the two experiments was the nearly 40% difference in accuracy on final stops for the NRus group, the result of their lower-than-expected accuracy in English and higher-than-expected accuracy in Korean.

This raises the question of why NRus listeners showed these unexpected patterns of performance. One potential explanation for NRus listeners' unexpectedly poor identification in English is an overgeneralization of burst occurrence in English. Perhaps, for example, the consistent realization of final stops as released in Russian biased NRus listeners to pick up on released tokens of final stops in English, resulting in L2 SPRs in which VC transitions were given inappropriately low PA. Using such ineffectual L2 SPRs to perceive the English stimuli would account for NRus listeners' poor English identification performance; however, under this account, they should also have underperformed in English discrimination and Korean identification (which they did not). In other words, NRus listeners' performance in Experiments 1 and 3 strongly suggests that they were capable of using VC transition cues, but this ability was blocked in Experiment 2 for some reason.

The reason that ASP would offer for this kind of blocking in Experiment 2, but not in Experiment 1, is the difference in task demands between Experiments 1 and 2: Experiment 2 was more difficult due to the more detailed identification response required, the non-word status of the target items, and the embedding of these items within a sentence-length utterance. Consequently, it is possible that NRus listeners performed relatively worse (including worse than NMnCh listeners) in Experiment 2 because of increased task demands that caused them to revert to (ill-suited) L1 SPRs. This could only make sense, however, if the NRus group was more affected by the demands of Experiment 2 than the other groups were, which would in turn imply that NRus listeners had lower English proficiency (and, thus, less ability to cope with higher demands in an English perceptual task). Unfortunately, formal proficiency scores for the participants are not available; however, it is worth noting that compared to the NMnCh group that outperformed them in Experiment 2, the NRus group was, on average, older at the time of study [Welch-corrected two-sample $t(30.2) = 3.773, p < .001$], older upon arrival in the U.S. [Welch-corrected two-sample $t(32.8) = 2.718, p < .05$], and more variable in age, age of arrival, and time speaking English (see Table 2). These facts are consistent with a scenario in which the NRus group had lower English proficiency, though without actual proficiency data we can only speculate on this point.

As for NRus listeners' exceptionally accurate identification in Korean, this result suggests that NRus listeners were not only capable of utilizing VC transition cues (as mentioned above), but in fact more attuned to VC transition cues than NEng listeners were in the perception of an unfamiliar L2. Given the comparative estimates of RFL and PA outlined at the beginning of this article, this reversal of the NRus and NEng groups on Korean is surprising. Note that a higher estimation of RFL of VC transitions in Russian (based on including plain-palatalized contrasts in the count of contrasts[12]) would predict only that NRus listeners should be more attuned to VC transitions across the board. However, the NRus group outperformed the NEng group only on Korean, suggesting that these two groups responded to the unfamiliarity of this language in different ways: whereas the NEng group appeared to transfer L1 SPRs from English, the NRus group appeared instead to retune their perception or revert to a language-general perceptual mode.

This disparity between the NEng and NRus groups raises a number of interesting questions. For example, what factors encourage the favorable perceptual adaptation seen in the NRus group but not the NEng group? Furthermore, given that NRus listeners seem not to transfer L1 SPRs from Russian to perceive Korean, why do they not transfer L2 SPRs from English? A burgeoning area of cross-language speech research is the investigation of third-language (L3) phonology (Gallardo del Puerto, 2007; Onishi, 2013; Wrembel, 2014), which points to an alternate possibility for perception of Korean (technically an L3 for the NRus group): L2 transfer rather than L1 transfer. The fact that perception of an L2 is positively correlated with perception of an L3 (Onishi, 2013) is consistent with

---

[12] Although Russian has the same major places of articulation in stops as English (i.e., labial, coronal, dorsal), there may be effectively more place contrasts in Russian because final labial and coronal stops can occur in both plain (“hard”) and palatalized (“soft”) versions (Timberlake, 2004). Since these secondary articulations can be regressively assimilated by preceding consonants (see, e.g., Barry, 1992; Daniels, 1972), it is possible that they leave a trace in a preceding vowel as well, which would increase the RFL of VC transitions in Russian.

