BUCLD 36 Proceedings
To be published in 2012 by Cascadilla Press
Rights forms signed by all authors

Distributional Learning of L2 Phonological Categories by Listeners with Different Language Backgrounds

Bożena Pająk and Roger Levy

1. Introduction

Learning a second language (L2) is a complex task that often takes years before functional communicative skills are attained, and that rarely results in native-like proficiency. One of the biggest challenges for an L2 learner is acquiring the sounds and sound patterns of a new language. Beginner learners struggle with both perception and production of novel sounds, and even years of practice do not generally erase a foreign accent.

In this paper we are concerned with the problem of acquiring a new sound system from the perspective of a naïve listener. Previous research in this area has shown that perception of novel sounds is largely affected by native language (L1) phonology. Specifically, L2 learners often have difficulty perceiving distinctions between L2 sounds that do not form separate speech categories in their L1: e.g., the English *r-l* contrast is hard for L1-Japanese listeners who only have one category fitting the same acoustic-phonetic space (Miyawaki et al. 1975). These difficulties have been explained as a result of L2 being processed through the L1-attuned perceptual filter (Kuhl and Iverson 1995, Trubetzkoy 1939/1969), which has generally been conceptualized as a process of mapping L2 sounds onto the representations of L1 sound categories that are acoustically or articulatorily most similar (Best 1995, Best and Tyler 2007, Flege 1995). In these theories (following the terminology introduced by Best and colleagues), discrimination of two L2 sounds is predicted to be easy if each sound maps, or assimilates, onto a different L1 category (two-category assimilation), or if at least one of the sounds is very distinct from any sound in the L1 inventory (non-assimilation). Discrimination is relatively impaired when two L2 sounds are perceived as imperfect exemplars of only one L1 category (category-goodness difference). Finally, discrimination is hard if two L2 sounds assimilate completely onto one L1 category (single-category assimilation).

* Bożena Pająk, UC San Diego, bpajak@ucsd.edu. Roger Levy, UC San Diego, rlevy@ucsd.edu. For their helpful feedback the authors thank Eric Baković, Klinton Bicknell, Sarah Creel, Carolyn Quam, Sharon Rose, and four anonymous BUCLD reviewers. K. Michael Brooks, Hannah Byers-Straus, Rafi Feliciano, and Nicole Pyon helped with data collection. This research was supported by NIH Training Grant T32-DC000041 from the Center for Research in Language at UC San Diego to the first author, and the UC San Diego Academic Senate Grant to the second author.

There is, however, some evidence that non-native speech perception is more fine-grained than assumed by the mapping accounts. In particular, learners do not just compute similarity between novel sound segments and their L1 counterparts, but are able to decompose novel sounds into familiar phonetic dimensions, with discrimination being facilitated if novel L2 sounds differ along dimensions that distinguish between other sounds in the learner's L1 (Pająk 2010a, Pająk 2010b, Pająk and Levy in prep.). For example, speakers of a language like Vietnamese, where length is a relevant cue for distinguishing between vowel categories (e.g., [bang] 'state' vs. [ba:ng] 'ice'; Winn et al. 2008), are better at discriminating length contrasts on consonants than speakers of a language like Mandarin where length is never relevant (Lin 2001). This result is surprising given the assumptions of the mapping theories, because – if we assume segment-to-segment mappings – L1 vowel categories should not affect how L2 consonants are mapped onto L1. Instead, for both Vietnamese and Mandarin speakers, novel short and long consonants should be assimilated to the most similar L1 categories, which in both cases are the corresponding short consonants. Regardless of whether this kind of mapping was analyzed as a case of single-category assimilation or a category-goodness difference, under the current assumptions of the mapping theories Vietnamese and Mandarin speakers are expected to perform equally in this regard. This result points to the need of incorporating in the accounts of non-native speech perception learners' sensitivity to more than just whole segments, such as their sensitivity to subsegmental cues, as well as their ability to generalize from L1 subsegmental phonetic properties to novel segment classes in L2.

