The role of abstraction in non-native speech perception

Bozena Pajak[a]*, Roger Levy[a]

*Corresponding author: bpajak@bcs.rochester.edu, Tel: +1 585 275 7171, Fax: +1 585 442 9216. Present address: Department of Brain & Cognitive Sciences, Meliora Hall, Box 270268, University of Rochester, Rochester, New York 14627-0268 USA.
[a] University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0108, USA

Abstract
The end-result of perceptual reorganization in infancy is currently viewed as a reconfigured perceptual space that is "warped" around native-language phonetic categories, which then acts as a direct perceptual filter on any non-native sounds: naïve-listener discrimination of non-native-sounds is determined by their mapping onto native-language phonetic categories that are acoustically/articulatorily most similar. We report results that suggest another factor in non-native speech perception: some perceptual sensitivities cannot be attributed to listeners' warped perceptual space alone, but rather to enhanced general sensitivity along phonetic dimensions that the listeners' native language employs to distinguish between categories. Specifically, we show that the knowledge of a language with short and long *vowel* categories leads to enhanced discrimination of non-native *consonant* length contrasts. We argue that these results support a view of perceptual reorganization as the consequence of learners' hierarchical inductive inferences about the structure of the language's sound system: infants not only acquire the specific phonetic category inventory, but also draw higher-order generalizations over the set of those categories, such as the overall informativity of phonetic dimensions for sound categorization. Non-native sound perception is then determined by sensitivities that emerge from these generalizations, rather than mappings of non-native sounds onto native-language phonetic categories.

Keywords
Non-native speech perception, sound discrimination, perceptual reorganization, naïve listeners, cross-linguistic influence, inductive inference

The role of abstraction in non-native speech perception

1. Introduction

The development of speech perception in the first year of life provides a critical foundation for future language learning. Infants undergo profound *perceptual reorganization* (Eimas, 1978): they transition from discriminating almost any speech sound distinction (including those absent from their ambient language) to a state of enhanced sensitivity to native-language (L1) distinctions, accompanied by a decline in sensitivity to many non-native distinctions (Werker & Tees, 1984; for reviews, see Werker, 1989; Kuhl, 2004). These results have led to the development of theories in which perceptual reorganization is understood as resulting from the acquisition of the specific inventory of native-language phonetic categories, and the end-state is a reconfigured ("warped") perceptual space, where innate perceptual sensitivity along natural auditory boundaries is replaced by sensitivity along boundaries of phonetic categories in the learner's native language (Kuhl, 1991, 2000).

As a consequence, the long-held assumption underlying the research on non-native speech perception has been that non-native speech is necessarily "filtered" through listeners' L1 phonetic category inventory. The "L1-category filter" metaphor can be traced back to Trubetzkoy (1939/1969), and the essence of this idea is present in current theories of non-native speech perception and learning: the Native Language Magnet model (NLM, Kuhl, 1992, 1994; Kuhl & Iverson, 1995; Kuhl, 2000; Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola, & Nelson, 2008), the Speech Learning Model (SLM, Flege, 1988, 1992, 1995), and the Perceptual Assimilation Model (PAM and PAM-L2, Best, 1993, 1994, 1995; Best & Tyler, 2007). These theories, while different in several respects, preserve the basic insight captured in the "L1-category filter" metaphor: that the perceptual space warped in accordance with the L1 phonetic category inventory – the end-result of perceptual reorganization in infancy – acts as a perceptual filter when processing non-native languages. Specifically, according to these theories, naïve-listener and second-language (L2) learner discrimination of non-native sounds is determined by their mapping onto specific L1 phonetic categories that are acoustically or articulatorily most similar, if such categories are available. Broadly speaking, discrimination of non-native contrasts is thought to be impaired when the stimuli are mapped (i.e., perceptually assimilated) onto the same L1 category (with varying performance depending on the goodness of fit to that category), relative to when they are mapped onto differing categories.

These classic theories have been very successful in explaining a wide range of perceptual difficulties in non-native speech perception and learning (Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975; Flege & Eefting, 1987; Best & Strange, 1992; Polka, 1991, 1992; Hallé, Best, & Levitt, 1999; Best, McRoberts, & Goodell, 2001; McAllister, Flege, & Piske, 2002; Best & Hallé, 2010, among others; for a review see Strange & Shafer, 2008), showing that the degree of similarity between native and non-native sounds – as assessed through acoustic and articulatory comparisons or direct measures of perceived similarity – can predict performance on discrimination of non-native sound pairs. That is, if two non-native sounds are both assessed as highly similar to a single L1 category, their discrimination is impaired. On the other hand, if each sound in the non-native pair is highly similar to a distinct L1 category, then their discrimination is facilitated. A widely cited example is the difficulty of L1-Japanese speakers in discriminating the English /ɹ/-/l/ distinction, which is generally attributed to Japanese only having one phonetic category in the same acoustic-phonetic range (Goto, 1971; Strange & Dittmann, 1984; Miyawaki et al., 1975). This type of examples have been used as evidence supporting the classic theories since perceptual difficulties can in this case be explained by L1-Japanese listeners' assimilating both of the non-native sounds onto a single L1 category.

However, recent evidence suggests that the theories of non-native speech perception might need to be extended to accommodate discrimination patterns that cannot be explained by specific L1 phonetic category inventory. In particular, it has been shown that native French, Dutch, and German listeners outperform English native speakers on discriminating the English /w/-/j/ contrast (Hallé et al., 1999; Bohn & Best, 2012). These results are extremely surprising for at least two reasons: (1) native speakers performed more poorly on discriminating sounds from their native language than did non-native listeners, and (2) this non-native perceptual advantage was observed for speakers of languages that do not even have /w/ in their inventory (Dutch and German). Bohn and Best suggested that these results could be explained by the influence of more general characteristics of the L1 inventory on the listener's perceptual system: French, Dutch, and German have a relatively rich vowel inventory, and – unlike English – use lip rounding contrastively for vowels; since lip rounding is one of the cues to distinguish /w/ from /j/, the practice with lip rounding to distinguish between L1 vowels might boost French, Dutch, and German listeners' performance on discriminating the /w/-/j/ contrast. In other words, phonological principles – such as whether or not L1 uses a given feature contrastively – may affect perception of non-native contrasts.

Bohn and Best's proposal echoes prior suggestions that phonological distinctive features may affect non-native speech perception and learning (e.g., Hancin-Bhatt, 1994; Brown, 1997, 2000; McAllister et al., 2002). For example, McAllister et al. (2002) developed one of the components of the SLM into the "feature hypothesis" stating that "[phonological] features not used to signal phonological contrast in L1 will be difficult to perceive for the L2 learner and this difficulty will be reflected in the learner's production of the contrast based on this feature" (p. 230). That is, under the feature hypothesis, the difficulty in learning a given L2 contrast based on feature X is determined by the role of feature X in the learner's L1. In particular, forming an L2 phonetic category may be more difficult (or even blocked) if it requires attending to a feature that is not exploited in the learner's L1. As a support for their hypothesis, McAllister and colleagues showed that L2 learners of Swedish are more successful at acquiring the short vs. long vowel distinctions if their L1 also employs the length feature to distinguish between vowels (as for L1-Estonian learners), relative to the case when duration is not a major cue to vowel contrasts in the learners' L1 (as for L1-Spanish or L1-English learners). In a similar vein, Brown (1997, 2000) argued for a model of non-native speech perception in which any phonological distinctive features used contrastively in L1 would be transferred to L2, which would in turn facilitate discrimination of any non-native sounds that are contrasted by those features.

