Difficulty in learning similar-sounding words: a developmental stage or a general property of

learning?

Bozena Pajak[1]

University of Rochester

Sarah C. Creel and Roger Levy

University of California, San Diego

9403 words (text + references + appendix)

Author Note

[1] Address correspondence to Bozena Pajak, Department of Brain & Cognitive Sciences, Meliora Hall, Box 270268,

University of Rochester, Rochester, NY 14627-0268 USA. Phone: (+1) 585-275-7171. Fax: (+1) 585-442-9216. E-

mail: bpajak@bcs.rochester.edu

Abstract

How are languages learned, and to what extent are learning mechanisms similar in infant native-language (L1) and adult second-language (L2) acquisition? In terms of vocabulary acquisition, we know from the infant literature that the ability to discriminate similar-sounding words does not guarantee successful word-meaning mapping (Stager & Werker, 1997). However, it is unclear whether this difficulty arises from developmental limitations of young infants (e.g., poorer working memory) or whether it is an intrinsic part of the initial word learning, L1 and L2 alike. Here we show that adults—just like young infants—have difficulty learning similar-sounding L2 words that they can nevertheless discriminate perceptually. This suggests that the early stages of word learning, whether L1 or L2, intrinsically involve difficulty in mapping similar-sounding words onto referents. We argue that this is due to an interaction between two main factors: (1) memory limitations that pose particular challenges for highly similar-sounding words, and (2) uncertainty regarding the language's phonetic categories, as these are being learned concurrently with words. Overall, our results show that vocabulary acquisition in infancy and in adulthood share more similarities than previously thought, thus supporting the existence of common learning mechanisms that operate throughout the lifespan.

Keywords: word learning, spoken word recognition, non-native speech perception, second language acquisition

Difficulty in learning similar-sounding words: a developmental stage or a general property of

learning?

Humans are able to learn languages throughout their lifespan. But how similar are the

learning mechanisms for infants acquiring their native language (L1) and adults learning a second

language (L2)? There has been little work trying to connect these two literatures, reflecting the

underlying assumption of a lack of developmental continuity in terms of language learning (see,

for example, discussion in White, Yee, Blumstein, & Morgan, 2013). Instead, infants and adults

have been assumed to use qualitatively different mechanisms to process and learn languages,

largely following the critical period hypothesis (Lenneberg, 1967; Johnson & Newport, 1989).

However, recent work has shown that while age of L2 acquisition negatively correlates with

achieved proficiency, there are signs of developmental continuity in language learning and

similarities between infant and adult acquisition (Birdsong & Molis, 2001; Hakuta, Bialystok, &

Wiley, 2003; Werker & Tees, 2005; Birdsong, 2009; White et al., 2013). For example, it has been

shown that infants and adults are similarly affected by word familiarity during lexical processing

of newly learned words (White et al., 2013). White et al. argued that these parallel results for

infants and adults might reflect common mechanisms that operate throughout development, thus

highlighting the need for greater interaction between the infant and the adult language learning

literatures.

In this paper we pursue a comparison between infant and adult language learning by

considering one aspect that is crucial at the initial stages of acquisition: the encoding of phonetic

detail during word learning. Learning words requires not only remembering a label for a given

referent, but also forming a phonetically-rich representation of that label by segmenting the word

into individual sounds. The detailed phonetic representation is especially important for similar-sounding words (e.g., *bin* vs. *pin*), because successful learning crucially relies on the ability to distinguish between the words based on subtle acoustic-phonetic cues. Thus, the learner must be able to perceptually discriminate the sounds that distinguish between the words (e.g., [b] vs. [p]), but also to ignore any irrelevant variability between instances of the same phonetic category (e.g., multiple exemplars of the word *bin*).

Indeed, we know from the infant literature that discrimination does not guarantee successful learning of similar-sounding words: despite the ability to perceptually discriminate between similar sounds (e.g., *b* and *d)*, 14-month-olds have been shown to confuse newly-learned words differentiated by those sounds (e.g., *bih* and *dih*) (Stager & Werker, 1997; Pater, Stager, & Werker, 2004), unless there is additional contextual information, or less demanding learning conditions (Swingley & Aslin, 2002; Fennell & Werker, 2003; Ballem & Plunkett, 2005; Fennell, Waxman, & Weisleder, 2007; Thiessen, 2007; Rost & McMurray, 2009; Yoshida, Fennell, Swingley, & Werker, 2009).

The initial explanation proposed for this result was a limited resource hypothesis (Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002): since attending to fine phonetic detail while learning new words is computationally very demanding, young infants—who have limited attentional and cognitive resources—might have difficulty accessing full phonetic detail when focusing their attention on learning meaning. Other explanations have emphasized the role of increased lexical competition in learning similar-sounding words (Swingley & Aslin, 2002, Swingley & Aslin, 2007), or suggested that the difficulty might arise from poorly defined phonetic category boundaries at that stage of infant development (Rost & McMurray, 2009) and limited experience with phonological categorization (Yoshida et al., 2009).

Regardless of the exact explanation, the consensus is that children outgrow this initial difficulty, and by 17-20 months of age succeed at learning new similar-sounding words (Werker et al., 2002). However, despite this acquired sensitivity to minimal differences between words in the learners' L1, phonological similarity continues to play a role in lexical processing in both older children and adults. This is indicated, for example, by robust and automatic activation of words that sound similar to the target word (e.g., Andruski, Blumstein, & Burton, 1994; Allopenna, Magnuson, & Tanenhaus, 1998; Swingley & Aslin, 2000; Magnuson, Dixon, Tanenhaus, & Aslin, 2007; White & Morgan, 2008; Mani & Plunkett, 2011). Adults are also slower at processing words that have a high neighborhood density (i.e., have a large number of minimal pairs) compared to words in sparse lexical neighborhoods (e.g., Vitevitch & Luce, 1998), and have increasing difficulty learning phonologically native-like nonsense words as the word similarity increases (Creel, Aslin, & Tanenhaus, 2006; Creel & Dahan, 2010). All these results suggest gradient effects of phonological similarity, where the encoding and the retrieval of similar-sounding minimal-pair words is impaired relative to dissimilar words.

Thus, adults are known to have difficulty learning novel similar-sounding words whose phonological form resembles their native language. These results do not, however, answer the question of whether adults are affected by phonological similarity during learning of an unfamiliar language, similarly to 14-month-old infants. We do know that adult L2 learners have extreme difficulty distinguishing—and therefore also learning—similar-sounding words that involve novel sound contrasts not found in their native language, such as *rake* vs. *lake* for native speakers of Japanese (e.g., Weber & Cutler, 2004; Hayes-Harb & Masuda, 2008; Escudero, Broersma, & Simon, 2012). The situation that we are concerned with here, however, is the case of learning similar-sounding words that the learners can nevertheless discriminate perceptually. In particular, the present paper examines two specific questions. First, is the mismatch between

discrimination and word learning a developmental phenomenon, or is it driven by the information being learned? That is, will adult L2 learners—who, like infants, are concurrently learning the language's phonological categories, but unlike infants, have vastly greater working memory capacity—show greater difficulty in word-learning tasks relative to discrimination tasks when acquiring new L2 vocabulary? Second, how does phonological similarity moderate discrimination vs. learning of similar-sounding L2 words?

If adults—just like 14-month-old infants—are found to have difficulty learning similar-sounding words that they can nevertheless discriminate perceptually, then this would provide another piece of evidence in favor of the existence of common language learning mechanisms that operate throughout development. Furthermore, a more detailed examination of the role of phonological similarity in word discrimination and learning, and how similarity interacts with different task demands, can help us shed more light onto the nature of those common mechanisms that underlie language learning.

The current study

In the current study, we examine how adults learn vocabulary in a new language that is phonologically unfamiliar (i.e., an L2), but composed of discriminable speech-sound categories. The situation of learning phonologically novel words that are similar-sounding, but that adults can nevertheless discriminate perceptually, is analogous to the situation of 14-month-old infants observed in Stager and Werker (1997). This allows us to assess whether the good-discrimination-without-learning pattern observed in Stager and Werker (1997) reflects a purely developmental phenomenon, or instead reflects general mechanisms of (language) learning. In addition, we include multiple sets of word pairs that differ in their degree of similarity, which lets us investigate how phonological similarity modulates discrimination and learning of L2 words.

We constructed a miniature language with minimal-pair words at three levels of similarity: (1) *dissimilar* (e.g., [tala]-[kenna]), (2) *similar* (e.g., [tala]-[taja]), and (3) *highly-similar*, where the words differed either in consonant *length* (e.g., [taja]-[tajja]) or in *place* of articulation between alveolo-palatal and retroflex sounds (e.g., [gotɕa]-[gotʂa]). We chose the *length* and the *place* dimensions because they have been shown to be differentially discriminable by two groups of speakers: L1-Korean and L1-Mandarin (Pajak, 2012; Pajak & Levy, in press). In particular, Korean speakers have an advantage over Mandarin speakers in discriminating consonant length contrasts, while Mandarin speakers have an advantage over Korean speakers in discriminating alveolo-palatal and retroflex consonant contrasts.[2] Therefore, we can investigate how these differential perceptual advantages on *highly-similar* word pairs compare with performance in a word-learning task. No differences in performance between the two groups were expected for *dissimilar* and *similar* items. Crucially, the language was phonologically novel to all participants in that all phonetic properties of the stimuli (e.g., voice onset time, vowel quality, stress, etc.) were taken from an unrelated language, Polish. Therefore, this scenario was more comparable to the situation of the 14-month-old infants than the studies in which adults learn phonologically native-like words.