Furthermore, the success of learning the sounds of a new language is not fully determined by learner's initial, L1-shaped, perceptual abilities. There is a high degree of plasticity in the adult speech processing system, as indicated by the fact that perception of novel sound contrasts improves with training (e.g., Goudbeek et al. 2008, Lim and Holt 2011, Logan et al. 1991, McClaskey et al. 1983, Pisoni et al. 1982). Furthermore, adults – just like infants – have been found to be highly sensitive to subtle language statistics, such as the statistical distribution of phonetic variation in the speech signal (Maye and Gerken 2001, Pająk and Levy 2011, Perfors and Dunbar 2010). Specifically, adults can pick up on distributional cues that indicate whether sounds along a given phonetic continuum belong to one ore two categories, simply by being exposed to sounds sampled from that continuum with either a unimodal or a bimodal distribution. For example, in our own study (Pająk and Levy 2011) we exposed English monolinguals to novel language sounds sampled from a length continuum (e.g., a continuum ranging from [aja] to [ajja]), using the distributional learning paradigm (Maye and Gerken 2000, Maye et al. 2002). For one group of participants, the sounds in the exposure phase were predominantly either short or long (bimodal distribution), and for another group of participants they were mostly of medium length (unimodal distribution), as illustrated in Fig. 1. Subsequently, we tested them on pairs of words with sounds from the endpoints of the continuum (e.g., [aja]-[ajja]), asking them to judge whether these were two different words in that language or the same word repeated twice. We found more 'different' responses for participants trained on the bimodal distribution than those trained on the unimodal distribution, indicating that learners were sensitive to the distributional cues along the length continuum: that is, they were more likely to infer two categories when exposed to the bimodally-distributed tokens, and one category when exposed to the unimodally-distributed ones.

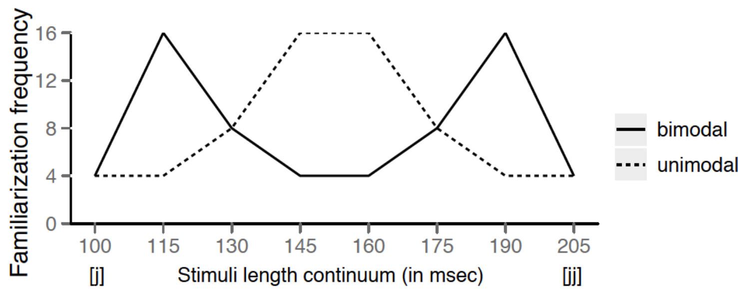

Figure 1: Critical training stimuli in Pająk and Levy (2011).

In Pająk and Levy (2011) we proposed a new model of L2 phonological acquisition that incorporates the recent findings about L2 phonetic category learning discussed above. The model views acquisition of phonetic categories as a general categorization process in which learners combine their L1 knowledge with statistical information from L2 input. According to the model, L1 knowledge provides learners with an inductive bias as to what categories might be expected in L2, prior to any actual L2 exposure. The other source of information is L2 statistics: when L2 exposure begins, learners extract statistical information from L2 input and use it to update their beliefs about L2 categories, thus combining their L1 biases with statistical information from L2. The model provides a framework to study the timecourse of how initial perception and categorization of L2 sounds by naïve listeners gradually change with L2 exposure. The current study constitutes our first attempt to examine this process by looking at how L1 biases affect interpretation of distributional information in a novel language.

2. Experiment

We recruited speakers of Korean and Mandarin, whose L1 perceptual biases we investigated in a previous study (Pająk 2010a, Pająk and Levy in prep.). We learned that speakers of Korean, in which length is a cue to distinguishing phonetic categories (e.g., [pul] 'fire' vs. [pu:l] 'blow'; Sohn 2001), are better at discriminating length contrasts than are speakers of Mandarin, in which there are no length contrasts (Lin 2001). The reverse is true for the place of articulation contrast between alveolo-palatal and retroflex sibilants, which Mandarin has (as an allophonic alternation; Lin 2001) but Korean does not (Sohn 2001). Given this finding, consider the sound distribution in a hypothetical language provided in Fig. 2 (LEFT) that could be interpreted as either a place distinction (alveolo-palatal vs. retroflex) or a length distinction (short vs. long). The evidence for a place contrast is suggested by strongly bimodal distribution and no overlap between the two data clusters. The evidence for a length contrast is suggested by a more weakly bimodal distribution with clear overlap.[1] Under our account, speakers of Korean should be biased toward inferring a length-based category distinction and against inferring a place-based category distinction, and thus interpret this phonetic input as two categories along the length dimension. Speakers of Mandarin, on the other hand, should be biased toward inferring a place-based category distinction and against inferring a length-based category distinction, thus interpreting the input as two categories along the place dimension. With input unimodally distributed in length (Fig. 2: RIGHT), neither group of speakers should infer that length is contrastive, but Korean speakers are expected to be less inclined to infer a place distinction than Mandarin speakers. The study reported here tested these predictions.