Further support for the idea that more general phonological principles might affect non-native speech perception comes from the literature on speech perception and learning by infant learners. In particular, it has been shown that exposing infants to particular sound categories leads not only to their enhanced sensitivity to those specific categories, but also to generally enhanced sensitivity to the underlying phonetic dimension that these categories are defined by (Maye, Weiss, & Aslin, 2008). More specifically, English-learning infants exposed to a non-native voice onset time (VOT) distinction (e.g., the prevoiced [d] vs. voiceless unaspirated [t]) showed enhanced perceptual sensitivity not only to that specific contrasts, but also to an unfamiliar contrast along the same dimension (e.g., prevoiced [g] vs. voiceless unaspirated [k]).

It is noteworthy that the emergence of such general dimension-based perceptual sensitivity (for dimensions or features such as voicing, length, or lip rounding) would be a natural consequence of language development from a rational perspective on learning (e.g., Chater & Manning, 2006; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). This is because languages re-use a limited set of phonetic dimensions or features for signalling multiple contrasts (Clements, 2003). Thus, enhancing sensitivity to relevant dimensions that encompass many potential contrasts is more adaptive for a learner than only enhancing sensitivity to specific categories that a learner has been directly exposed to. Note that learning need not be sequential (i.e., first learning [d]-[t], and then [g]-[k], as in the Maye et al. study) in order for learners to benefit from general enhanced sensitivity to phonetic dimensions or features. When learning all phonetic categories at once, as is the case in naturalistic language acquisition, noticing featural relationships between categories would make learning more efficient because less data may be needed to accurately identify and represent each individual category.

Taken together, recent evidence suggests that speech perception – both native and non-native – is affected not only by specific L1 phonetic category inventory, but also by more general phonological principles. But what exactly are those phonological principles? Bohn and Best (2012) showed that contrastive lip rounding in L1 vowels translates into enhanced sensitivity to a non-native lip-rounding contrast between semivowels ([w]-[j]). This result is suggestive of a phonological principle that generates sensitivity to the lip-rounding cue. However, there are still many open questions. For example, it is unclear what kinds of cues are the building blocks of such principles (phonological features, acoustic and/or articulatory cues, etc.), or how such principles operate: Do they apply only in cases when the corresponding contrasts are phonetically similar (as is the case with vowels and semivowels)? Or are they more abstract principles that apply regardless of the degree of phonetic similarity between known and non-native contrasts?

In this paper we begin to address these questions by examining the phonetic dimension of length. Length is a relatively salient cue that has been suggested to be universally accessible to L2 learners, irrespective of their L1 background (Bohn & Flege, 1990; Bohn, 1995). The reason for this suggestion was the observation that L2 learners tend to over-rely on length when distinguishing between non-native vowels for which length is only a secondary cue (e.g., the English lax-tense distinction). Critically, this is the case even for L2 learners whose L1 does not use length contrastively (e.g., native speakers of Spanish or Mandarin). However, even though length is indisputably salient, others have shown that L1 background does affect perception and learning of length contrasts (Hayes, 2002; McAllister et al., 2002; Escudero & Boersma, 2004; Goudbeek, Cutler, & Smits, 2008; Hayes-Harb & Masuda, 2008; Heeren & Shouten, 2008; Pajak, 2013). Furthermore, differential perception of length contrasts can already be observed in 18-month-old infants learning a language with phonemic vs. non-phonemic length, such as Dutch (Dietrich, Swingley, & Werker, 2007) or Japanese (Mugitani, Pons, Fais, Werker, & Amano, 2009). Given these recent studies, naïve-listener discrimination of length contrasts could well be modulated by the overall role of length in the listener's L1.

The main advantage of using length as the dimension of investigation is that it can be applied contrastively to a wide range of segments, both vowels and consonants. This means that a length contrast in one place of the L1 phonetic category inventory might potentially reveal an influence on perception of sounds in a completely different part of the perceptual space. Therefore, studying length allows us to investigate the limits of phonological abstractness that might guide non-native speech perception. Specifically, we ask whether sensitivity to length as a cue for distinguishing between L1 *vowels* translates into enhanced sensitivity to length as a cue for *consonants*, even when length is not used to distinguish between consonants in the listener's L1. This is of particular interest given that vowels and consonants as a group are acoustically and functionally very distinct. For example, vowels are perceived differently from consonants (e.g., vowel perception is less categorical than consonant perception; e.g., Schouten & van Hessen, 1992), and the two types of segments carry different kinds of information (Bonatti, Peña, Nespor, & Mehler, 2005).

Here we test discrimination of non-native consonant length contrasts by naïve listeners of different L1 backgrounds: Korean, where length is a highly informative contrastive cue for both vowels and consonants; Vietnamese and Cantonese, where length is informative, but more limited (only for vowels, and – for Cantonese – only as an additional cue together with changes in vowel quality); and Mandarin Chinese, where length is uninformative for segmental distinctions (see detailed language characteristics in section 2.2). We use an AX discrimination task, which can inform us about perceptual sensitivities of naïve listeners. If non-native speech perception is affected by a phonological principle that produces general enhanced sensitivity to those cues that are informative in L1, irrespective of the exact segments involved, then we expect all length-familiar participants (Korean, Vietnamese, Cantonese) to outperform

Mandarin speakers on discrimination of short versus long consonants. In addition, if the degree of cue informativity plays a role, then we expect gradient discrimination performance that corresponds to length informativity in L1: best for L1-Korean, medium for L1-Vietnamese and L1-Cantonese, and worst for L1-Mandarin. On the other hand, if this kind of phonological principle is restricted to segments that are acoustically/articulatory fairly similar, then we expect L1-Korean listeners to show relatively good discrimination of short and long consonants (since Korean uses length contrastively for some consonants), but the other three groups (Vietnamese, Cantonese, Mandarin) should pattern together and show relatively worse discrimination (given that none of these languages uses length contrastively for consonants).

As a control, we included *sibilant* place of articulation contrasts from Polish (alveolo-palatal vs. retroflex consonants) distinguished by dimensions familiar to speakers of Mandarin (where similar alveolo-palatal and retroflex consonants exist as allophones), but not to speakers of the other languages. Here, we expected a reverse performance pattern: Mandarin speakers outperforming other participants on discriminating between sibilants.