the view that L2 transfer is one type of transfer that can occur in L3 acquisition. Nevertheless, L2 transfer was not readily identifiable in NRus listeners' performance, as evident in the lack of similarity between their outcomes in Experiments 2 and 3. Thus, while there is at least one proposal in the L2 speech literature for how L1 transfer interacts with universal processes in L2 acquisition (Ontogeny Phylogeny Model; Major, 2001), more research is needed to understand how L2 transfer interacts with both of these factors over the course of L3 acquisition.

5. Conclusion

To return to the direct phonotactic and cue-centric views articulated at the beginning of this paper, recall that ostensibly phonotactic transfer (as in Davidson, 2011b) was also able to be explained in terms of attentional transfer—namely, transfer of SPRs shaped by the RFL of acoustic cues in the L1. In the case of Russian listeners' superior cluster/non-cluster discrimination, for example, this finding could be attributed to either of two kinds of advantage that Russian listeners have over Catalan/English listeners: (1) relative freedom from constraints against consonant clusters (the direct phonotactic view), or (2) greater perceptual attunement to acoustic cues contained in the vocalic interval between consecutive consonants, which distinguish between consonant adjacency vs. non-adjacency (the cue-centric view). In fact, insofar as the phonological patterns of L2 listeners' L1 conspire to either limit or enhance the amount of perceptual attention paid to the crucial cues associated with an L2 target, it will generally be possible to reframe apparent cases of phonotactic transfer in L2 perception as cases of cue-based attentional transfer.

Despite this empirical overlap, however, this study has shown that direct phonotactic and cue-centric views of transfer do not necessarily converge on the same predictions; in particular, the direct phonotactic view may not lead to the right predictions without being supplemented by the insights of the cue-centric view. If transfer must be able to occur at a cue-based level to make the right predictions, though, this raises the question of whether L2 perception is ever influenced by transfer at an unambiguously phonotactic (i.e., truly abstract) level. Some researchers suggest that abstract—and even innate—phonological knowledge must play a role in L2 perception (e.g., Berent et al., 2007; Berent & Lennertz, 2010); however, previous findings interpreted in terms of an abstract effect may often not reflect abstract knowledge per se (cf. Peperkamp, 2007), and it is clear that L2 perception must, in any case, engage attention to subphonemic details (Wilson et al., 2014). Addressing this question satisfactorily may thus require languages that show larger mismatches between phonotactics and the RFL of cues, which are likely to involve typologically unusual patterns. In future work, for example, it would be interesting to examine listeners of two unusual types of L1s (see Table 7): (1) Type A, which disallows /p t k/ finally, but otherwise allows many place contrasts among final sonorants without strong internal cues (e.g., nasals), and (2) Type B, which allows /p t k/ finally, but only released, and no other final place contrasts.[13] The properties of Type A languages are disadvantageous at the level of phonotactics, but advantageous at the level of the cue, while the properties of

---

[13] By crowding the space of coda contrasts with several places of articulation within one manner of articulation instead of maximally utilizing a few places across (multiple) manners, Type A languages run counter to a typological preference for featural economy (Clements, 2003; Martinet, 1968). By allowing coda consonants but limiting these to obstruents, Type B languages run counter to a typological preference for sonorants—in particular, nasals—as syllable codas (Blevins, 2004).

Table 7: Summary of phonotactic and cue-centric properties of Type A and B languages in terms of their potential consequences for perceiving L2 final voiceless stops (/p t k/) that are unreleased. The crucial variable at the level of phonotactics is whether or not /p t k/ are allowed finally; the crucial variable at the level of the cue is the relative functional load (RFL) of VC transitions.