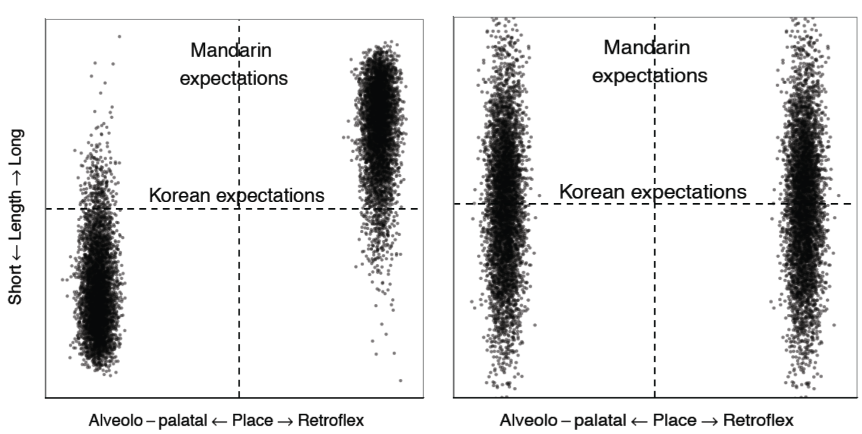

Figure 2: Schematic representation of segment statistics in a novel language. LEFT: strongly bimodal *place* and weakly bimodal *length*. RIGHT: strongly bimodal *place* and unimodal *length*.

1 Despite no overlap along the place dimension, the between-cluster confusability is expected to be high due to the fact that this distinction is acoustically very subtle (Nowak 2006) and, as confirmed by our earlier study (Pająk 2010a, Pająk and Levy in prep.), poorly discriminable by both Mandarin and Korean speakers. Length distinctions, on the other hand, are discriminated relatively more easily (perhaps since temporal cues are more salient than spectral ones; Hall et al. 2002), as also confirmed by our study with Korean and Mandarin speakers (Pająk 2010a, Pająk and Levy in prep.). Thus, the gradient distribution and overlap along the length dimension might reduce the relative salience of the length cue and increase confusability between tokens along that dimension. Based on this reasoning, we expected that the relative confusability along each of the two dimensions would be roughly comparable.

2.1. Participants

144 undergraduate students at UC San Diego participated in the experiment for course credit or payment. Half were Korean-English bilinguals, and the other half were Mandarin-English bilinguals. All learned Korean or Mandarin from birth, and reported to be at least competent speakers of those languages. In most cases they had some limited high school and/or college exposure to Spanish or French. Some Mandarin-English bilinguals were also familiar with Taiwanese, mostly through family exposure. All participants reported no history of speech or hearing problems.

2.2. Materials

The materials consisted of nonce words recorded in a soundproof booth by a phonetically-trained native speaker of Polish. The critical items included segments from two classes: alveolo-palatals ([ɕ], [tɕ]) and retroflexes ([ʂ], [tʂ]). They were recorded as words with long intervocalic consonants: [aɕɕa], [atɕtɕa], [aʂʂa], [atʂtʂa]. Subsequently, two recordings of each word were chosen, and the consonant length in each word was manipulated to create length continua, each with eight tokens, where durations of consonants ranged from short (140msec) to long (280msec) in a 2:1 duration ratio (cross-linguistically, the long-to-short consonant ratio varies between 1.5 to 3; Ladefoged & Maddieson 1996). Each token adjacent on the continuum differed in duration by 20msec. For affricates, the frication portion was held constant throughout the continuum (90msec), and only the closure duration was manipulated (ranging from 50 to 190msec). The fillers resembled the critical items, but included different consonants: [afa], [ava], [axa], [aχa], [aba], [aβa], [asa], [aθa], [ada], [aða], [aʁa], [aʕa], [atsa], [adza], [aka], [aqa]. Eight different recordings of each filler word were used in the experiment. There were no length manipulations on fillers.