2. Materials and methods

2.1 Participants

96 undergraduate students at the University of California, San Diego participated in the experiment for course credit. Each participant was from one of four language groups: Korean, Vietnamese, Cantonese (all *length-familiar*), and Mandarin (*sibilant-familiar*). All Cantonese speakers also spoke some Mandarin, learned at school. Participants learned the target language from birth, and were bilingual in English. We collected detailed language background information through a questionnaire, and we recruited both L1-dominant and English-dominant participants (as self-reported) to ensure that the results would not be driven by language dominance. The questionnaire data revealed no major differences between language groups (see the table in the Appendix for detailed information). Participants reported no history of speech or hearing problems.

2.2 Language characteristics

Korean, Vietnamese, and Cantonese have segmental length distinctions, but no place of articulation distinctions between alveolo-palatals and retroflexes. Korean (Choi, 1995; Lee, 1999; Sohn, 1999; Hahm, 2007) has long vowels and consonants.[1] Long consonants mostly arise from phonological assimilation processes, and tense obstruents have sometimes been analyzed as long (Kim, 2002). In Vietnamese (Winn, Blodgett, Bauman, Bowles, Charters, Rytting, & Shamoo, 2008), length is phonologically contrastive for two sets of vowels, and consonants are always short. In Cantonese, (Kao, 1971; Bauer & Benedict, 1997; Zhang, 2011) length is one of the cues (often primary) to distinguish between vowels. Consonants are always short.

Mandarin (Lin, 2001) does not have segmental length contrasts, although Mandarin tones vary in length, and some listeners have been reported to use length to distinguish between tones when the main cue – the F0 pattern – is ambiguous (Tseng, Massaro, & Cohen, 1986; Blicher, Diehl, & Cohen, 1990; Jongman, Wang, Moore, & Sereno, 2006). As for sibilants, Mandarin has voiceless alveolo-palatals and retroflexes (as allophones), and the voiced retroflex fricative is a between-speaker variant of the retroflex approximant.

2.3 Materials

The materials consisted of phonologically-Polish nonce words recorded in a soundproof booth

[1] Phonemic vowel length contrasts are currently being lost from Korean (e.g., Magen & Blumstein, 1993). However, we tested participants who largely grew up in the United States, thus learning the Korean of their parents' generation. Consonant length contrasts are currently still present in Korean.

by a phonetically-trained Polish native speaker. Polish was chosen as the target language because it has both consonant length contrasts and alveolo-palatal vs. retroflex sibilant contrasts (Sawicka, 1995). Note that the Polish and the Mandarin alveolo-palatals and retroflexes differ in the exact place of articulation, but they share similar spectral cues that distinguish between the two (Ladefoged & Maddieson, 1996).
A complete list of sound segments used is provided in Table 1. The critical length items included short and long consonants. The control sibilant items included alveolo-palatal and retroflex consonants, which differ in the spectral shape of the frication noise. Each sound segment was recorded embedded in seven contexts: [pa_a], [pe_a], [po_a], [ta_a], [te_a], [ka_a], [ke_a], with five repetitions of each word. The words were produced in isolation in a random order. Then, the stimuli were manipulated through splicing to ensure that the minimal-pair words differed only in length or place of articulation, with no irrelevant differences present elsewhere in a word. One token of each word type was chosen as a frame (a short-consonant word for length contrasts and a retroflex-word for sibilant contrasts), from which the target segments were removed. The missing segments were spliced out from other recorded tokens and placed in the frames.

For length items, only short segments were spliced in; long segments were created from short ones by either doubling their length (for sonorants: [j], [w], [l], [m], [n]) or elongating them by half their length (for fricatives: [f], [s]). This difference was introduced to mimic natural production, reflecting the fact that intervocalic length contrasts are perceptually harder for sonorants than for fricatives due to more blurred segment boundaries (Kawahara, 2007). The duration of short segments was not manipulated. Instead, naturally produced short consonants from each word type were spliced in. The exact consonant duration values for the length items are listed in Table 2.

For sibilant items, both alveolo-palatals and retroflexes were spliced in. The formant transitions into and out of the medial consonants were not manipulated. In order to handle differences in formant transitions between alveolo-palatals and retroflexes, we first checked which frame type (an alveolo-palatal or a retroflex word) produced more natural-sounding stimuli. We chose retroflex-words as frames, which meant that the formant transitions were appropriate for the retroflex consonants, but not for the alveolo-palatal consonants. Consequently, one of the cues to the sibilant contrast – formant transitions out of alveolo-palatals – was partially removed, thus making this contrast perceptually less salient than in natural speech.

2.4 Procedure

The experiment consisted of a same-different AX discrimination task. On each trial, a pair of words was presented auditorily over headphones. The words were either “different” (e.g., pama-pam:a; all tested contrasts are listed in Table 3) or “same” (e.g., pama-pama). “Same” words in each pair were physically identical. “Different” words in each pair always shared a physically identical frame (i.e., the words were identical except for artificial lengthening for length contrasts and a spliced consonant for sibilant contrasts) to ensure that “different” responses resulted only from the manipulation of interest, and not due to irrelevant differences present elsewhere in a word. The words in each pair were separated by a 750 ms interstimulus interval. Each pair was repeated twice throughout the experiment, which yielded 392 pairs (196 length pairs, 196 sibilant and filler pairs; half “different”, half “same”), divided into seven 56-trial blocks separated by self-terminated breaks. In each trial, a word-pair was played once without a replay option, and the response to one trial triggered presentation of the subsequent trial with a 500 ms delay. Trial order was randomized for every participant. The testing was preceded by a 16-trial no-feedback practice session.

Table 1 ***Sound segments used as stimuli (in IPA), and the occurrences of corresponding sounds in Korean, Vietnamese, Cantonese, and Mandarin***

Segment			Kor	Viet	Cant	Mand
length stimuli	*short*	m	√	√	√	√
		n	√	√	√	√
		l	√	√	√	√
		s	√	√	√	√
		j	√	√	√	√
		w	√	√	√	√
		f		√	√	√
	long	m:	√[a]			
		n:	√[a]			
		l:	√[a]			
		s:	√[b]			
		j:				
		w:				
		f:				
		(vowels)[c]	(√)	(√)	(√)	
sibilant stimuli	*alv.-pal.*	ɕ	√			√[d]
		tɕ				√
		ʑ				
		dʑ				
	retroflex	ʂ				√
		tʂ				√
		ʐ				√
		dʐ				
filler stimuli		x		√	√	√
		χ				
		ʁ				
		ʝ				

[a] Almost exclusively as a result of phonological assimilation.
[b] Korean tense [s] is sometimes analyzed as long.
[c] Long vowels were not included in the stimuli.
[d] Alveolo-palatals and retroflexes are allophones in Mandarin.