We had two main sets of predictions. The first concerned the overall effect of phonological similarity on discrimination vs. learning of L2 words that are all discriminable by learners, either because the differences are salient (*dissimilar*, *similar*) or because a related distinction is used in the learners' L1 (*highly-similar*: *length* for Korean, *place* for Mandarin). Given the gradient acoustic similarity between the different sets of word pairs, we expect discrimination to also be gradient: best for *dissimilar* words, intermediate for *similar* words, and

---

[2] Pajak & Levy (in press) argued that these differential perceptual sensitivities follow from the fact that Korean has some length distinctions, while Mandarin has none, whereas Mandarin has some alveolo-palatal and retroflex sounds, while Korean only has alveolo-palatal but no retroflex sounds (Sohn, 1999; Lin, 2001).

poorest for *highly-similar* words. As for word learning, prior work using native-like words has shown gradience in performance that was modulated by word similarity (e.g., Creel et al, 2006; Creel & Dahan, 2010). But how does phonological similarity interact with word learning in the case of learning a new language with overall non-native phonology? If it works similarly to learning vocabulary in an L1 phonological system, then we would expect learners' performance to change as a function of similarity between the word pairs, matching the discrimination performance: best when identifying the referent in the context of two *dissimilarly*-named possible referents, intermediate for *similarly*-named referents, and poorest for *highly-similar* pairs. On the other hand, it is possible that learning words with an unfamiliar phonology is not affected by similarity in the same way that discrimination is. That is, we might expect a mismatch between the discrimination and the word-learning task: for example, gradient performance in discrimination, but no differences in performance when identifying word referents; or more exaggerated gradient effects in one task than the other.

The second set of predictions is in some sense a more focused version of the first set of predictions, as both are examining the influence of perceptual similarity. However, here we focus on the *highly-similar* words whose underlying speech sound categories are acoustically overlapping, thus being most comparable to infants' nascent speech sound categories. More specifically, we compare two groups of speakers with differential perceptual sensitivities to these word differences: Korean speakers better on *length*, and Mandarin speakers better on *place*. Therefore, we test word learning vs. discrimination of *highly-similar* words that either (a) are relatively easily discriminable because the differences are based on a phonetic dimension informative in L1 (even though the learners do not actually know any words distinguished by some of the specific sounds used, thus resembling the situation of young infants), or (b) are not

easily discriminable because the differences are based on a phonetic dimension not informative in L1. The contrasts in (a) are *length* for Korean speakers and *place* for Mandarin speakers, while the contrasts in (b) are *place* for Korean speakers and *length* for Mandarin speakers. If L2-learning adults are like L1-learning 14-month-olds in this learning situation, we should observe a difference between Korean and Mandarin speakers in the discrimination task (Korean better at *length*; Mandarin better at *place*), thus replicating previous findings, but no difference (or a much smaller difference) in the word-learning task. On the other hand, if adults can encode similar-sounding words better than young infants, then we should see comparable differences between Korean and Mandarin speakers in both discrimination and word learning. Additionally, given that word learning is arguably a more difficult task than discrimination, it is possible that we will observe individual differences in how efficiently learners make use of their perceptual abilities in the word-learning task. Therefore, we also examine potential individual differences in performance.

Method

Participants. Ninety undergraduate students at UC San Diego participated in the experiment for course credit or payment. Half were speakers of Korean, and the other half were speakers of Mandarin. We recorded participants' language background information, including self-reported proficiency in both Korean/Mandarin and English, current language exposure, as well as a measure of English proficiency through the Shipley Vocabulary Test (Shipley, 1967). Participants varied in terms of their length of residence in the USA: some were born in the US, while others immigrated at some point after birth or were international students who arrived very recently. Consequently, they varied in English proficiency. Importantly, however, they all learned

Korean or Mandarin from birth, reported high proficiency in those languages, and still used them regularly, predominantly with family. In most cases they had some high school and/or college exposure to Spanish or French. Some Mandarin speakers were also familiar with Taiwanese, mostly through family exposure. All participants reported no history of speech or hearing problems. We collected individual measures of participants' nonverbal IQ using the Matrices subtest of the Kaufman Brief Intelligence Test (KBIT-2, Kaufman & Kaufman, 2004). All participant characteristics are shown in Table 1.

Table 1. Individual measures: participants in the discrimination vs. the word-learning task.

MEASURE	DISCRIMINATION TASK PARTICIPANTS		WORD-LEARNING TASK PARTICIPANTS		T-TEST
	mean	sd	mean	sd	
Age	20	1.6	21	2.0	t(84.1)=-1.23, p=.22
L1 proficiency: speaking[a]	8.1	1.7	7.8	1.9	|t|<1
L1 proficiency: understanding[a]	8.3	1.6	8.3	1.5	|t|<1
% time current L1 exposure	33	19.0	30	19.3	|t|<1
Age when began regular English exposure	6.3	4.0	5.5	4.1	|t|<1
Age of arrival in the USA	8.6	7.0	7.5	7.5	|t|<1
Length of residence in the USA[b]	11.6	7.6	13.1	6.7	|t|<1
English proficiency: speaking[a]	8.1	1.7	8.3	1.9	|t|<1
English proficiency: understanding[a]	8.4	1.4	8.8	1.5	t(78.8)=-1.24, p=.22
English Vocabulary Test (% correct)	72	10.8	72	13.7	|t|<1
% time current English exposure	66	19.1	69	19.5	|t|<1
Nonverbal IQ test (% correct)	88	7.0	88	7.8	|t|<1

[a] On a 0-10 scale (0-none, 10-perfect).
[b] If born in the USA, coded as 0.

In order to avoid potential carryover effects from one task to the other task, we chose to test discrimination and word learning in a between-participant design. Fifty-four participants were assigned to the word-learning task, and 36 to the discrimination task. In each group, half were speakers of Korean, and the other half were speakers of Mandarin. Comparing participants assigned to the discrimination vs. the word-learning task revealed no significant differences in any of the measures we collected, as shown in Table 1.

Materials. The materials consisted of 16 bisyllabic consonant-vowel-consonant (CVCV) nonce words (see Table 2), where each was in a minimal pair differing only in the middle consonant (a subset of contrasts tested by Pajak, 2012; Pajak & Levy, in press). The minimal pairs differed either in *length* (a short vs. a long middle consonant) or *place* of articulation (an alveolo-palatal vs. a retroflex sibilant consonant in the middle position).[3] The materials were constructed using the sound inventory and other phonological properties of Polish, and were recorded by a phonetically-trained Polish native speaker.

---

[3] We chose the middle consonants in our stimuli in such a way that half of the corresponding sound distinctions had their analog in the listener's L1, and the other half did not. This was done in order to allow for a comparison in performance between distinctions that were relatively familiar to our participants from their L1s vs. completely unfamiliar distinctions that yet varied along familiar dimensions. In the previous work on *length* and *place* discrimination by Korean and Mandarin speakers (Pajak, 2012; Pajak & Levy, in press), Korean speakers outperformed Mandarin speakers on discriminating all *length* contrasts, whether familiar or not, while Mandarin speakers outperformed Korean speakers on discriminating all *place* contrasts (note that the stimuli in that study were also based on the Polish *length* and *place* contrasts). However, there was a trend in that earlier study for both groups to perform slightly better at the distinctions familiar from their L1s compared to the unfamiliar distinctions that varied along familiar dimensions (e.g., familiar [m]-[mm] > unfamiliar [j]-[jj] for Korean speakers). Therefore, it is possible that a similar difference will hold in a word learning task: that is, both Korean and Mandarin speakers will be better at learning similar-sounding words that include familiar categories than those that include unfamiliar categories that nevertheless vary along a familiar dimension (see Table 2 for the list of word contrasts based on (a) familiar categories, where the specific distinction exists in Korean/Mandarin, and (b) unfamiliar category contrasts, but familiar phonetic dimensions). In order to obtain enough power for such a comparison, but at the same time keeping the total number of words relatively small in order to assure their learnability in a single experimental session, we decided to focus our analysis on *length* words (12 words in total), and included a much smaller number of *place* words (4 words in total). No such difference was borne out in the current results.

The inventories of Korean and Mandarin include some sound distinctions along the dimensions of *length* and *place*, respectively, that are similar but not identical to the distinctions used in the experiment. Korean employs the dimension of *length*, distinguishing between short and long sounds, but not the dimension of *place*. We found previously that Korean speakers are better than Mandarin speakers at discriminating consonant *length* contrasts, while Mandarin speakers are better than Korean speakers at discriminating alveolo-palatal vs. retroflex *place* contrasts (Pajak, 2012; Pajak & Levy, in press). This follows from the fact that Korean has some length distinctions, while Mandarin has none; whereas Mandarin has some alveolo-palatal and retroflex sounds, while Korean does not (Sohn, 1999; Lin, 2001).

Table 2. Stimuli (in IPA).

LENGTH WORDS			PLACE WORDS		
short	*long*	*specific distinction exists in Korean?*	*alveolo-palatal*	*retroflex*	*specific distinction exists in Mandarin?*
tala	talla	yes(?)[a,b]			
kema	kemma	yes[b]	gotɕa	gotʂa	yes[c]
kena	kenna	yes[b]			
diwa	diwwa	no, but familiar dimension			
difa	diffa	no, but familiar dimension	goʑa	goʐa	no, but familiar dimension
taja	tajja	no, but familiar dimension			

[a] In Korean, intervocalic [l] is realized as a flap.
[b] In Korean, the long sound results almost exclusively from phonological assimilation.
[c] While the general distinction exists, the exact place of articulation in Mandarin is different for both the alveolo-palatal and the retroflex, and the distinction is only allophonic.