2.3. Procedure

We followed the general procedure of the distributional learning paradigm (Maye and Gerken 2000, Maye et al. 2002), as applied by Maye and Gerken (2001) in a study with adult participants, where the main idea is that by manipulating the frequency of exposure to sounds that vary along a given dimension, participants can recover the underlying structure along that dimension and, for example, infer two categories when the input is bimodally-distributed, but only one category when the input is unimodal (e.g., as in Pająk and Levy 2011, discussed above).

The general overview of the critical part of the experiment is the following: In training, participants were exposed to a novel language by listening to tokens that varied along the length and place dimensions, as was depicted in Fig. 2. The place contrast was indicated by including naturally recorded tokens of both alveolo-palatals and retroflexes. The evidence for the length contrast was provided by varying frequency of exposure to different tokens along the length continuum. In testing, participants heard pairs of words that were clear place or length contrasts, and were asked to judge whether these were two different words or two repetitions of the same word. A detailed description of the study is provided below.

Each participant was randomly assigned to one of four conditions: (1) *discrimination* (13 Korean, 13 Mandarin), (2) *filler training* (13 Korean, 13 Mandarin), (3) *bimodal-length training* (23 Korean, 23 Mandarin), (4) *unimodal-length training* (23 Korean, 23 Mandarin). The first two conditions were introduced in order to assess baseline performance. In each condition participants were presented with the same exact testing. The conditions differed only in instructions and/or training provided prior to and in the middle of testing.

The instructions included a short practice. In the *discrimination* condition, the practice consisted of acoustically identical ('same') pairs and acoustically distinct ('different') pairs of words from the new language that were not included in the subsequent training nor testing. In the *training* condition, the practice consisted of English words, where 'different' words were minimal pairs (e.g., *mass – miss*), and 'same' words were repetitions of the same word pronounced with different intonations.

In the *discrimination* condition, participants were told that the goal of the experiment was to assess how well they can hear differences between sounds in a new language. There was no exposure to the language besides the testing trials.

In the *training* conditions, participants were told that they would first listen to words in a new language (training) and then would be asked to use what they learned in testing. In training, participants listened to single words presented over headphones and were asked to push a button after hearing each word. The response to a given stimulus triggered the presentation of the following stimulus with a delay of 1sec. There were two training sessions: one prior to testing, and another after the first half of testing. The first training session consisted of a total of 384 words (four repetitions of one training block) and lasted about 10min. The second training session consisted of a total of 192 words (two repetitions of one training block) and lasted about 5 minutes. Stimulus order was randomized for each participant, and there was a self-terminated break after each block.

In the *filler training* condition, participants were exposed to 12 filler words ([afa], [ava], [axa], [aχa], [aba], [aβa], [asa], [aθa], [ada], [aða], [aʁa], [aʕa]) with no variability along the length dimension (i.e., all segments were short). One training block consisted of 96 items: 8 repetitions of each word, where each repetition was a different recording of the word.

In the *bimodal-length training* condition, participants were exposed to words that were either critical or filler items. One training block consisted of 64 critical items (8 tokens from each length continuum type: [ɕ]-[ɕɕ], [tɕ]-[tɕtɕ], [ʂ]-[ʂʂ], [tʂ]-[tʂtʂ]) and 32 fillers (8 repetitions each of the words [afa], [ava], [axa], [aχa], where each repetition was a different recording of the word). The critical items from the length continua were presented with different frequencies, as illustrated in Fig. 3 (LEFT): alveolo-palatals were most frequently short, and retroflexes were most frequently long, suggesting a bimodal distribution along the length dimension.

The *unimodal-length training* condition differed from the *bimodal-length training* only in the frequencies of critical items, as in Fig. 3 (RIGHT): both alveolo-palatals and retroflexes were most frequently of medium length, indicating a unimodal distribution along the length dimension.