Table 2 ***Consonant duration values (in ms) for the length items***

Segment type	Middle consonant durations (short/long)							Mean long/short duration ratio
	[pa_a]	[pe_a]	[po_a]	[ta_a]	[te_a]	[ka_a]	[ke_a]	
j/j:	73/153	81/160	78/154	82/164	77/153	87/180	76/149	2.01
w/w:	88/176	79/165	66/130	88/184	64/126	61/120	73/146	2.01
l/l:	71/142	76/151	78/154	70/142	72/143	73/146	95/193	2.00
m/m:	81/163	79/155	83/170	92/186	87/176	91/180	89/177	2.00
n/n:	74/148	74/157	73/156	71/147	86/172	82/166	74/145	2.04
s/s:	139/208	152/228	147/220	150/225	139/209	160/240	155/234	1.50
f/f:	128/192	142/213	127/191	130/195	128/192	135/202	149/223	1.50

Table 3 ***Tested contrasts***

Length	Sibilant	Filler
j-jː	ɕ-ʂ	x-χ
w-wː	tɕ-tʂ	χ-ʁ
l-lː	ʑ-ʐ	j-ʝ
m-mː	dʑ-dʐ	
n-nː		
s-sː		
f-fː		

3. Results

The results are plotted in Figures 1-2. For each tested contrast and each participant we calculated d-prime scores, which is a measure of contrast sensitivity based on the principles of Signal Detection Theory (Macmillan & Creelman, 2005).[2] We analysed the scores using repeated-measures ANOVAs.

First, we compared length-familiar participants (Korean, Vietnamese, Cantonese speakers) to sibilant-familiar participants (Mandarin speakers) using an ANOVA with the factors Language Group (*length-familiar*, *sibilant-familiar*) and Contrast (*length*, *sibilant*). There was a significant interaction between Language Group and Contrast [$F(1,94)=51.7$; $p<.001$]: length-familiar participants were more sensitive to length differences, and sibilant-familiar participants were more sensitive to sibilant contrasts. The results also revealed a main effect of Language Group [$F(1,94)=6.5$; $p<.05$]: Mandarin speakers performed overall worse than length-familiar participants as a group. Mandarin speakers' diminished sensitivity to length contrasts cannot, however, be attributed to worse overall performance, since the result was reversed for sibilant contrasts[3] (also supported by a significant interaction between Language Group and *sibilant* vs. *filler* Contrast [$F(1,94)=24.0$; $p<.001$]), and the differences between the two Language Groups on length and on filler contrasts were of different magnitudes (as indicated by a significant interaction between Language Group and *length* vs. *filler* Contrast [$F(1,94)=33.4$; $p<.001$]).

Crucially, the main result was not driven just by Korean performance, but also held for each relevant pairwise Language comparison, as indicated by significant interactions between Language and Contrast (Korean-Mandarin: [$F(1,46)=63.7$; $p<.001$]; Vietnamese-Mandarin: [$F(1,46)=33.1$; $p<.001$], Cantonese-Mandarin: [$F(1,46)=18.7$; $p<.001$]). The results reveal an extremely robust pattern: Korean, Vietnamese, and Cantonese speakers were consistently better at discriminating length contrasts than Mandarin speakers for each tested segment (Figure 2).

[2] D-prime scores are calculated by taking into account the standardized proportion of Hits (correct 'different' responses) and False Alarms (incorrect 'different' responses), with the following formula: $d' = z(FA) - z(H)$. D-prime values near zero indicate chance performance.

[3] Note that Mandarin speakers' relatively low performance extended even to the sibilant contrasts, which were discriminated slightly worse than the filler contrasts. This result might be due to two main factors: (1) one of the cues to the contrast – the formant transition out of the consonant – was partially removed due to splicing, and (2) the tested Polish alveolo-palatal vs. retroflex sibilant contrast differed in the exact place of articulation from the analogous Mandarin contrast (Ladefoged & Maddieson, 1996). The first factor made the contrast overall perceptually less salient, thus making it hard for all participants, while the second factor possibly diminished Mandarin speakers' perceptual advantage relative to other listeners on the sibilant stimuli. Indeed, when the formant transition cues are left intact, the discrimination of the Polish alveolo-palatal vs. retroflex sibilant contrast by Mandarin listeners is much higher (around 90% accuracy in an ABX task; Pajak, Creel, & Levy, in prep.).

Figure 1 ***Overall performance on perception of three types of contrasts: length, sibilant, and filler (error bars are standard errors)***

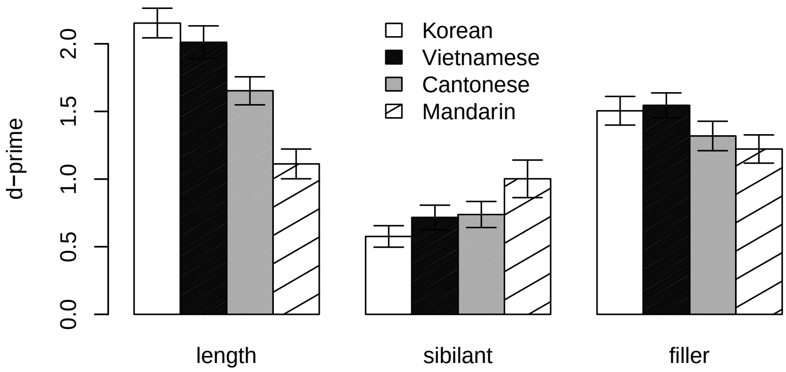
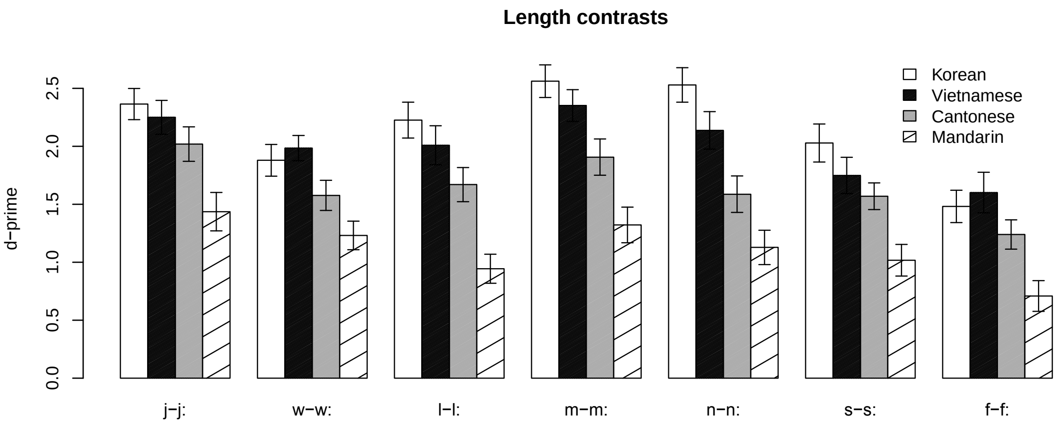


Figure 2 ***Performance on perception of length contrasts split by segment (error bars are standard errors)***