Language type	Variable	Potential effect
Type A	PHONOTACTIC: final /p t k/ disallowed	*disadvantageous*
	CUE-CENTRIC: high RFL of VC transitions, due to many final place contrasts among sonorants without strong internal cues (e.g., /m n̪ n ɳ ɲ ŋ ɴ/)	*advantageous*
Type B	PHONOTACTIC: final /p t k/ allowed	*advantageous*
	CUE-CENTRIC: low RFL of VC transitions, due to few final place contrasts (limited to released /p t k/)	*disadvantageous*

Type B languages are essentially the reverse. According to the cue-centric view, Type A speakers should be better at perceiving L2 unreleased stops because they are biased to attend to VC transitions, and their advantage over Type B speakers should be greater when the L2 is unfamiliar as opposed to familiar. Importantly, however, this is the prediction only in case the L1 transitions are similar enough to the L2 transitions that the L1 bias is in fact helpful. As shown by Tsukada et al. (2007), L2 listeners who speak various L1s with unreleased stops do not show the same degree of native-language transfer benefit in perception of L2 final stops, which may be related to crosslinguistic variability in patterns of coarticulation and, thus, in the phonetic quality of VC transitions in the L1.

To be clear, it is not the claim of this paper that phonotactic constraints of the L1 play no role in L2 learning. There is abundant evidence that L1 phonological patterns, including phonotactics, influence L2 production (for a review, see Broselow & Kang, 2013), and whether abstract L1 patterns clearly distinct from processing biases may additionally influence L2 perception remains an open question (though cf. Boomershine et al., 2008). The point is rather that a cue-centric view of transfer makes better predictions in regard to L2 perception than a direct phonotactic view. Thus, it is argued here that the question to ask in regard to L1 influence in L2 perception is not *whether* the target is permitted in the L1, but rather *how much* the relevant acoustic cues are attended to in the L1 (which involves considering their RFL). This level of analysis is different from that in current frameworks of L2 perception, such as the gestural level in PAM-L2 (Best & Tyler, 2007) and the position-specific allophonic level in SLM (Flege, 1995). Furthermore, it involves broad consideration of a cue's function across the L1. For example, as outlined in Section 1.1, estimating the RFL of VC transition cues involves considering not just the VC transition cues to the exact target structure (final voiceless stops), which may not occur in the L1, but rather VC transitions in general (i.e., in any context that may increase their unique linguistic burden).

In closing, the contribution of the present study to the literature on L2 acquisition and nonnative speech perception is in the cue-centric view of L1 transfer as an issue about gradient biases to attend to acoustic cues as well as acquired L2 knowledge. For the comparative purposes of this study, perceptual attention to a cue, and its basis in the cue's RFL, were considered mainly in comparative terms, based on a working quantitative definition of RFL. However, recent work quantifying the notion of functional load for contrasts (Kang & Johnson, 2014; Wedel et al., 2013) provides some insight into how RFL for cues might be quantified more precisely in future work. The examination of RFL for cues as a factor shaping SPRs for the L1, the transfer of L1 SPRs to L2 perception, and the interaction of L1 SPRs with L2 SPRs and universal processes in L3 perception promises to shed new light on both native and cross-language speech development and the ways in which native perceptual processes can and cannot be adapted to suit the requirements of a new language.

Acknowledgments

The author gratefully acknowledges funding from the Center for Advanced Study of Language and logistical assistance from the Department of Hearing and Speech Sciences and Second Language Acquisition Program at the University of Maryland and from the Department of Linguistics at New York University. The paper benefited from the feedback of Taehong Cho, Karthik Durvasula, and several anonymous reviewers, as well as discussions with Nick Fleisher, Slava Gorbachov, Kevin Roon, Geoff Schwartz, and audiences at the CUNY Graduate Center, the University of Cambridge, University College London, MIT, the 7th International Symposium on the Acquisition of Second Language Speech, and the 167th Meeting of the Acoustical Society of America (Chang, 2014).

References

Altenberg, E. P. (2005). The judgment, perception, and production of consonant clusters in a second language. *International Review of Applied Linguistics in Language Teaching*, *43*(1), 53–80.

Barry, M. C. (1992). Palatalisation, assimilation and gestural weakening in connected speech. *Speech Communication*, *11*(4–5), 393–400.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv*, *1506.04967*, 1–27.