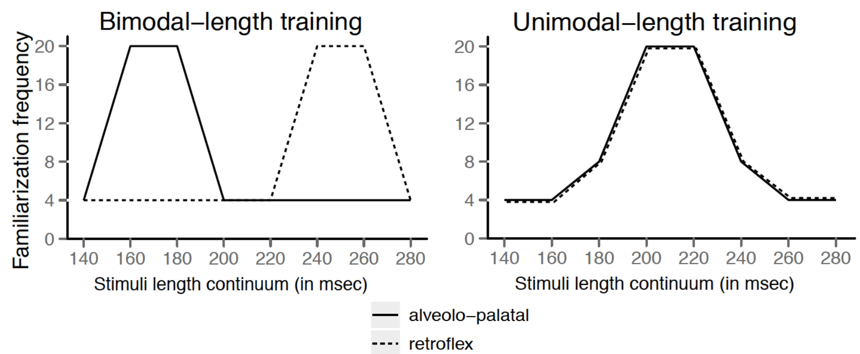

Figure 3: Critical training stimuli in bimodal-length and unimodal-length training conditions.

The testing was identical for all participants, and consisted of a same-different AX discrimination task. Participants listened to pairs of words, and were asked to answer whether these were 'same' or 'different' by pushing one of two buttons. In the *discrimination* condition, participants were instructed to answer 'different' whenever they heard any kind of difference between the two words. In the *training* conditions, on the other hand, participants were asked to make an intuitive judgment, based on what they learned during training, about what differences counted as 'different' in this language and whether the words in a pair were two different words or two repetitions of the same word. There were two critical contrasts: *length* and *place*. The 'different' critical pairs are illustrated in Fig. 4. For *length*, these were endpoints of each length continuum differing only in length (e.g., [aɕa]-[aɕɕa]), but each word in a pair originated from a different recording of the word. For *place*, these were items of medium length that differed only in place (e.g., [aɕa]-[aʂa]). The 'same' pairs were always two different recordings of a word from the same point along the length and place dimensions (e.g., $[\text{aɕa}]_{rec1}$-$[\text{aɕa}]_{rec2}$). Just like for 'different' pairs, only items from the endpoints and the middle of the length continuum were used. For filler 'different' pairs, these were two words that differed by one segment: the contrasts were either in voicing ([afa]-[ava], [atsa]-[adza]), place of articulation ([axa]-[aχa], [asa]-[aθa], [aʁa]-[aʕa], [aka]-[aqa]), or place and/or manner ([aba]-[aβa], [ada]-[aða],). The 'same' pairs were again always two different recordings of the same word.

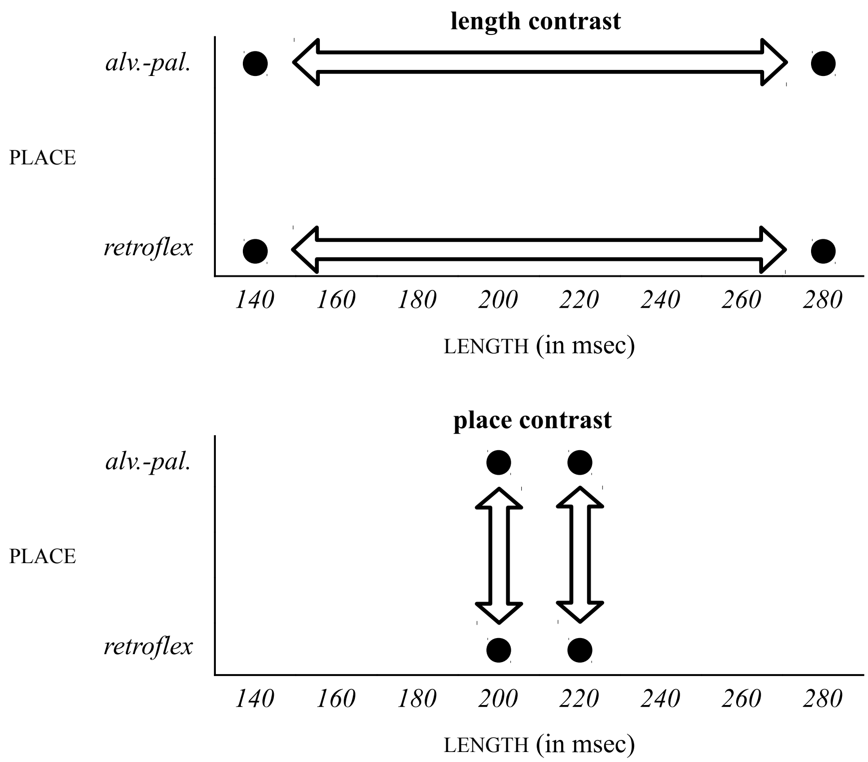

Figure 4: Critical testing 'different' word pairs.