As for the comparisons within the length-familiar group, there was no significant difference on length contrasts for the Korean-Vietnamese pair [$F<1$], but there was a significant difference for both Korean-Cantonese [$F(1,46)=11.0$; $p<.01$] and Vietnamese-Cantonese [$F(1,46)=5.1$; $p<.05$], with Cantonese speakers performing worse. Therefore, the observed pattern of performance was: Korean, Vietnamese > Cantonese > Mandarin (although Vietnamese speakers performed numerically worse than Korean speakers). This result does not align exactly with our predictions, since we expected either (1) all length-familiar participants patterning together: Korean, Vietnamese, Cantonese > Mandarin, or (2) a gradient pattern: Korean > Vietnamese, Cantonese > Mandarin. However, this result is consistent with the idea that sensitivity to a given phonetic dimension is mediated by the degree of informativity that this dimension has in the learner’s native language. Recall that Korean uses length contrastively on both vowels and consonants, and Vietnamese uses it contrastively on vowels but not consonants. In Cantonese, on the other hand, length is only one of the cues to vowel distinctions (in addition to changes in vowel quality). Therefore, it appears that the major factor affecting perceptual sensitivities is the contrastiveness of a cue in L1, regardless of the exact segments it applies to (whether vowels or consonants). On the other hand, when a dimension is only one of the cues to a phonemic contrast (as length in Cantonese vowels), perceptual sensitivities to other contrasts along that dimension are also enhanced, but to a lesser degree.

4. Discussion

In this paper we investigated the extent to which general phonological principles affect non-native speech perception. Hallé et al. (1999) and Bohn and Best (2012) showed that when the listeners' L1 uses the lip-rounding cue to distinguish between vowels, their discrimination is also enhanced for a non-native lip-rounding contrast between semivowels ([w]-[j]). This particular result could be explained, as Bohn and Best suggested, by a simple phonological principle: contrastive lip rounding in L1 enhances perceptual sensitivity to the lip-rounding cue, which in turn leads to overall better discrimination of any non-native lip-rounding contrasts. Here we tested the extent of abstractness of such phonological principles: Do they apply only in cases when the corresponding contrasts are phonetically similar (as is the case with vowels and semivowels)? Or are they more abstract, applying regardless of the degree of phonetic similarity between known and non-native contrasts? We tested a different cue, length, which can span a wide range of segments, both vowels and consonants. We investigated whether sensitivity to L1 length differences on vowels will lead to enhanced perceptual sensitivity to non-native length differences on a range of semivowels and consonants (glides, liquids, nasals, and voiceless fricatives). We recruited participants of different L1 backgrounds: Korean, where length is a highly informative contrastive cue for both vowels and consonants; Vietnamese and Cantonese, where length is informative, but more limited (only for vowels, and – for Cantonese – only as an additional cue together with changes in vowel quality); and Mandarin Chinese, where length is uninformative for segmental distinctions.

The results revealed differential discrimination of non-native length contrasts: all length-familiar participants (Korean, Vietnamese, Cantonese) outperformed participants not familiar with length (Mandarin). Furthermore, there were differences between the length-familiar group: performance was best for Korean speakers, slightly worse for Vietnamese speakers, and worst for Cantonese speakers. These results suggest that experience with a phonetic dimension on a limited subset of segments can lead to enhanced perceptual sensitivity to any distinctions along that dimension, even those that are acoustically and functionally very dissimilar from the previously learned categories. This result cannot be attributed to better task performance, as the pattern was reversed for sibilant contrasts, which are more familiar to Mandarin speakers. Furthermore, we found some evidence of gradience: discrimination appears to be mediated by the degree of informativity that this dimension has in the learner's native language. Note that while we did not expect a difference between Vietnamese and Cantonese speakers, it may be that length is in fact less informative in Cantonese than in Vietnamese due to the fact that it is only one of the cues to vowel contrasts.

These results add to the previous findings suggesting that existing models of non-native speech perception and learning need to be extended, incorporating the role of abstract phonological principles. The influence of such phonological principles might be captured in one existing framework: Processing Rich Information from Multidimensional Interactive Representations (PRIMIR; Werker & Curtin, 2005; Curtin, Byers-Heinlein, & Werker, 2011). The basic idea of PRIMIR is that linguistic knowledge is represented at non-hierarchically organized multidimensional spaces, including a General Perceptual space, a Word Form space, and a Phoneme space, with abstract phonemes emerging after the infant has acquired some of the lexicon, integrating information from the other two spaces. Werker, Curtin, and Byers-Heinlein do not explicitly discuss the influence of general phonological principles on non-native speech perception (such as which cues are generally informative or contrastive in L1). However, this influence might be captured in PRIMIR by assuming, for example, a separate learning mechanism available to infants that dynamically evaluates the general informativity of different phonetic cues in L1. This kind of mechanism would aid the acquisition of the L1 sound system because noticing featural relationships between categories would allow learners to pool the data from multiple categories along the same dimension, and thus require less data from each individual category to form accurate category representations. Crucially, once established for L1, the generalizations about cue informativity would affect perceptual sensitivities in a non-native language as well, thus providing an explanation for the results reported here, as well as those by Hallé et al. (1999) and Bohn and Best (2012).

While we view PRIMIR as a promising framework, these ideas could also be developed within an alternative framework that builds on insights from recent rational approaches to learning (Tenenbaum et al., 2011). In particular, we propose to view perceptual reorganization as the consequence of learners' hierarchical inductive inferences about the structure of the language's sound system: infants not only acquire the specific phonetic category inventory, but also draw higher-order generalizations over the set of those categories, such as the overall informativity of phonetic dimensions for sound categorization. Non-native sound perception is then determined by sensitivities that emerge from these generalizations, rather than mappings of non-native sounds onto native-language phonetic categories. Below we motivate and develop this account in more detail. We hope that by adopting a different framework we will contribute additional insights to our understanding of perceptual reorganization and non-native speech perception.

4.1 Hierarchical inductive inference in perceptual reorganization

Current theories view perceptual reorganization in infancy as resulting from the acquisition of the specific inventory of L1 phonetic categories, and yielding a "warped" perceptual space that then acts as a direct filter in naïve-listener perception of non-native sounds (Kuhl, 1991, 2000). We propose to adopt an alternative framework, in which perceptual reorganization is viewed as a consequence of hierarchical inductive inferences over the sound inventory of a language: in addition to learning individual phonetic categories, learners are posited also to make higher-order generalizations about the general properties of the set of those categories. On our proposal, then, perceptual reorganization leads not only to perceptual sensitivity to specific L1 phonetic categories, but also to sensitivity induced by these higher-order generalizations.

Our proposal is based on rational approaches to learning, and in particular recent work in the computational modelling of acquisition of abstract knowledge, suggesting that learners find regularities in the input through (implicit) inductive inferences and constructing hierarchical representations with deep abstract structure (Chater & Manning, 2006; Tenenbaum et al., 2011). More specifically, learning has been shown to occur not only at a single, flat level of representation but rather hierarchically. This means that the learner makes simultaneous inductive inferences about both particular categories and higher-level category structure. To take one example from the language domain: when acquiring individual verbs in a language, learners also infer general properties of verbal classes (Perfors, Tenenbaum, & Wonnacott, 2010). In terms of recent work in phonetic category induction (de Boer & Kuhl, 2003; Vallabha, McClelland, Pons, & Werker, 2007; McMurray, Aslin, & Toscano, 2009; Feldman, Griffiths, & Morgan, 2009a; Feldman, Griffiths, Goldwater, & Morgan, 2013) – where native-speaker knowledge is represented as a set of distributions over perceptual space, one distribution for each phonetic category – our proposal can be viewed as adding a higher-level distribution responsible for generating specific category-level distributions.