More specifically, Korean uses length distinctions mostly on vowels (e.g., [pul] 'fire' vs. [pu:l] 'blow'), but some long consonants ([ll], [nn], [mm]) arise from phonological assimilation

processes (Sohn, 1999), and Korean tense obstruents ([p'], [t'], [k'], [s'], [tɕ']) have sometimes

been analyzed as long (Choi, 1995). In terms of place of articulation, Korean has some alveolo-

palatal sounds ([ɕ], [tɕ]), but no retroflex sounds, thus lacking the *place* contrast as defined in this

paper.

Mandarin, on the other hand, has both alveolo-palatal and retroflex sounds that are

distinguished by spectral shape in the frication noise (the *place* dimension), but does not use the

length dimension. In particular, Mandarin has voiceless alveolo-palatals ([ɕ], [tɕ]) and retroflexes

([ʂ], [tʂ]) as allophones of the same phonemic category. In addition, the voiced retroflex fricative

([ʐ]) is a between-speaker variant of the retroflex approximant ([ɻ]). Other voiced sibilants are

assumed to be absent because Mandarin has obstruent distinctions in aspiration, not in voicing

(Lin, 2001). Note, however, that the analogous place distinction in Polish, which we used in the

stimuli, is not exactly the same as the one in Mandarin, differing somewhat in the placement of

the tongue tip.

The materials were recorded in a soundproof booth by a phonetically-trained native

speaker of Polish. There were 10 tokens recorded for each word. For *length* words, two tokens of

each word with long consonants were chosen for the experiment. Subsequently, words with short

consonants were created by shortening the tokens with long consonants in a way that, for each

word and each recording, the naturally-recorded long consonant was reduced to half its duration

so as to maintain a constant 2:1 duration ratio (cross-linguistically, the long-to-short consonant

ratio varies between 1.5 to 3; Ladefoged and Maddieson, 1996). For *place* words, given that the

alveolo-palatal vs. retroflex distinction is intrinsically already very subtle, even in natural speech

(Nowak, 2006), we used natural recordings of both alveolo-palatals and retroflexes with no

additional manipulations. Two tokens each were chosen for the experiment with the goal of maximizing the similarity between the words in minimal pairs with regards to how vowels were pronounced, but at the same time choosing tokens with clearly enunciated sibilants. This was a departure from how the stimuli were constructed by Pajak (2012) and Pajak and Levy (in press), where both alveolo-palatals and retroflexes were spliced into an identical word frame. Pajak and Levy's procedure removed one of the cues to the contrast (the formant transition into the following vowel), thus making it extremely subtle. In the current study, we left this cue intact so that stimuli in the word-learning task were not overly difficult.

These auditory stimuli were used for both the discrimination and the word-learning task. For the word-learning task, each word was paired with a picture of a different kind of mushroom (see two examples in Figure 1), which were chosen in order to include objects that were unfamiliar to our participants, but not so unfamiliar that participants would find them bizarre and hard to remember. We selected pictures that varied in shape and color so as to maximize visual differences between them. We created four different one-to-one word-to-picture mappings that were counterbalanced between participants in order to make sure that the results were not driven by any peculiarities in the mappings we chose.

Figure 1. Example of a screen shot from the word-learning task.

Procedure. Participants sat in front of a computer, and responded by using a mouse. They

were instructed that in this experiment they would be listening to a novel language, and,

specifically, either (i) learn to distinguish this language's sounds (in the discrimination task), or

(ii) learn the language's words for different types of mushrooms (in the word-learning task). The

experiment was completed in a single session, and each participant took part in only one of the

tasks. The discrimination and the word-learning tasks were made equal in the total auditory

exposure to each stimulus in order to keep them as parallel as possible.

Discrimination task. Discrimination was tested in an ABX task. In each trial, three words

were presented auditorily through headphones: A *[500ms]* B *[750ms]* X (e.g., [taja] [tajja]

[taja]). The task was to assess whether X sounded more like A or more like B. The X word was

always acoustically different from both A and B words to make sure that the matching of X onto

A or B was not based on pure acoustical identity of two tokens. This was a different procedure

than in Pajak (2012) and Pajak and Levy (in press), where an AX task was used instead. In this

study, however, we wanted to maintain a close parallel between the discrimination and the word-

learning tasks, which was achieved with the ABX procedure. The AB word order was

counterbalanced, and the trial order was randomized for each participant. There were 4 blocks,

each with 64 trials and lasting about 5 minutes. Blocks were separated by self-terminated breaks.

There were four types of trials depending on the AB contrast, as illustrated in Table 3: (i)

dissimilar word pairs (e.g., [tala]-[kenna]), which differed in all sounds but the last vowel (16

trials), (ii) *similar* word pairs (e.g., [tala]-[taja]), which shared the initial CV sequence, but the

middle consonants differed along multiple phonetic dimensions (16 trials), (iii-iv) *highly-similar*

(e.g., *length*: [tala-talla] or *place*: [gotɕa]-[gotʂa]), where the initial CV sequence was identical

and the middle consonants differed minimally, either in *length* or in *place* (32 trials per block).

Table 3. Trial types in discrimination (AB=words presented auditorily) and testing in word learning (AB=labels for visually presented pictures). For each pair, both orders of presentation were tested.

DISSIMILAR	SIMILAR	HIGHLY-SIMILAR	
A B	A B	A B	
tala-kenna	diwa-difa	diwa-diwwa	
talla-goża	diwwa-diffa	difa-diffa	
taja-gotça	taja-tala	taja-tajja	*length*
tajja-kema	tajja-talla	tala-talla	
diwa-kemma	kema-kena	kema-kemma	
diwwa-goża	kemma-kenna	kena-kenna	
diffa-kena	gotṣa-goża	goża-goża	*place*
difa-gotṣa	gotça-goża	gotça-gotṣa	

Word learning task. In this task, participants learned to associate words with pictures of mushrooms. There were 4 training blocks (each with 128 trials, about 10-15 minutes long) and 4 testing blocks (each with 64 trials, about 5 minutes long), interleaved. Thus, each train+test combination contained $128 + 64 = 192$ exposures to the words, the same as a block of the discrimination task ($64 \times 3 = 192$). Blocks were separated by self-terminated breaks. In each trial, two pictures were presented on a computer screen (see Figure 1), and a word was played through headphones with a delay of 500 ms. Participants were asked to click on the picture that they

thought went with the word. In training, feedback was provided following the response in the form of the correct picture staying on the screen. A mouse click triggered the start of the next trial. Because participants were learning via feedback presented after each response, the early responses were necessarily random. Participants were told to guess at first, and that through feedback they would eventually learn the correct word-to-picture mappings. In testing, no feedback was provided.

The training trial types consisted of picture pairs that were always associated with *dissimilar* word pairs (e.g., [taja]-[diwa], [gotɕa]-[kemma]) so that participants were not directly alerted to the distinctions of interest. The testing trial types were always different from the training trials, and were completely analogous (in form and number) to trials in the discrimination task, as illustrated in Table 3. Each picture pair in the word-learning task was an analog of an AB word pair in the discrimination task, and the auditorily presented word in the word-learning task corresponded to the X word in the discrimination task. The picture position was counterbalanced. The trial order was pseudo-randomized: we created four randomized lists, and then altered them manually so that the same word was never repeated in two consecutive trials. Furthermore, the minimal-pair trials were always separated by at least two other trials. Each participant heard each list once, with a different list for each block. Block order was counterbalanced across participants.

Results

We investigated two main questions in our data analysis, based on the two sets of predictions that guided our study's design. First, how does phonological similarity moderate discrimination vs. learning of similar-sounding words in an L2 with clearly non-native

phonology? Second, when testing phonologically highly similar word pairs, does good discrimination predict good learning of similar-sounding words in adult L2 learners, or is there a disconnect between discrimination and word learning like that observed in infant L1 learners (Stager & Werker, 1997)?

To answer these questions, we analyzed accuracy scores from both discrimination and testing in word learning with mixed-effects logit models (Jaeger, 2008). We included random intercepts for participants and items, and random slopes for participants and items for all effects of interest (including interaction effects) that were manipulated within participants or within items. We controlled for participants' nonverbal IQ and L1 proficiency and use (through a combined score on various measures) by adding them as fixed effects to the models. All the binary and continuous predictor variables were centered; three-level variables were coded using successive differences contrast coding. The reported p values came from the z statistic. For *highly-similar* trials, the difference between familiar categories vs. familiar dimensions only (as shown in Table 2, p. 12, and discussed it footnote 3, p. 11) was not significant in either task, and so we do not report it in the analysis.