There was a total of 384 word pairs in testing, which included 6 repetitions of the testing block. The block consisted of 32 critical pairs (16 'same' and 16 'different') and 32 filler pairs (16 'same' and 16 'different'). The content of the block was balanced with each pair occurring twice. The words in each pair were separated by an interstimulus interval of 750msec. As with training, stimulus order was randomized for each participant, and there was a self-terminated break after each block. Testing lasted about 20min.

2.4. Results

The results from 'same' trials are provided in Tab. 1. Participants rarely responded 'different' on 'same' trials, and there were no significant differences between CONDITIONS. Therefore, we only analyzed responses from 'different' trials, using 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 that were manipulated within participants or within items. We controlled for main effects of participants' dominant language, length of residence in the US, and – for bimodal- vs. unimodal-length comparisons – performance on filler items by adding them as fixed effects to the models.

Table 1: Proportion of 'different' responses on 'same' trials (standard errors in parentheses).

LANGUAGE	DISCRIMINATION		FILLER-TRAINING		BIMODAL-LENGTH TRAINING		UNIMODAL-LENGTH TRAINING	
	Length	*Place*	*Length*	*Place*	*Length*	*Place*	*Length*	*Place*
Korean	.11 (.02)	.15 (.03)	.10 (.02)	.12 (.02)	.12 (.02)	.14 (.02)	.12 (.02)	.15 (.03)
Mandarin	.13 (.03)	.14 (.04)	.09 (.03)	.10 (.03)	.12 (.02)	.13 (.02)	.13 (.02)	.14 (.02)

The results from 'different' trials are illustrated in Fig. 5. First, we predicted that – in agreement with their L1 biases – Mandarin speakers should overall give more 'different' responses than Korean speakers on the place trials, while the reverse should be true for the length trials. We examined this in a model with fixed effects of LANGUAGE (*Korean*, *Mandarin*) and CONTRAST (*place*, *length*), and found a significant interaction between the two effects ($p<.001$) in the predicted direction: Mandarin speakers responding more 'different' on *place*, and Korean speakers responding more 'different' on *length*. In addition, there was a significant main effect of CONTRAST ($p<.001$), with more 'different' responses for *place* than for *length*, suggesting that the place contrast was perhaps relatively more salient than the length contrast. Finally, there was a significant main effect of LANGUAGE ($p<.01$): Mandarin speakers gave overall more 'different' responses than Korean speakers.

As the next step, we looked at the data from the two baseline conditions, *discrimination* and *filler-training* to assess how perceptual sensitivity compared to phonetic category judgments with no prior training on *place* or *length* items. We examined this in a model with fixed effects of CONDITION (*discrimination*, *filler-training*), LANGUAGE (*Korean*, *Mandarin*), and CONTRAST (*place*, *length*). As a sanity check, we expected at least as many 'different' responses in the *discrimination* as in the *filler-training* condition, since perceptual sensitivity should constitute a ceiling for category judgments. We found a significant main effect of CONDITION ($p<.001$) with more 'different' responses in *discrimination* than in *filler-training*, which was consistent with our prediction. Furthermore, we expected an interaction between LANGUAGE and CONTRAST, as already found in the overall model, which was indeed significant ($p<.001$). In addition, as in the overall model, we found a significant main effect of CONTRAST ($p<.001$) with more 'different' responses on *place* than on *length*. Finally, there was an unexpected significant interaction between CONDITION and CONTRAST ($p<.001$): for *place*, 'different' responses were only slightly less frequent in the *filler-training* than in the *discrimination* condition; for *length*, on the other hand, the 'different' responses were considerably lower in the *filler-training* than in the *discrimination* condition. This result suggests that after exposure to only short segments in training, the expectations for a length contrast decreased significantly with respect to participants' perceptual sensitivity. The expectations for a place contrast, on the other hand, did not seem considerably affected by the filler training, and were maintained at nearly the same level as perceptual sensitivity.