One type of higher-order generalization that learners might make about their native-language sound system involves the underlying set of informative phonetic dimensions from which the system is constructed.[4] Positing dimension-based generalizations is motivated by evidence that infants encode speech sounds in terms of their subsegmental properties (e.g., Saffran & Thiessen, 2003; Maye et al., 2008; Cristià & Seidl, 2008; Cristià, Seidl, & Francis, 2011), suggesting that the information necessary to derive generalizations about phonetic dimension informativity in the native language is available to infant learners. The dimensions that infants

[4] Alternatively, it can be thought of as a set of phonological features that are active in the language. In this paper we do not make any theoretical commitment as to the exact content of subsegmental properties.

learn to attend to can include a range of phonetic cues with different degrees of informativity. The most informative dimensions are *contrastive*: they uniquely distinguish between phonemic categories (e.g., VOT differentiating English /b/-/p/, /t/-/d/, /k/-/g/) and are necessary to discriminate lexical items (e.g., *bin-pin*), but other properties may also be attended to (e.g., secondary cues to phonemic categories, cues distinguishing between allophones). Thus, any higher-order generalizations about phonetic dimensions are likely modulated by their gradient native-language informativity. Crucially, these generalizations will differ depending on the person's language background because the exact configuration of informative dimensions varies cross-linguistically (Ladefoged & Maddieson, 1996): for example, VOT distinguishes categories in English, but plays no role in Hawaiian; segmental length is contrastive for many Japanese categories (e.g., /p/-/p:/, /t/-/t:/), but is only a secondary cue to some English vowel distinctions (e.g., /ɪ/-/i/) and is uninformative in languages like Spanish.

Therefore, we propose that perceptual reorganization involves higher-order generalizations about phonetic dimension informativity. The crucial prediction of this account is that the end-state of perceptual reorganization is *not* just a perceptual space warped around L1 phonetic categories, as in the current theories. Instead, the end-state includes (1) the knowledge of the inventory of specific phonetic categories in the language (which in turn gives rise to perceptual warping effects around specific category boundaries for reasons such as those given by Feldman, Griffiths, & Morgan, 2009b), but also (2) heightened overall sensitivity and precision of encoding for those phonetic dimensions that determine category contrasts within the language as a whole. Thus, sensitivity is based on subsegmental properties abstracted away from individual categories.

Our proposal follows naturally from assuming a rational language learning and processing architecture, because for perceptual reorganization to proceed along these lines would be adaptive for the language learner. In particular, given that languages extensively re-use a limited set of phonetic dimensions (Clements, 2003), learning one distinction (e.g., /b/-/p/) might help learn analogous distinctions (e.g., /d/-/t/, /g/-k/). Hence, if the learner's perceptual sensitivity to a given phonetic dimension is enhanced once the dimension is determined to be informative for one set of L1 phonetic categories, this would allow for perceptual bootstrapping and more efficient learning of L1 phonetic categories overall. Existing experimental evidence supports this view: both infants and adults generalize newly learned phonetic category distinctions to untrained sounds along the same dimension (McClaskey, Pisoni, & Carrell, 1983; Maye et al., 2008; Perfors & Dunbar, 2010; Pajak & Levy, 2011a, 2011b; see also Pajak, Bicknell, & Levy, 2013, for a computational implementation). Our proposal attributes these learning results to the mechanisms underlying perceptual reorganization, which – as we propose – induce sensitivity to whole phonetic dimensions, not just individual phonetic categories.

Let us now turn to the consequences of this view of perceptual reorganization for non-native speech perception. If perceptual reorganization in infancy yields general sensitivity to phonetic dimensions that are informative in L1, then listeners' perception should be enhanced for any contrast along those informative dimensions. This means that naïve-listener perception of non-native sounds would be determined by perceptual sensitivities emerging from the higher-order generalizations made during perceptual reorganization, rather than direct mappings of non-native sounds onto L1 phonetic categories. That is, learning a single VOT distinction in L1 (e.g., /b/-/p/) would lead to enhanced sensitivity not only to that particular distinction, but also to analogous VOT distinctions (e.g., /d/-/t/, /g/-k/), even if they do not in fact occur in L1. Similarly, learning lip rounding as a cue for distinguishing between L1 vowels would lead to enhanced sensitivity to other lip-rounding contrasts, such as [w]-[j], as reported by Hallé et al. (1999) and Bohn and Best (2012). Finally, learning that length is an informative cue to distinguish between some categories (e.g., vowels) would lead to enhanced sensitivity to other length contrast (e.g., consonants), again, even if they do not in fact occur in the learner's L1. Therefore, the hierarchical inductive inference framework allows us to capture the present results, as well as the results of Hallé et al. (1999) and Bohn and Best (2012), thus incorporating the idea that non-native speech perception is affected by abstract phonological principles.

Although the results reported here only bear on one type of possible inference about learners' native-language sound system, the hierarchical inductive inference framework predicts that learners make other types of higher-order generalizations about linguistic structures that should affect non-native language processing, both in the sound domain (e.g., likely phonotactic patterns) and in other aspects of language (e.g., likely word and morpheme orderings). Thus, this framework fits within the broader literature investigating generalization of linguistic knowledge in different language domains (e.g., Xu & Tenenbaum, 2007; Wonnacott, Newport, & Tanenhaus, 2008; Gerken, 2010), and it provides theoretical unification with many other domains in which inductive approaches to learning have proven fruitful.

5. Conclusion

In this paper we investigated the role of abstract phonological principles in shaping non-native speech perception. We reported results demonstrating that some perceptual sensitivities cannot be attributed to listeners' warped perceptual space alone, as assumed in prior accounts, but rather to enhanced general sensitivity along phonetic dimensions that the listeners' native language employs to distinguish between categories. Specifically, we showed that knowledge of a language with short and long *vowel* categories leads to enhanced discrimination of non-native *consonant* length contrasts. These results suggest that abstract phonological principles – such as whether a phonetic dimension is overall informative in L1 – influence perception of non-native sounds. Critically, such principles operate at an abstract level given that they apply across sound groups that are acoustically and functionally very distinct. To account for these results we developed a novel approach within a hierarchical inductive inference framework. We proposed viewing perceptual reorganization in infancy as the consequence of learners' hierarchical inductive inferences about the structure of the language's sound system. That is, we argued that infants not only acquire the specific phonetic category inventory of their native language, but also draw higher-order generalizations over the set of those categories. One such generalization captures the overall informativity of phonetic dimensions for sound categorization in L1. We further argued that this re-conceptualization of perceptual reorganization has consequences for non-native speech perception: perceptual sensitivities of naïve listeners emerge from higher-order generalizations formed during the L1 acquisition, rather than from mappings of non-native sounds onto native-language phonetic categories. We believe that this accounts contributes new insights to our understanding of perceptual reorganization and non-native speech perception by offering a rational perspective that has been successful in many other domains.