Gradient phonological competition in L2 learning? We begin by addressing the second question concerning the effects of phonological similarity on discrimination and learning of L2 words. We analyzed trials from both discrimination and word-learning tasks for word pairs that were all discriminable by learners: *dissimilar* (e.g., [tala]-[kenna]), *similar* (e.g., [tala]-[taja]), and a subset of *highly-similar* pairs, depending on language group: *length* for Korean speakers (e.g., [tala-talla]) and *place* for Mandarin speakers (e.g., [gotɕa]-[gotʂa]). Based on prior work, we expected discrimination to be gradient: best for *dissimilar* words, intermediate for *similar* words,

and poorest for *highly-similar* words. Of most interest here was the comparison between

discrimination and the word learning in order to evaluate how phonological similarity moderates
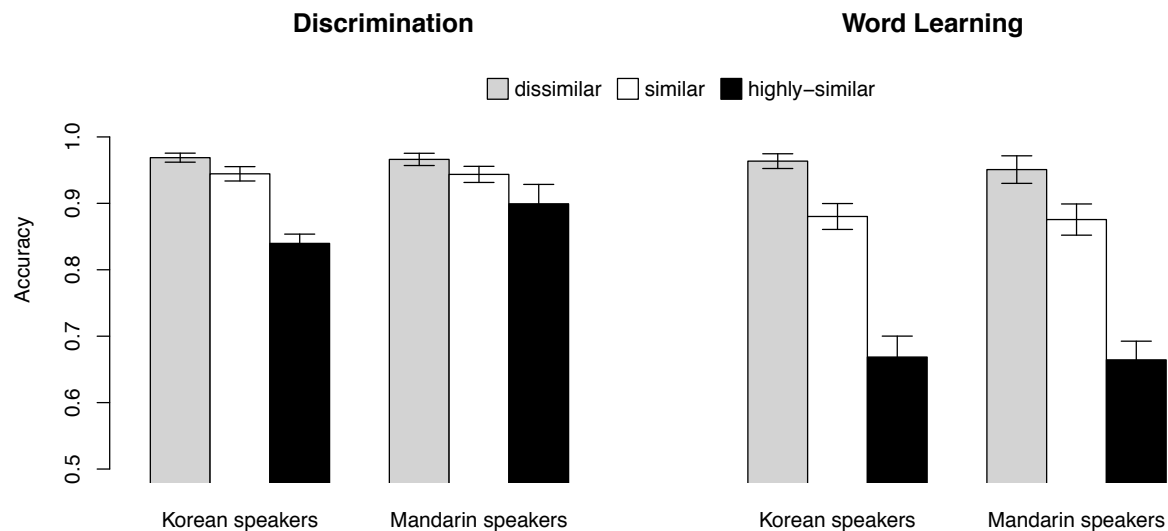
performance across different tasks.

Figure 2. Results for *dissimilar*, *similar* and *highly-similar* (*length* for Korean, *place* for

 Mandarin) trials. Accuracy scores indicate proportion of correct responses and error bars are

 standard errors.

The results are illustrated in Figure 2. We began the analysis by evaluating a model with

fixed effects of TASK (*discrimination, word-learning*), TRIAL-TYPE (*dissimilar, similar, highly-*

similar) and LANGUAGE (*Korean, Mandarin*). TRIAL-TYPE was coded with *similar* trials as the

reference level so that we would be able to directly compare *similar* and *dissimilar* trials, as well

as *similar* and *highly-similar* trials. As expected, there was a significant effect of TRIAL-TYPE in

that accuracy, pooled across the discrimination and the word learning tasks, varied in accordance

to the similarity between words: the responses on *dissimilar* trials were significantly higher than on the *similar* trials (*p*<.001), which in turn were higher than *highly-similar* trials (*p*<.001).

Furthermore, there was a significant interaction between TASK and TRIAL-TYPE (*p*s<.001), suggesting that accuracy on each type of trial was moderated by the task: discrimination vs. word learning. To examine this further, we directly compared performance in discrimination and word learning separately for each TRIAL-TYPE using models with the fixed effect of TASK (*discrimination*, *word-learning*). For *dissimilar* trials, there were no main effects of TASK (*p*=.59), suggesting that performance did not differ across tasks (although it is possible that an underlying difference between the tasks was masked by ceiling effects, given that the overall performance was above 95%). However, for both *similar* and *highly-similar* trials, we found significant main effects of TASK (*p*s<.001): higher overall performance in discrimination than in the word-learning task.

No other effects in the full model reached significance, including effects or interactions involving LANGUAGE (*p*s>.2). This suggests that the two language groups had similar overall response patterns on the word pairs that were predicted to be relatively well discriminable by all participants (i.e., all word pairs excluding *place* trials for Korean speakers and *length* for Mandarin speakers).

In sum, these results suggest that there is a gradient effect of phonological similarity in both word discrimination and word learning: performance decreases as similarity grows. However, this effect is moderated by the specific task: relative to discrimination, performance in word learning suffers substantially more as similarity increases.

Discrimination vs. word learning: are L2 learners like infants? Addressing whether L2

learners show a discrimination-word learning asymmetry like that observed in L1-learning

infants (Stager & Werker, 1997) entails a specific comparison between the two different *highly-

similar* trial types—*length* and *place*—for both language groups and across the two different

tasks. Given previous studies with similar stimuli (Pajak, 2012; Pajak & Levy, in press), we

expected differential discrimination of *length* and *place* contrasts by the two participant groups:

Korean speakers more accurate than Mandarin speakers on discriminating *length* trials, and

Mandarin speakers more accurate than Korean speakers on discriminating *place* trials (a

FEATURE-TYPE X LANGUAGE interaction). The question was whether this interaction would also

extend to the word-learning data.

The results are illustrated in Figure 3. We analyzed these data in a model with fixed

effects of TASK (*discrimination*, *word-learning*), FEATURE-TYPE (*length*, *place*) and LANGUAGE

(*Korean*, *Mandarin*). (See Appendix for additional analyses that include BLOCK as a fixed effect,

demonstrating that the main result of a FEATURE-TYPE X LANGUAGE interaction was consistent

throughout the experiment.) First, there was a significant main effect of TASK (p<.001):

performance was overall higher in discrimination than in word learning, indicating that the latter

task was more difficult. Furthermore, there was a significant FEATURE-TYPE X LANGUAGE

interaction (p<.001): as expected, Korean speakers performed better on *length* trials, and

Mandarin speakers performed better on *place* trials. Critically, however, there was also a

significant three-way TASK X FEATURE-TYPE X LANGUAGE interaction (p<.001), indicating that

performance as a function of FEATURE-TYPE was different across the two tasks. In order to

interpret this interaction, we continued by analyzing each task separately.

For the discrimination task (Figure 3, left), there was a significant interaction between

FEATURE-TYPE and LANGUAGE (p=.001): as predicted, Korean listeners performed better on

length trials, while Mandarin listeners performed better on *place* trials. For the word-learning

task (Figure 3, right), however, there was no interaction between the two variables (*p*=.21; nor

any main effects, *p*s>.5), suggesting that the respective perceptual advantages of Korean and

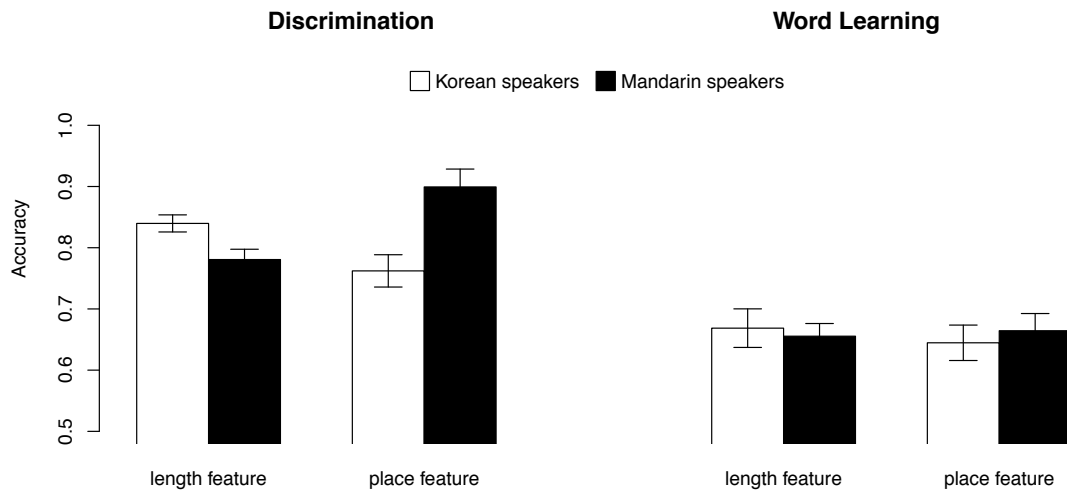Mandarin speakers did not translate into an advantage during word learning.

Figure 3. Results for all *highly-similar* trials: *length* and *place*. Accuracy scores indicate

   proportion of correct responses and error bars are standard errors.

Therefore, the three-way interaction in the main model indicates that good discrimination

does not necessarily yield better learning. Overall, these results suggest that learners did not take

full advantage of their perceptual abilities in a word-learning task. This pattern is highly similar

to 14-month old infants learning their native language (Stager & Werker, 1997), suggesting that

discordant performance between discrimination and mapping is not a developmental

phenomenon, but a more general feature of learning words in a new phonology.

Do individual differences affect use of perceptual advantages in word learning? Word learning is a complex task that involves a combination of cognitive abilities and attention. Therefore, it is possible that there is a high degree of individual differences in how efficiently learners make use of their perceptual abilities in the word-learning task. To answer that question we examined the word-learning data separately for two groups of participants: higher and lower performers.

We split participants into higher- and lower-performer groups based on their accuracy scores on *dissimilar* and *similar* trials. These trials consisted of more salient distinctions (see Table 3), and were independent from the variables of interest. The median accuracy on all these trials combined was 94.5%. There were 7 participants who scored exactly at 94.5%. We performed two separate analyses of the data, where the 7 participants were either all included in the higher-performer group or in the lower-performer group, later referred to as split-1 and split-2, respectively. The results were equivalent in both cases, but for simplicity reasons, we only illustrate the split-1 results. The distribution of participants was fairly equal across language background (Table 4). The scores on *dissimilar* and *similar* trials for both higher and lower performers are provided in Table 4. Both groups were highly accurate on these trials, but there was much more variability among lower performers, as indicated by the higher standard deviations.

Figure 4 illustrates the word-learning results for length and place trials split into higher and lower performers. By visual inspection alone, it can be seen that participants in the higher-performer group were clearly learning the minimal-pair words, as indicated by their much higher levels of accuracy. In the lower-performer group, on the other hand, participants' responses were close to chance.