Berent, I. & Lennertz, T. (2010). Universal constraints on the sound structure of language: Phonological or acoustic? *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 212–223.

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, *104*(3), 591–630.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 171–204). Baltimore, MD: York Press.

Best, C. T. & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege* (pp. 13–34). Amsterdam, The Netherlands: John Benjamins Publishing.

Blevins, J. (2004). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge, UK: Cambridge University Press.

Boersma, P. & Weenink, D. (2011). Praat: Doing phonetics by computer. Version 5.3. http://www.praat.org.

Bohn, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 279–304). Baltimore, MD: York Press.

Bohn, O.-S. & Best, C. T. (2012). Native-language phonetic and phonological influences on perception of American English approximants by Danish and German listeners. *Journal of Phonetics*, *40*(1), 109–128.

Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (2008). The impact of allophony versus contrast on speech perception. In P. Avery, B. E. Dresher & K. Rice (Eds.), *Contrast in Phonology: Theory, Perception, Acquisition* (pp. 145–171). Berlin, Germany: Mouton de Gruyter.

Bradlow, A. R. & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America*, *106*(4), 2074–2085.

Broselow, E. & Kang, Y. (2013). Phonology and speech. In J. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge Handbook of Second Language Acquisition* (pp. 529–554). Cambridge, UK: Cambridge University Press.

Brown, C. A. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research*, *14*(2), 136–193.

Brown, C. [A.] (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (Ed.), *Second Language Acquisition and Linguistic Theory* (pp. 4–63). Malden, MA: Blackwell Publishers.

Byrd, D. (1993). 54,000 American stops. *UCLA Working Papers in Phonetics*, *83*, 97–116.

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, *40*(2), 249–268.

Chang, C. B. (2014). Transfer effects in perception of a familiar and unfamiliar language. *Journal of the Acoustical Society of America*, *135*(4), 2355.

Chang, C. B. (2015). Determining cross-linguistic phonological similarity between segments: The primacy of abstract aspects of similarity. In E. Raimy & C. E. Cairns (Eds.), *The Segment in Phonetics and Phonology* (pp. 199–217). Chichester, UK: John Wiley & Sons.

Chang, C. B. (2016). Bilingual perceptual benefits of experience with a heritage language. *Bilingualism: Language and Cognition*, *19*(4), 791–809.

Chang, C. B. & Mishler, A. (2012). Evidence for language transfer leading to a perceptual advantage for non-native listeners. *Journal of the Acoustical Society of America*, *132*(4), 2700–2710.

Chang, C. [B.] & Yao, Y. (2007). Tone production in whispered Mandarin. In Trouvain, J. & Barry, W. J. (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1085–1088). Dudweiler, Germany. Pirrot.

Chang, C. B., Yao, Y., Haynes, E. F., & Rhodes, R. (2011). Production of phonetic and phonological contrast by heritage speakers of Mandarin. *Journal of the Acoustical Society of America*, *129*(6), 3964–3980.

Cho, T. & McQueen, J. M. (2006). Phonological versus phonetic cues in native and non-native listening: Korean and Dutch listeners' perception of Dutch and English consonants. *Journal of the Acoustical Society of America*, *119*(5), 3085–3096.

Clements, G. N. (2003). Feature economy in sound systems. *Phonology*, *20*(3), 287–333.

Cutler, A. (2001). Listening to a second language through the ears of a first. *Interpreting*, *5*(1), 1–23.

Cutler, A., Garcia Lecumberri, M. L., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: Effects of local context. *Journal of the Acoustical Society of America*, *124*(2), 1264–1268.

Daniels, W. J. (1972). Assimilation in Russian consonant clusters: I. *Paper in Linguistics*, *5*(3), 366–380.

Davidson, L. (2011a). Characteristics of stop releases in American English spontaneous speech. *Speech Communication*, *53*(8), 1042–1058.

Davidson, L. (2011b). Phonetic, phonemic, and phonological factors in cross-language discrimination of phonotactic contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 270–282.