Acknowledgements

We thank Eric Bakovic, Klinton Bicknell, Sarah Creel, Tamar Gollan, Dennis Norris, Amy Perfors, and Arthur Samuel for useful feedback. We are also grateful to the associate editor, Ocke-Schwen Bohn, and two anonymous reviewers, whose comments and suggestions have been extremely helpful. This research was supported by NIH Training Grant T32-DC-000041 from the Center for Research in Language at UC San Diego to BP and NIH Training Grant T32-DC000035 from the Center for Language Sciences at University of Rochester to B.P.

Appendix

Table ***Mean and standard deviation of self-reported participant characteristics***

	Korean				Vietnamese				Cantonese				Mandarin			
	L1-dominant N=12		Eng-dominant N=12		L1-dominant N=7		Eng-dominant N=17		L1-dominant N=12		Eng-dominant N=12		L1-dominant N=12		Eng-dominant N=12	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age	21	3.0	20	1.0	20	1.2	20	1.4	20	0.9	20	1.5	21	1.4	20	1.3
Age when immigrated to the US	14	5.3	4	5.3	5	4.5	0	0.3	14	4.4	2	4.4	12	4.4	4	4.3
Self-rated L1 proficiency (0-none, 10-perfect)	9.1	1.0	7.4	2.1	8.1	0.9	7.0	0.9	9.2	1.3	8.3	1.2	9.4	1.0	8.1	1.0
% time current L1 exposure	44	13.8	34	17.1	25	14.7	20	12.2	41	24.6	33	8.9	46	14.7	16	11.8
L1 use w/family (0-never, 10-always)	9.9	0.3	8.6	1.5	8.7	1.8	8.1	1.8	9.1	1.9	9.6	0.8	9.6	0.9	8.6	1.7
L1 use w/friends (0-10)	6.8	1.7	4.0	2.3	2.6	2.5	1.7	1.8	7.1	3.1	4.0	2.1	5.5	3.2	2.8	2.3
% time preferred L1 use	61	15.6	35	21.5	46	9.4	26	14.0	58	25.1	33	9.4	56	24.7	26	15.1
Age when began regular Eng exposure	10	4.7	4	3.8	6	3.5	3	1.8	5	2.8	4	1.7	8	3.6	5	2.7
Self-rated Eng proficiency	7.3	1.1	9.4	0.8	7.7	0.49	9.4	0.6	7.3	1.0	9.2	0.9	7.3	0.7	9.0	1.1
% time current Eng exposure	53	13.6	65	16.0	74	15.1	77	13.5	45	22.9	57	11.4	50	14.7	81	18.6
Eng use w/family (0-10)	1.0	1.76	4.3	2.4	3.6	2.9	6.1	2.1	1.1	1.2	4.0	2.9	2.3	1.9	3.3	2.1
Eng use w/friends (0-10)	5.8	2.2	9.0	1.6	9.1	1.5	9.7	0.6	7.3	2.1	8.6	2.0	7.4	1.9	9.6	0.7
% time preferred Eng use	35	16.2	63	20.9	55	10.3	69	15.6	28	20.3	51	196	43	24.2	73	15.3

[a] If born in the US, coded as 0.
[b] Mean proficiency speaking and understanding.
[c] "If you could freely choose a language to speak, what percentage of time would you choose to speak each language?"

Reference List

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. Berlin/New York: Mouton de Gruyter.

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: a window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: speech and face processing in the first year of life*. Dordrecht, The Netherlands: Klewer.

Best, C. T. (1994). The emergence of native-language phonological influence in infants: A perceptual assimilation model. In J. Goodman & H. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). Cambridge: MIT Press.

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). Timonium, MD: York Press.

Best, C. T., & Hallé, P. A. (2010). Perception of initial obstruent voicing is influenced by gestural organization. *Journal of Phonetics, 38*, 109–126.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of nonnative consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, *109*, 775–794.

Best, C. T., & Strange,W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, 20, 305–330.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. 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/Philadelphia: John Benjamins.

Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: evidence of auditory enhancement. *Journal of Phonetics*, *18*, 37–49.

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. 275–300). Timonium, 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*, 109–128.

Bohn, O. S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*, 303–328.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: the role of consonants and vowels in continuous speech processing. *Psychological Science*, *16*(6), 451–459.

Brown, C. (1997). *Acquisition of segmental structure: Consequences for speech perception and second language acquisition*. Unpublished doctoral dissertation, McGill University.

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibal (Ed.), *Second language acquisition and linguistic theory* (pp. 4–63). Malden, MA: Blackwell.

Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Science*, *10*(7), 335–344.

Choi, D.-I. (1995). Korean "tense" consonants as geminates. *Kansas Working Papers in Linguistics*, *20*, 25–38.

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

Cristià, A. & Seidl, A. (2008). Is infants' learning of sound patterns constrained by phonological features? *Language Learning and Development*, *4*(3), 203–227.

Cristià, A., Seidl, A., & Francis A. L. (2011). Phonological features in infancy. In G. N. Clements & R. Ridouane (Eds.), *Where do phonological features come from? Cognitive, physical and developmental bases of distinctive speech categories* (pp. 303–326). Amsterdam/Philadelphia: John Benjamins.