Table 4. *Dissimilar* and *similar* trial scores from the word-learning task split by higher and lower performers (standard deviations are in parentheses)

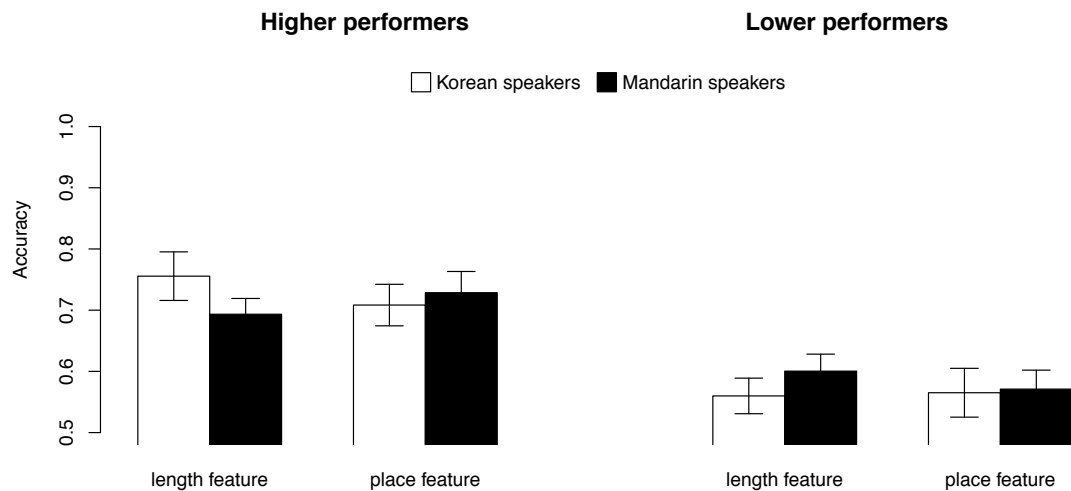| TRIAL TYPE	HIGHER PERFORMERS		LOWER PERFORMERS	
	Korean	*Mandarin*	*Korean*	*Mandarin*
	$n = 16 / 11$[a]	$n = 15 / 13$	$n = 11 / 16$	$n = 12 / 14$
dissimilar	.99 (.10) / .99 (.07)	.98 (.11) / .99 (.09)	.93 (.26) / .93 (.25)	.90 (.30) / .92 (.27)
similar	.95 (.21) / .96 (.20)	.94 (.24) / .95 (.22)	.79 (.41) / .81 (.39)	.78 (.41) / .82 (.38)

[a] Split-1 / Split-2.

Figure 4. Word-learning task results (*length* and *place* trials) split by higher and lower performers (split-1; results for split-2 were equivalent). Accuracy scores indicate proportion of correct responses and error bars are standard errors.

We analyzed these results with models with fixed effects of FEATURE-TYPE (*length*, *place*), LANGUAGE (*Korean*, *Mandarin*), and PERFORMER-TYPE (*high*, *low*), separately for split-1 and split-2. In both cases we found a significant effect of PERFORMER-TYPE (ps<.001), reflecting higher accuracy in the higher-performer group than in the lower-performer group. Critically, there were also significant three-way interactions (in both split-1 and split-2) between FEATURE-TYPE, LANGUAGE, and PERFORMER-TYPE (ps<.05), indicating distinct response patterns for Korean vs. Mandarin speakers on *length* and *place* trials depending on their overall success rate in learning, as measured by their accuracy on *dissimilar* and *similar* trials.

To assess the nature of the 3-way interaction, we analyzed the effects of FEATURE-TYPE and LANGUAGE separately for each PERFORMER-TYPE. Higher performers showed a pattern more consistent with taking advantage of their perceptual biases: Korean speakers were more accurate on *length* trials than Mandarin speakers, but the reverse was true on *place* trials, as indicated by significant interactions (in both split-1 and split-2) between FEATURE-TYPE and LANGUAGE (ps<.05).[4] The lower performers, on the other hand, showed no significant interactions between FEATURE-TYPE and LANGUAGE (and no other significant effects).

Therefore, analyzing the word-learning data split by accuracy on more salient distinctions revealed a difference between higher and lower performers, with the former group taking more advantage of their perceptual abilities than the latter group. Was this higher-performer advantage driven by the learners' better *discrimination* abilities? If so, then lower performers in the discrimination task should also lack the L1-feature advantage. However, we found no evidence in support of this hypothesis. With a similar split of participants (Figure 5), the FEATURE-TYPE X

---

[4] One might wonder if these results only showed up after better learners were alerted to the nature of the task. However, the interaction was already numerically present (marginal on split-1, p=.08; split-2, p=.11) in the first block of testing. This suggests that the effect of FEATURE-TYPE was present before test trials "tipped off" participants to the presence of the *length* and *place* minimal pairs, since the pictures corresponding to these minimal pairs were never shown together in training.

LANGUAGE interactions were significant for both higher-performer and lower-performer groups (p*s*<.01). Therefore, these results suggest that the higher-performer advantage in the word-learning task is driven by something other than better discrimination abilities. We outline possible explanations in the discussion.
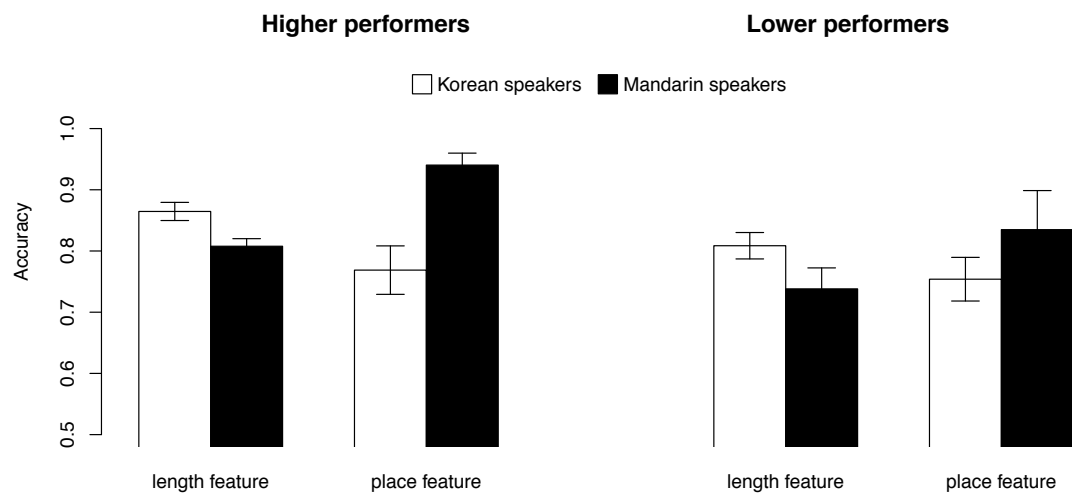

Figure 5. Discrimination task results (*length* and *place* trials) split by higher and lower performers. Accuracy scores indicate proportion of correct responses and error bars are standard errors.

Discussion

In this paper we asked two specific questions: (1) do adult L2 learners have difficulty learning similar-sounding words that they can nevertheless discriminate, just like 14-month-old infants?, and (2) how does phonological similarity moderate discrimination vs. learning of similar-sounding L2 words? To answer these questions we asked participants of two different

language backgrounds either to discriminate similar-sounding words in a new language, or to

map similar-sounding words onto novel referents. We used three levels of word similarity:

dissimilar word pairs (differing in all sound segments but the last vowel), *similar* word pairs

(same first syllable, but salient differences in the middle consonant), and *highly-similar* word

pairs (same first syllable and only subtle differences in the middle consonant). Furthermore, the

highly-similar word pairs consisted of two distinction types that the two participant groups had

differential perceptual sensitivities to (as shown by previous work, and confirmed by our

discrimination task). In particular, some distinctions varied along dimensions familiar to

participants from their L1, and others did not. Comparing the three levels of word pairs

(*dissimilar*, *similar*, *highly-similar*) let us assess the role of phonological similarity in

discrimination vs. word learning (question 2), while investigating the differences between

language groups on the two types of *highly-similar* word pairs allowed us to answer the question

regarding the potential mismatch between discrimination and learning of similar-sounding words

in adult L2 learning (question 1).

Regarding the effects of phonological similarity on discrimination vs. word learning, we

found—as in prior work on learning phonologically native-like words—gradient performance. In

both discrimination and word-learning tasks, performance was best on *dissimilar* word pairs,

intermediate for *similar* word pairs, and worst on *highly-similar* word pairs. However, we also

found a mismatch between discrimination and learning: the gradient effects were more

exaggerated in the word-learning than in the discrimination task, with performance dropping

disproportionally as word similarity increased. Therefore, our results suggest that word learning

is particularly difficult when words are similar-sounding, even when the differences are fairly

salient (as in our *similar* word pairs).

The results from *highly-similar* word pairs revealed that participants in the word-learning task were unable to effectively use their perceptual abilities word learning. However, there was some evidence that higher performers (as independently assessed by performance on unrelated trials) were somewhat more successful. This was in sharp contrast to the discrimination results, where L1-based perceptual advantages were observed for all participants. This result thus reveals an intermediate effect between failure to learn similar-sounding words (as observed for 14-month-old infants) and a full ability to use existing perceptual abilities in learning (which should mimic the discrimination data).

What factors led to better performance in learning similar-sounding words? The observed differences in the word-learning task performance for higher vs. lower performers raises the question of whether they can ultimately be attributed to individual differences in such measures as IQ or L1 proficiency. In order to investigate these potential factors, we examined more closely the individual measures collected from our participants, as shown in Table 5 (split-1 only; minor differences between split-1 and split-2 are reported in table footnotes).