Davidson, L. & Roon, K. (2008). Durational correlates for differentiating consonant sequences in Russian. *Journal of the International Phonetic Association*, *38*(2), 137–165.

Davies, M. (2008-). The corpus of contemporary American English: 425 million words, 1990–present. Available online at http://corpus.byu.edu/coca/.

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language*, *59*(4), 447–456.

Duanmu, S. (2007). *The Phonology of Standard Chinese* (2nd ed.). Oxford, UK: Oxford University Press.

Duanmu, S. (2014). Syllable structure and stress. In C.-T. J. Huang, Y.-H. A. Li, & A. Simpson (Eds.), *The Handbook of Chinese Linguistics* (pp. 422–442). Chichester, UK: Wiley-Blackwell.

Dupoux, E., Hirose, Y., Kakehi, K., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1568–1578.

Ernestus, M., Kouwenhoven, H., & van Mulken, M. (2017). The direct and indirect effects of the phonotactic constraints in the listener's native language on the comprehension of reduced and unreduced word pronunciation variants in a foreign language. *Journal of Phonetics*, *62*, 50–64.

Escudero, P. (2009). Linguistic perception of "similar" L2 sounds. In P. Boersma & S. Hamann (Eds.), *Phonology in Perception* (pp. 151–190). Berlin, Germany: Mouton de Gruyter.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233–272). Baltimore, MD: York Press.

Flege, J. E. (2003). A method for assessing the perception of vowels in a second language. In E. Fava & A. Mioni (Eds.), *Issues in Clinical Linguistics* (pp. 19–43). Padova, Italy: Unipress.

Gallardo del Puerto, F. (2007). Is L3 phonological competence affected by the learner's level of bilingualism? *International Journal of Multilingualism*, *4*(1), 1–16.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia*, *9*(3), 317–323.

Hallé, P. A., & Best, C. T. (2007). Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters. *Journal of the Acoustical Society of America*, *121*(5), 2899–2914.

Hallé, P. A., Best, C. T., & Levitt, A. (1999). Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics*, *27*(3), 281–306.

Hallé, P. A., Segui, J., Frauenfelder, U., & Meunier, C. (1998). Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 592–608.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47–B57.

Iwasaki, S. (2013). *Japanese* (revised ed.). Volume 17 of *London Oriental and African Language Library*. Amsterdam, The Netherlands: John Benjamins Publishing.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

Jones, D. & Ward, D. (1969). *The Phonetics of Russian*. Cambridge, UK: Cambridge University Press.

Kang, S. & Johnson, K. (2014). Effects of linguistic structure on perceptual attention given to different speech units. *Journal of the Acoustical Society of America*, *135*(4), 2225.

Kang, Y. (2003). Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology*, *20*(2), 219–273.

Katz, J. (2012). Compression effects in English. *Journal of Phonetics*, *40*(3), 390–402.

Kim, H. & Jongman, A. (1996). Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean. *Journal of Phonetics*, *24*(3), 295–312.

Kruskal, W. H. & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*(260), 583–621.

Lado, R. (1957). *Linguistics Across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor, MI: University of Michigan Press.

Levy, E. S. & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics*, *36*(1), 141–157.

Lindblom, B. & MacNeilage, P. (2011). Coarticulation: A universal phonetic phenomenon with roots in deep time. *TMH - QPSR*, *51*(1), 41–44.

Lisker, L. (1999). Perceiving final voiceless stops without release: Effects of preceding monophthongs versus nonmonophthongs. *Phonetica*, *56*(1–2), 44–55.

Macmillan, N. A. & Creelman, C. D. (2005). *Detection Theory: A User's Guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Maddieson, I. (1985). Phonetic cues to syllabification. In V. A. Fromkin (Ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged* (pp. 203–221). New York, NY: Academic Press.

Major, R. C. (2001). *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. Mahwah, NJ: Lawrence Erlbaum Associates.

Malécot, A. (1958). The role of releases in the identification of released final stops: A series of tape-cutting experiments. *Language*, *34*(3), 370–380.