Curtin, S., Byers-Heinlein, K., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics, 39*, 492–504.
de Boer, B., & Kuhl, P. K. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustic Research Letters Online*, *4*(4), 129–134.
Dietrich, C., Swingley, D., & Werker, J. F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences*, *104*, 16027–16031.
Eimas, P. D. (1978). Developmental aspects of speech perception. In R. Held, H. W. Leibowitz, & H. L. Teuber (Eds.), *Handbook of sensory physiology* (Vol. 8). Berlin: Springer-Verlag.
Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*, 551–585.
Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009a). Learning phonetic categories by learning a lexicon. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2208–2213). Austin, TX: Cognitive Science Society.
Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009b). The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, *116*(4), 752–782,
Feldman, N. H., Griffiths, T. L., Goldwater, S., & Morgan, J. L. (2013). A role for the developing lexicon in phonetic category acquisition. *Psychological Review*, *120*(4), 751–778.
Flege, J. E. (1988). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human communication and its disorders: a review – 1988* (pp. 224–401). Norwood, NJ: Ablex Publishing.
Flege, J. E. (1992). The intelligibility of English vowels spoken by British and Dutch talkers. In R. D. Kent (Ed.), *Intelligibility in speech disorders: theory, measurement, and management* (pp. 157–232). Amsterdam, The Netherlands: John Benjamins.
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. 229–273). Timonium, MD: York Press.
Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics, 15*, 67–83.
Gerken, L. (2010). Infants use rational decision criteria for choosing among models of their input. *Cognition*, *115*, 362–366.
Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L" and “R". *Neuropsychologia*, *9*, 317–323.
Goudbeek, M., Cutler, A., & Smits, R. (2008). Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication*, *50*, 109–125.
Hahm, H.-J. (2007). The effects of following vowel on Korean fricatives. *Linguistic Research*, *24*(1), 57–82.
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*, 281–306.
Hancin-Bhatt, B. J. (1994). Segment transfer: A consequence of a dynamic system. *Second Language Research*, *10*, 241–269.
Hayes, R. 2002: The perception of novel phoneme contrasts in a second language: a developmental study of native speakers of English learning Japanese singleton and geminate consonant contrasts. Coyote Papers 12, 28–41.
Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, *24*, 5–33.
Heeren, W. F. L., & Schouten, M. E. H. (2008). Perceptual development of phoneme contrasts: how sensitivity changes along acoustic dimensions that contrast phoneme categories. *Journal of the Acoustical Society of America*, *124*(4), 2291–2302.
Jongman, A., Wang, Y., Moore, C. and Sereno, J. (2006). Perception and production of Mandarin tone. In Li, P., Tan, L.H., Bates, E., and Tzeng, O.J.L. (Eds.), *Handbook of East Asian Psycholinguistics* (Vol. 1: Chinese). Cambridge, UK: Cambridge University Press.
Kao, D. L. (1971). *Structure of the syllable in Cantonese*. The Hague: Mouton.

Kawahara, S. (2007). Sonorancy and geminacy. In *University of Massachusetts Occasional Papers in Linguistics 32: Papers in Optimality III* (pp. 145–186). Amherst: GLSA.
Kim, Y.-S. (2002). On non-moraic geminates. *Studies in Phonetics*, *8*(2), 187–200.
Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, *50*(2), 93–107.
Kuhl, P. K. (1992). Psychoacoustics and speech perception: internal standards, perceptual anchors, and prototypes. In L. A. Werner & E. W. Rubel (Eds.), *Developmental psychoacoustics* (pp. 293–332). Washington, DC: American Psychological Association.
Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, *4*(6), 812–822.
Kuhl, P. K. (2000). Language, mind, and brain: Experience alters perception. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences (2nd ed.)* (pp. 99–115). Cambridge, MA: MIT Press.
Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831–843.
Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society*, *363*, 979–1000.
Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 121–154). Timonium, MD: York Press.
Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, UK; Cambridge, MA: Blackwell.
Lee, H. B. (1999). Korean. In *Handbook of the International Phonetic Association* (pp. 120–123). Cambridge: Cambridge University Press.
Lin, H. (2001). *A grammar of Mandarin Chinese*. München: Lincom Europa.
Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: a user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
Magen, H. S., & Blumstein, S. E. (1993). Effects of speaking rate on the vowel length distinction in Korean. *Journal of Phonetics*, *21*, 387–409.
Maye, J., Weiss, D., & Aslin, R. N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science*, *11*(1), 122–134.
McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English, and Estonian. *Journal of Phonetics*, *30*, 229–258.
McClaskey, C. L., Pisoni, D. B., & Carrell, T. D. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception and Psychophysics*, *34*(4), 323–330.
McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science*, *12*(3), 369–378.
Miyawaki, K., Strange, W., Verbrugge, R. R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics*, *18*, 331–340.
Mugitani, R., Pons, F., Fais, L., Werker, J. F., & Amano, S. (2008). Perception of vowel length by Japanese- and English-learning infants. *Developmental Psychology*, *45*(1), 236–247.
Pajak, B. (2013). Non-intervocalic geminates: typology, acoustics, perceptibility. In L. Carroll, B. Keffala, & D. Michel (Eds.), *San Diego Linguistics Papers 4* (pp. 2–27). San Diego, CA: UC San Diego.
Pajak, B., Creel, S. C., & Levy, R. (in prep.). Difficulty in learning similar-sounding words: a developmental stage or a general property of learning? Manuscript in preparation for journal submission.
Pajak, B. & Levy, R. (2011a). How abstract are phonological representations? Evidence from distributional perceptual learning. In *Proceedings of the 47th Annual Meeting of the Chicago Linguistic Society*. Chicago, IL: University of Chicago.
Pajak, B. & Levy, R. (2011b). Phonological generalization from distributional evidence. In L. Carlson, C. Hölscher & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the*

Cognitive Science Society (pp. 2673–2678). Austin, TX: Cognitive Science Society.

Pajak, B., Bicknell, K., & Levy, R. (2013). A model of generalization in distributional learning of phonetic categories. In V. Demberg & R. Levy (Eds.), *Proceedings of the 4th Workshop on Cognitive Modeling and Computational Linguistics* (pp. 11–20). Sofia, Bulgaria: Association for Computational Linguistics.

Perfors, A., & Dunbar, D. (2010). Phonetic training makes word learning easier. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1613–1618). Austin, TX: Cognitive Science Society.

Perfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language*, *37*, 607–642.

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.

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, *39*(3), 484–494.

Sawicka, Irena. 1995. Fonologia. In H. Wróbel (Ed.), *Gramatyka współczesnego języka polskiego. Fonetyka i fonologia* (pp. 105–195). Kraków: Instytut Języka Polskiego PAN.

Schouten, M. E. H., & van Hessen, A. J. (1992). Modeling phoneme perception. I: Categorical perception. *Journal of the Acoustical Society of America*, *92*(4), 1841–1855.

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

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics*, *36*(2), 131–145.

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: the re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 153–191). Amsterdam/Philadelphia: John Benjamins.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. (2011). How to grow a mind: statistics, structure, and abstraction. *Science*, *331*, 1279–1285.

Trubetzkoy, N. (1939/1969). *Principles of phonology [Grundzu□ge der Phonologie]*. Berkeley, CA: University of California Press.

Tseng, C.-Y., Massaro, D. W., & Cohen, M. M. (1986). Lexical tone perception in Mandarin Chinese: Evaluation and integration of acoustic features. In H. S. R. Kao & R. Hoosain (Eds.), *Linguistics, psychology, and the Chinese language* (pp. 91–104). Hong Kong, China: Centre of Asian Studies, University of Hong Kong.

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences*, *104*(33), 13273–13278.

Werker, J. F. (1989). Becoming a native listener. *Scientific American*, *77*(1), 54–59.

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

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49–63.

Winn, M., Blodgett, A., Bauman, J., Bowles, A., Charters, L., Rytting, A., & Shamoo, J. (2008). Vietnamese monophthong vowel production by native speakers and American adult learners. In *Proceedings of Acoustics '08* (pp. 6125–6130).

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: distributional learning in a miniature language. *Cognitive Psychology*, *56*, 165–209.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245–272.

Zhang, L. (2011). Vowel length perception in Cantonese. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 2292–2295). Hong Kong, China.