A series of t-tests revealed that the higher-performer and the lower-performer groups did not significantly differ in nonverbal IQ or L1 proficiency. However, they did differ in English proficiency: higher performers reported higher overall proficiency than lower performers. Higher English proficiency was likely the consequence of the higher performers' earlier age of arrival to the USA and a longer length of residence relative to the lower-performer group. This result suggests that the higher-performer group might have been composed of early bilinguals who were perhaps more balanced in both languages (i.e., fluent in both Korean/Mandarin and English), while the lower-performer group included more speakers strongly dominant in their L1.

Table 5. Higher vs. lower performers in the word-learning task (significant differences in **bold**).

Measure	Higher Performers		Lower Performers		T-Test
	mean	sd	mean	sd	
Age	20	1.6	21	2.3	$\|t\|<1$
L1 proficiency: speaking[a]	7.5	1.9	8.1	1.9	t(47.3)=1.02, p=.31
L1 proficiency: understanding[a]	8.2	1.4	8.4	1.6	$\|t\|<1$
% time current L1 exposure	**23**	**16.6**	**39**	**19.2**	**t(43.3)=3.03, p<.01**** [c]
Age when began regular English exposure	4.8	3.9	6.3	4.1	t(46.4)=1.34, p=.18
Age of arrival in the USA	**5.6**	**6.5**	**10.2**	**8.0**	**t(41.1)=2.19, p<.05***
Length of residence in the USA[b]	**14.9**	**6.2**	**10.8**	**6.7**	**t(45.4)=-2.23, p<.05***
English proficiency: speaking[a]	**8.9**	**1.4**	**7.4**	**2.1**	**t(35.4)=-2.83, p<.01****
English proficiency: understanding[a]	**9.2**	**1.2**	**8.1**	**1.7**	**t(38.2)=-2.60, p<.05***
English Vocabulary Test (% correct)	75	12.7	68	14.1	t(44.4)=-1.69, p=.10
% time current English exposure	**75**	**16.2**	**62**	**21.0**	**t(39.8)=-2.38, p<.05*** [d]
Nonverbal IQ test (% correct)	89	7.6	86	7.9	t(46.5)=-1.26, p=.21

[a] On a 0-10 scale (0-none, 10-perfect).
[b] If born in the USA, coded as 0.
[c] Effect only marginal in split-2: t(51.4)=1.74, p=.087.
[d] Effect not significant in split-2: t(51.8)=-1.36, p=.181.

Therefore, one possible explanation for the better performance of some participants in the word-learning task is that this is an effect of early bilingualism: more balanced bilinguals might be better equipped to use their perceptual abilities when learning novel words. This is consistent with other findings suggesting that early simultaneous bilingual adults have a general cognitive advantage over monolinguals for novel word learning, whether phonologically familiar or unfamiliar, and independently of phonological memory capacity (Kaushanskaya & Marian, 2009,

Kaushanskaya, 2012). An alternative explanation is that higher English proficiency of some participants might be an indicator (more precise than IQ) of better learning skills that led to superior performance in our word-learning task.

Why is word learning difficult?

Irrespective of individual differences, the results reported in this paper suggest that there is something inherently hard about the early stage of word learning that precludes attention to fine phonetic detail that is otherwise available during phonetic processing – this is true for adult L2 learners in the same way it seems to be true for 14-month-old L1 learners. Together with the results of White et al. (2013), showing that infants and adults are similarly affected by lexical familiarity during word learning, these findings provide evidence that there might be common learning mechanisms operating throughout development. But what are these mechanisms, and what is the source of difficulty in learning similar-sounding words?

One answer is that learning novel words is simply a highly complex task that leads to information processing overload. This explanation was originally proposed by Stager & Werker (1997) and Werker et al. (2002), who argued that young infants struggle with accessing sufficient phonetic detail in their lexical representations to successfully differentiate between words that are phonetically highly similar. This means that—relative to task difficulty—only individuals with better attentional or general cognitive abilities might effectively manage concurrent information at multiple levels of processing. This would explain why older, 17 to 20-month-old infants outperform 14-month-olds, as well as why adults with some general cognitive advantages (such as early simultaneous bilinguals; e.g., Bialystok, 1999) might outperform other adults, as with the results reported in this paper.

Note, however, that it does not seem to be the case that word learning is simply overall harder than discrimination – our results indicate that for highly dissimilar word pairs, performance is equally high in both tasks (although an underlying difference might be hidden due to ceiling effects). Instead, performance in word learning drops disproportionately as the word pairs get phonologically more similar, suggesting that the task is especially taxing when detailed phonetic representations are needed to distinguish between words, even when the words are perceptually easily discriminable (as our *similar* word pairs). This is consistent with other work showing that confusability between newly-learned—but phonologically native-like—words is modulated by the phonetic distance between the sounds that differentiate between them (e.g., Creel et al., 2006; Creel & Dahan, 2010; White et al., 2013). Thus, the overall difficulty of word learning in general lies most likely not (just) in the task itself, but (also) in pinning down the correct label/referent pairings.

What might then contribute to the difficulty of learning the correct label/referent pairing? One potential factor is that beginner learners, both infant and adult, might have noisy phonetic representations, reflecting low confidence in the fidelity of phonetic encoding of the newly-learned words or in the exact location of phonetic category boundaries (Rost & McMurray, 2009; Yoshida et al., 2009). Another factor might involve memory limitations, which pose particular challenge for highly similar-sounding words. Below we describe in more detail how these factors might lead to the results observed in the experiments reported in this paper.

A conceptual model of mapping difficulties

Learners acquiring words need to rely on their memory representations of label/referent pairs, where each label can be described as a sequence of sounds sampled from the language's phonetic categories. Precise encoding of the label's phonetic form thus requires establishing what

categories the sounds were sampled from, a difficult task at the early stages of language learning. Word learning is then likely affected by two sources of noise: (1) noise and uncertainty associated with categorization of each individual sound, and (2) noise associated with retrieving a memory trace of the phonetic input and of the label/referent pairing. Thus, one way of thinking about the difference between the discrimination and the word-learning results in our experiment is that the quality of phonetic representations for individual input exemplars is lower for word learning than for discrimination due to heavier long-term memory demands in the word-learning task: in word learning listeners have to simultaneously keep track of the referent and try to form phonetic representations, while a discrimination task only requires comparing memory traces of phonetic input, without the need to link them to referents.

As more data are obtained over time, successful word learning requires integrating multiple memory traces to arrive at the correct label/referent pairing. This would correspond to narrowing down the effective variance around the memory representation of the label's phonetic form. In cases when the words being learned are composed of highly dissimilar sequences of sounds – that is, the label's phonetic form variance is small relative to the phonetic distance between sound categories – the learner's performance should not be impeded because a small number of samples would be sufficient to learn the distinctions among the categories. This is indeed what we observed: participants' performance was near ceiling when identifying the referent of a word in the context of two potential referents with highly dissimilar labels.

However, when the words are highly similar—that is, the phonetic form variance is large relative to the distance between sound categories—it should be much harder for learners to separate the sounds into categories and pinpoint the right label/referent pairings. This is consistent with our results, showing that learners have difficulty identifying the referent of a word in the context of two potential referents with highly-similar-yet-discriminable labels. It is

expected that much more data (i.e., more learning instances) is needed in this case before learners can accumulate a sufficient number of exemplars to learn phonetic category distinctions and form correct label/referent mappings. Furthermore, if the learner is unable to integrate the information accumulated from a number of exemplars of the label/referent pairing, he or she will have great difficulty in learning the pairing reliably at all.

In our study we observed that only some learners were able to reliably identify referents in the highly-similar-label context in the course of the experiment. What then differentiates more successful learners from the less successful ones? In the view discussed above, the critical difference between higher and lower performers is that higher performers can keep track of and/or integrate evidence from more exemplars than the lower performers can. In addition, higher performers might be able to store exemplars better (i.e., with higher phonetic fidelity) than the lower performers from the outset of language exposure, and thus be faster at forming new phonetic categories. This does not hurt lower performers when asked to identify dissimilar words, but impacts them more substantially when identifying label-referent pairings of similar-sounding words.

This model predicts that performance in learning similar-sounding words would be improved by decreasing memory demands in the word-learning task, as well as providing additional support for storing the phonetic detail of word exemplars and forming new phonetic categories. The latter idea is consistent with recent findings that auditory training on difficult phonetic distinctions improves subsequent word learning, and is particularly beneficial for learners with low pre-training auditory sensitivity (Cooper & Wang, 2013; Ingvalson, Barr, & Wong, 2013).

Implications for L2 acquisition

The results presented in this paper contribute to our understanding of L2 learning, as well as provide more practical implications for second-language teaching. The current view is that L2 phonetic category learning is largely hindered due to perceptual difficulties arising from prior acquisition of L1 phonology (e.g., Best, 1995, Flege, 1995, Kuhl & Iverson, 1995). While we do not dispute the importance of L1 influence on L2 phonological acquisition, our results suggest that at least some of the learners' difficulties in distinguishing between novel L2 sounds might be due to their introduction in the context of highly similar-sounding lexical items. In such a context, when attention is directed at forming label-referent mappings, learners might be unable to properly separate similar-sounding categories. It thus seems that the introduction of highly similar L2 phonetic categories might be more effective either outside of a word-learning task (e.g., as a focused non-native sound discrimination practice) or when learning words that are overall fairly dissimilar. The latter conclusion is independently supported by the results from infant studies, showing that prior experience with sounds in non-minimal-pair lexical contexts improves infants' ability to later learn minimal-pair words distinguished by those sounds (Thiessen, 2007). A similar conclusion can be drawn from computational modeling work, showing that non-minimal-pair lexical contexts improve distributional learning of overlapping sound categories (Feldman, Griffiths, Goldwater, & Morgan, 2013).