Martinet, A. (1933). Remarques sur le système phonologique du français. *Bulletin de la Société Linguistique de Paris*, *34*, 192–202.

Martinet, A. (1968). Phonetics and linguistic evolution. In B. Malmberg (Ed.), *Manual of Phonetics* (pp. 464–487). Amsterdam, The Netherlands: North-Holland Publishing.

Nábělek, A. K. & Donahue, A. M. (1984). Perception of consonants in reverberation by native and non-native listeners. *Journal of the Acoustical Society of America*, *75*(2), 632–634.

Newell, A. & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive Skills and Their Acquisition* chapter 1 (pp. 1–55). Hillsdale, NJ: Lawrence Erlbaum Associates.

Odlin, T. (1989). *Language Transfer: Cross-Linguistic Influence in Language Learning*. Cambridge, UK: Cambridge University Press.

Onishi, H. (2013). *Cross-Linguistic Influence in Third Language Perception: L2 and L3 Perception of Japanese Contrasts*. PhD thesis, University of Arizona, Tucson, AZ.

Park, H. & de Jong, K. J. (2017). Perceptual category mapping between English and Korean obstruents in non-CV positions: Prosodic location effects in second language identification skills. *Journal of Phonetics*, *62*, 12–33.

Parlato-Oliveira, E., Christophe, A., Hirose, Y., & Dupoux, E. (2010). Plasticity of illusory vowel perception in Brazilian-Japanese bilinguals. *Journal of the Acoustical Society of America*, *127*(6), 3738–3748.

Peperkamp, S. (2007). Do we have innate knowledge about phonological markedness? Comments on Berent, Steriade, Lennertz, and Vaknin. *Cognition*, *104*(3), 631–637.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic and acoustic contributions. *Journal of the Acoustical Society of America*, *89*(6), 2961–2977.

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception and Psychophysics*, *52*(1), 37–52.

R Development Core Team. (2015). R: A language and environment for statistical computing. Version 3.2.1. Vienna, Austria: R Foundation for Statistical Computing. http://www.r-project.org.

Rositzke, H. A. (1943). The articulation of final stops in General American speech. *American Speech*, *18*(1), 39–42.

Shapiro, S. S. & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3–4), 591–611.

Sheldon, A. & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243–261.

Sohn, H.-M. (1999). *The Korean Language*. Cambridge Language Surveys. Cambridge, UK: Cambridge University Press.

Steriade, D. (2009). The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In K. Hanson & S. Inkelas (Eds.), *The Nature of the Word: Studies in Honor of Paul Kiparsky* (pp. 151–179). Cambridge, MA: MIT Press.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, *39*(4), 456–466.

Sumner, M. & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, *60*(4), 487–501.

Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., & Munhall, K. G. (2008). Training English listeners to perceive phonemic length contrasts in Japanese. *Journal of the Acoustical Society of America*, *123*(1), 397–413.

Timberlake, A. (2004). *A Reference Grammar of Russian*. Cambridge, UK: Cambridge University Press.

Tsukada, K., Nguyen, T. T. A., Roengpitya, R., & Ishihara, S. (2007). Cross-language perception of word-final stops: Comparison of Cantonese, Japanese, Korean and Vietnamese listeners. In Trouvain, J. & Barry, W. J. (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1781–1784). Dudweiler, Germany. Pirrot.

Wang, W. S.-Y. (1959). Transition and release as perceptual cues for final plosives. *Journal of Speech and Hearing Research*, *2*(1), 66–73.

Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, *128*(2), 179–186.

Werker, J. F. & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*(2), 197–234.

Wilson, C., Davidson, L., & Martin, S. (2014). Effects of acoustic-phonetic detail on cross-language speech production. *Journal of Memory and Language*, *77*, 1–24.

Wrembel, M. (2014). VOT patterns in the acquisition of third language phonology. *Concordia Working Papers in Applied Linguistics*, *5*, 750–770.

Zsiga, E. C. (2003). Articulatory timing in a second language: Evidence from Russian and English. *Studies in Second Language Acquisition*, *25*(3), 399–432.