Conclusion

The results presented in this paper show that adults, just like young infants, have difficulty learning similar-sounding words that they can nevertheless distinguish perceptually. This parallel between infants and adults point to a common mechanism underlying the initial

stage of lexical acquisition throughout development, whether in the native language or any additional language acquired in adulthood. Together with recent results showing other parallels between infant and adult lexical acquisition (White et al., 2013), these findings highlight the necessity for greater interaction between infant and adult language learning literatures that would investigate the commonalities and the differences between native-language development in infancy and second-language learning in adulthood, thus shedding more light onto the degree of developmental continuity in language learning.

Appendix A

Learning during the task. In both discrimination and word learning it is reasonable to

expect that performance could improve with exposure to the materials and greater experience in

the test task; hence we analyzed both sets of *highly-similar* trials including test block as an

additional variable (Figure A1). Throughout the analysis BLOCK was treated as a mean-centered

continuous covariate, which is a simple way of imposing the constraint that temporal changes

should be monotonic. For both tasks we found significant main effects of BLOCK (discrimination:

$p<.05$; word-learning: $p<.001$), indicating that participants improved throughout the experiment.

There were also significant interactions between FEATURE-TYPE and BLOCK (discrimination:

$p<.05$; word-learning: $p<.01$): in word learning, the improvement was more prominent for the

length trials than for the *place* trials; the opposite seemed to be the case for discrimination – more

improvement on *place* than on *length* trials. Crucially, adding the block information revealed that

the main result – a difference between Korean and Mandarin speakers in discrimination, but not

in word learning – was consistent throughout the experiment. Figures A2 and A3 illustrate the

by-block results split by higher and lower performers in the word-learning and the discrimination
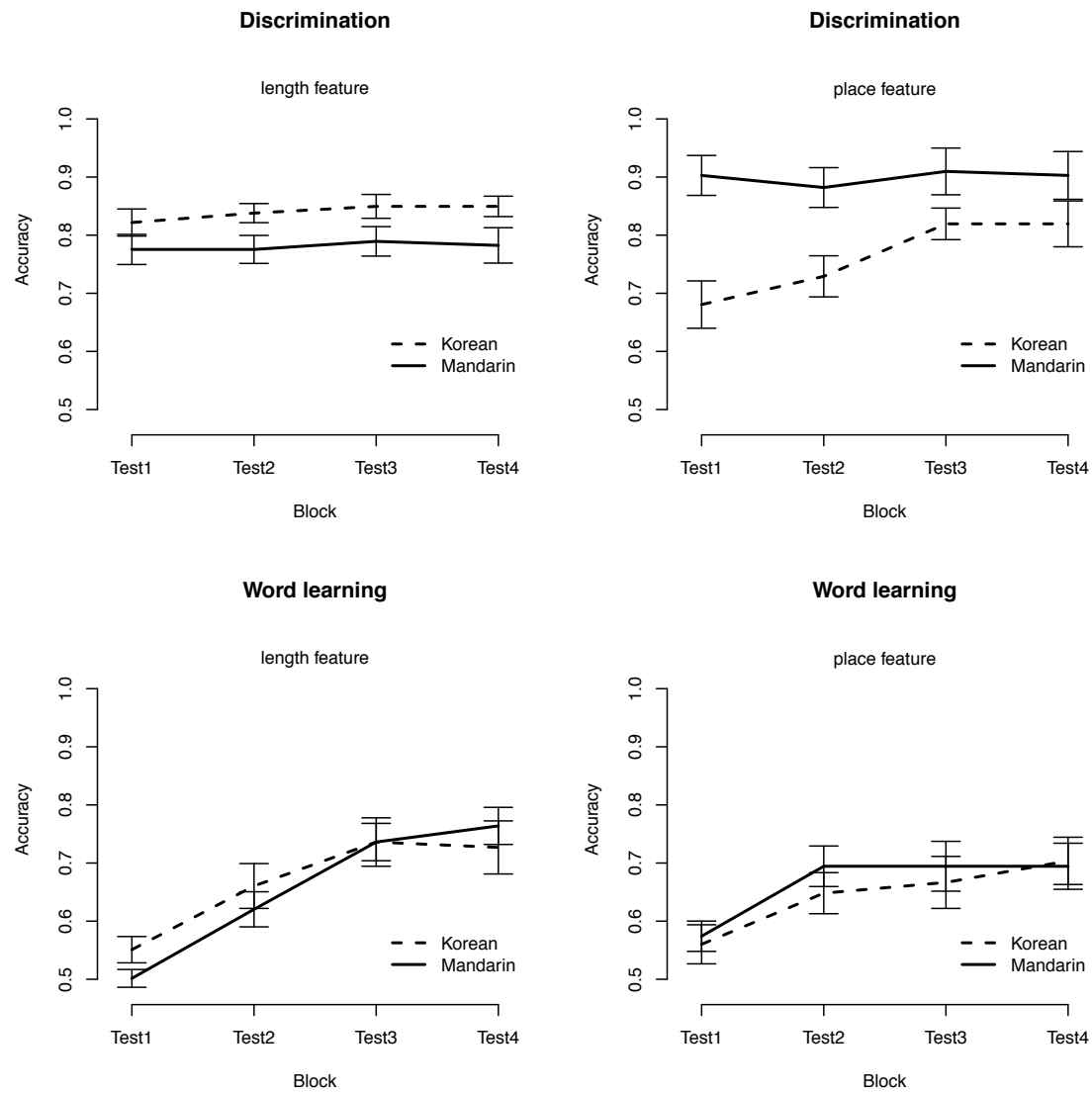
tasks, respectively.

Figure A1. Results for all *highly-similar* trials by block: *length* and *place*. Accuracy scores

indicate proportion of correct responses and error bars are standard errors.
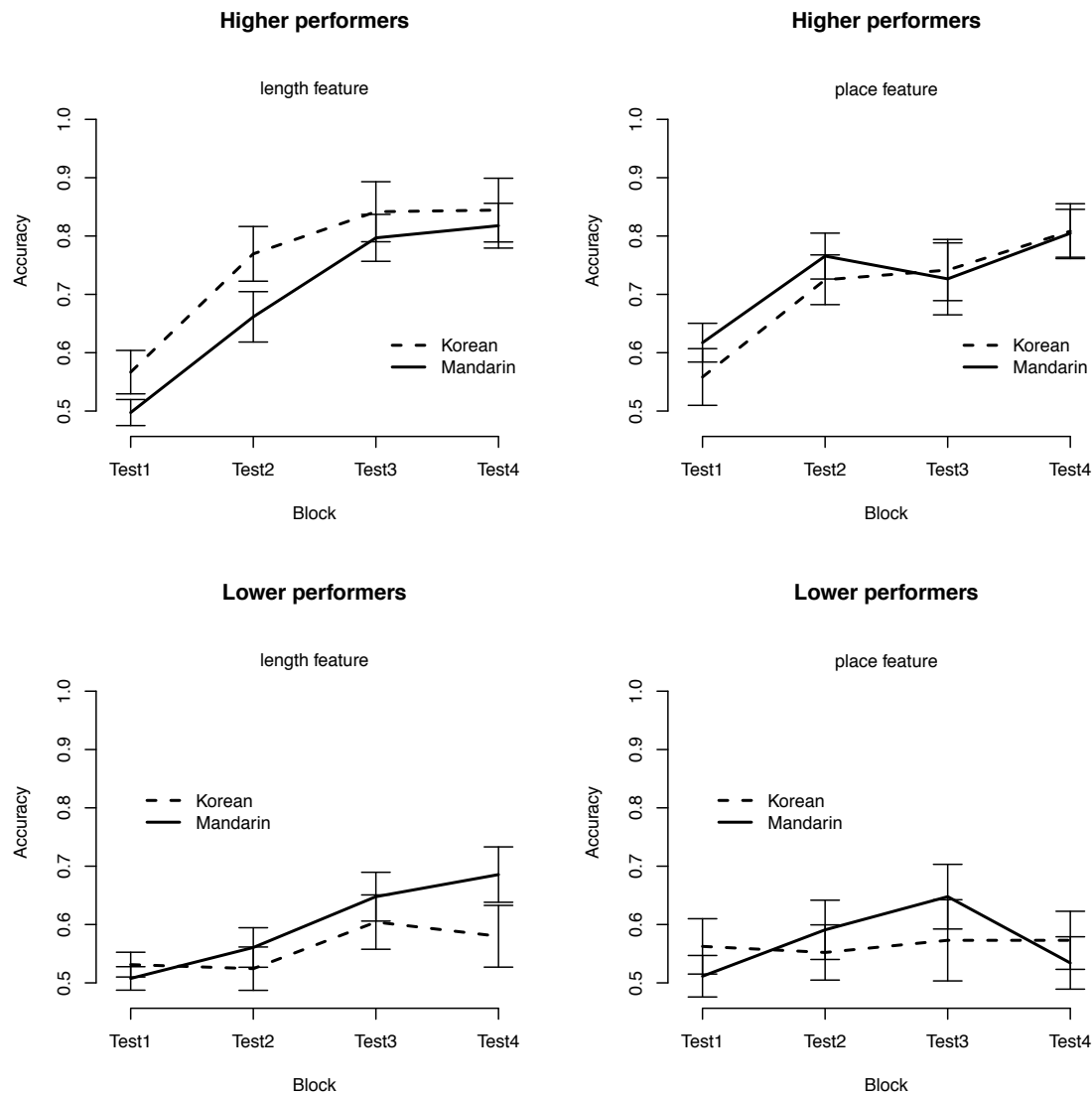
Figure A2. Word-learning task results by block (*length* and *place* trials) split by higher and lower

performers (split-1; results for split-2 were equivalent). Accuracy scores indicate proportion

of correct responses and error bars are standard errors.

Figure A3. Discrimination task results by block (*length* and *place* trials) split by higher and lower

performers (split-1; results for split-2 were equivalent). Accuracy scores indicate proportion
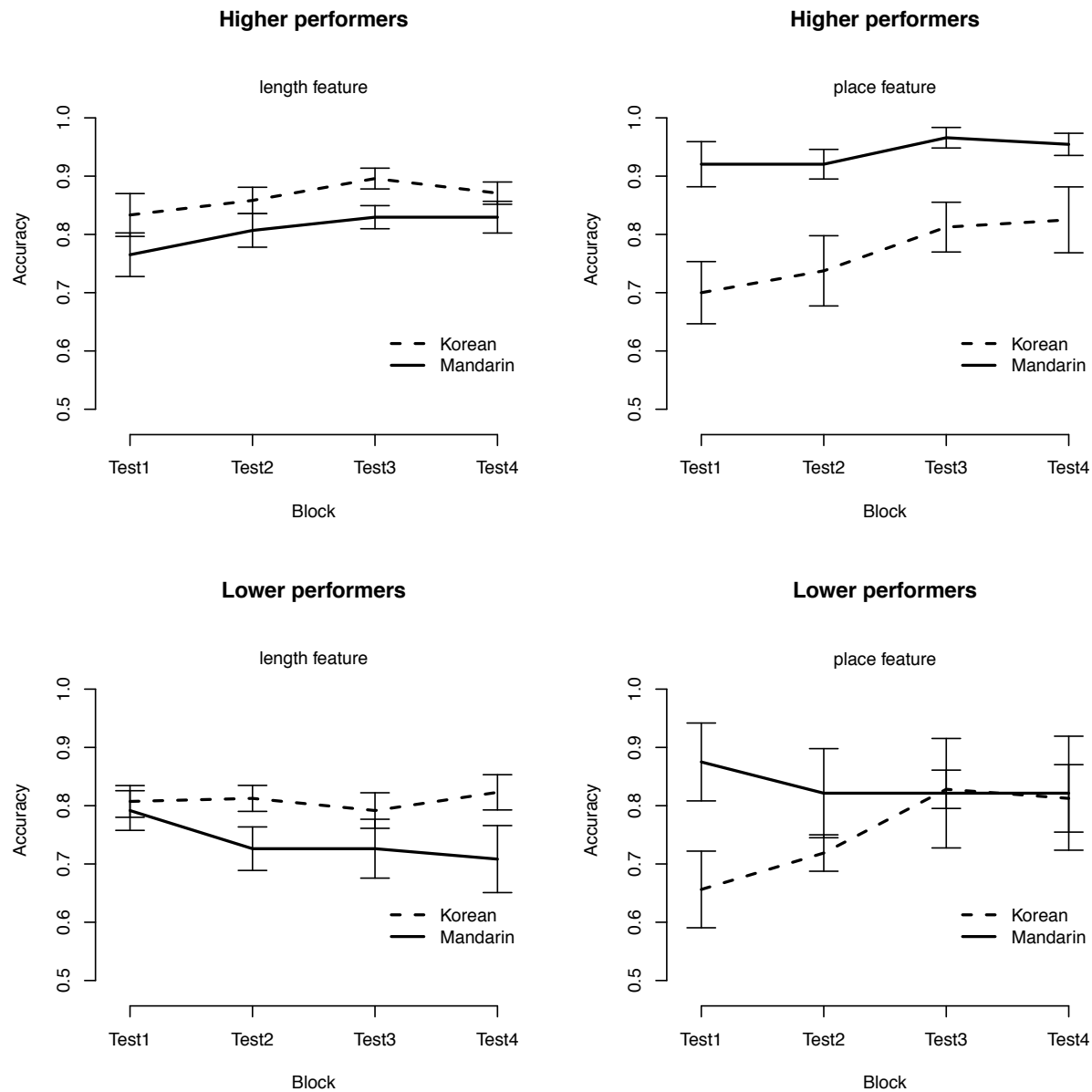
of correct responses and error bars are standard errors.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of

spoken word recognition using eye movements: Evidence for continuous mapping models.

Journal of Memory and Language, *38*(4), 419–439.

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on

lexical access. *Cognition*, *52*, 163–187.

Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of*

Child Language, *32*, 159–173.

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.

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child*

Development, *70*, 636–644.

Birdsong, D. (2009). Age and the end state of second language acquisition. In W. Ritchie & T.

Bhatia (Eds.), *The new handbook of Second Language Acquisition* (pp. 401–424).

Amsterdam: Elsevier.

Birdsong, D., & Molis, M. (2001). On the Evidence for Maturational Constraints in Second-

Language Acquisition. *Journal of Memory and Language*, *44*, 235–249.

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

Linguistics, *20*, 25–38.

Cooper, A., & Wang, Y. (2013). Effects of tone training on Cantonese tone-word learning.

Journal of the Acoustical Society of America, *134*(2), EL133–EL139.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2006). Acquiring an artificial lexicon: segment
      type and order information in early lexical entries. *Journal of Memory and Language, 54,*
      1-19.

Creel, S. C., & Dahan, D. (2010). The influence of the temporal structure of spoken words on
      paired-associate learning. *Journal of Experimental Psychology: Learning, Memory, and
      Cognition, 36,* 110-122.

Escudero, P., Broersma, M., & Simon, E. (2012). Learning words in a third language: Effects of
      vowel inventory and language proficiency. *Language and Cognitive Processes*, *iFirst*, 1–
      16.

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.

Fennell, C. T., Waxman, S. R., & Weisleder, A. (2007). With referential cues, infants
      successfully use phonetic detail in word learning. In *Proceedings of the 31st Boston
      University Conference on Language Development.* Cascadilla Press.

Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in
      well-known words. *Language and Speech*, *46*(2-3), 245–264.

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.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical Evidence: A Test of the Critical-Period
      Hypothesis for Second-Language Acquisition. *Psychological Science*, *14*(1), 31–38.

Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel
      second language phonemic contrasts. *Second Language Research*, *24*(1), 5–33.

Ingvalson, E. M., Barr, A. M., & Wong, P. C. M. (2013). Poor phonetic perceivers show greater benefit in phonetic-phonological speech learning. *Journal of Speech, Language, and Hearing Research*, *56*, 1045–1050.

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

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60–99.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test – Second Edition (KBIT-2)*. Circle Pines, MN: American Guidance Service.

Kaushanskaya, M. (2012). Cognitive mechanisms of word learning in bilingual and monolingual adults: The role of phonological memory. *Bilingualism: Language and Cognition*, *15*(3), 470–489.

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, *16*(4), 705–710.

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.

Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.

Lin, H. (2001). *A grammar of Mandarin Chinese*. München: Lincom Europa.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, *31*, 1–24.

Mani, N., & Plunkett, K. (2011). Does size matter? Subsegmental cues to vowel

    mispronunciation detection. *Journal of Child Language*, *38*, 606–627.

Nowak, P. M. (2006). The role of vowel transitions and frication noise in the perception of Polish

    sibilants. *Journal of Phonetics*, *34*, 139–152.

Pajak, B. (2012). *Inductive inference in non-native speech processing and learning*. PhD

    dissertation, University of California, San Diego.

Pajak, B., & Levy, R. (in press). The role of abstraction in non-native speech perception. *Journal*

    of Phonetics.

Pater, J., Stager, C., & Werker, J. (2004). The perceptual acquisition of phonological contrasts.

    Language , *80*(3), 384–402.

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in

    early word learning. *Developmental Science*, *12*(2), 339–349.

Shipley, W. C. (1967). *Shipley Institute of Living Scale: Vocabulary Test and Abstraction Test*.

    Los Angeles, CA: Western Psychological Services.

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

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception

    than in word-learning tasks. *Nature*, *388*, 381–382.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very

    young children. *Cognition*, *76*, 147–166.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations

    of 14-month-olds. *Psychological Science*, *13*(5), 480–484.

Swingley, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning.

    Cognitive Psychology, *54*, 99–132.

Thiessen, E. D. (2007). The effect of distributional information on children's use of phonemic

contrasts. *Journal of Memory and Language*, *56*, 16–34.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: levels of processing in perception

of spoken words. *Psychological Science*, *9*(4), 325–329.

Weber, A., & Cutler, A. (2004). Lexical competition in nonnative spoken word recognition.

Journal of Memory and Language, *50*, 1-25.

Werker, J. F., Fennell, C. T., Christopher T., Corcoran, K. M., & Stager, C. L. (2002). Infants'

ability to learn phonetically similar words: effects of age and vocabulary size. *Infancy*,

3(1), 1–30.

Werker, J. F. & Tees, R. C. (2005). Speech perception as a window for understanding plasticity

and commitment in language systems of the brain. *Developmental Psychobiology*, *46*(3),

233-251.

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations.

Journal of Memory and Language, *59*, 114–132.

White, K. S., Yee, E., Blumstein, S. E., & Morgan, J. L. (2013). Adults show less sensitivity to

phonetic detail in unfamiliar words, too. *Journal of Memory and Language*, *68*(4), 362–

378. doi:10.1016/j.jml.2013.01.003

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants

learn similar-sounding words. *Developmental Science*, *12*(3), 412–418.