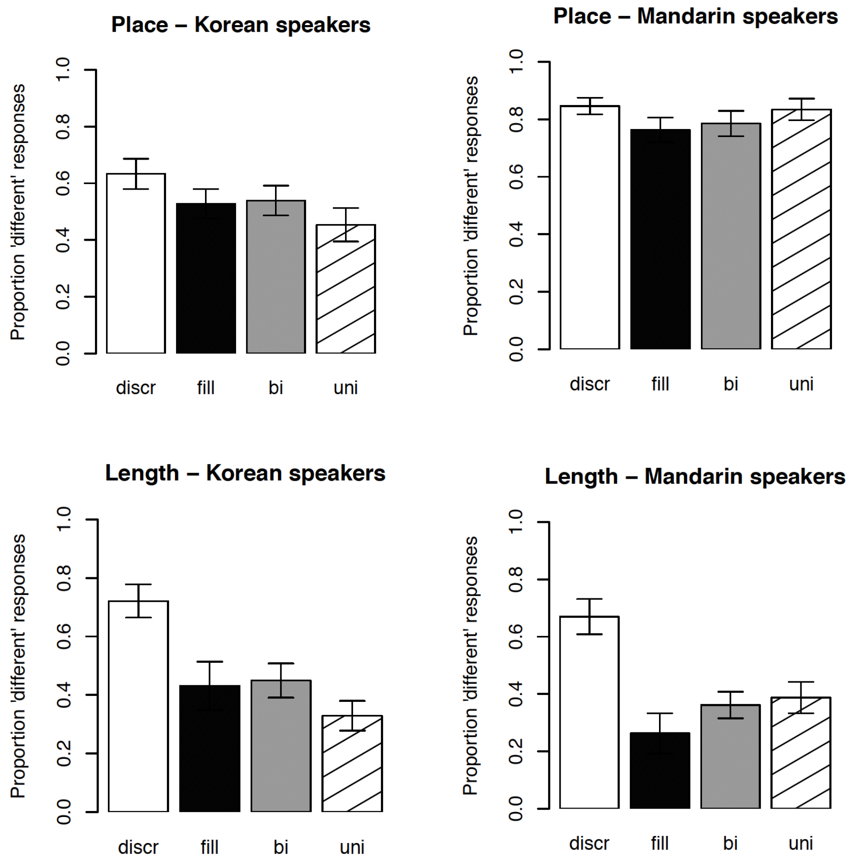

Figure 5: Proportion of 'different' responses on 'different' trials (error bars are standard errors).

Next, we examined the data from all training conditions for fixed effects of CONDITION (*filler*, *bimodal-length*, *unimodal-length*), LANGUAGE (*Korean*, *Mandarin*), and CONTRAST (*place*, *length*).[2] We predicted a three-way

2 The model with the full random effects structure failed to converge. Thus, we removed the interaction between CONDITION and LANGUAGE from random effects for items.

interaction, because speakers of Korean – but not Mandarin – should be highly sensitive to the distribution of length. For length, we expected more 'different' responses for bimodal-length training than for unimodal-length training, and we expected the reverse for place. However, the analysis revealed no significant three-way interaction. Instead, there was a two-way interaction of LANGUAGE and CONTRAST ($p<.001$), but it did not interact with CONDITION. The key reason why the three-way interaction did not come out as expected is the fact that the Korean/place/unimodal response was – unlike what we predicted – lower than the Korean/place/bimodal response.

We followed up on this result by running pairwise CONDITION by CONTRAST comparisons within each language for all training conditions. As found in previous models, in all tests there was a significant main effect of CONTRAST (p*s*<.05). For Korean speakers, the only other nearly significant difference was the marginal main effect of CONDITION for bimodal-length vs. unimodal-length ($p=.052$): more 'different' responses after training on bimodal-length than on unimodal-length. When the data were examined separately for length and for place trials, there was a significant effect of CONDITION for *length* ($p<.05$), but not for *place*. For length trials, the responses in the filler-training condition did not differ significantly from either bimodal-length or unimodal-length (which was perhaps due to smaller number of participants in the filler-training condition). However, if we interpret the results numerically, responses for bimodal-length were slightly higher than for filler-training, and responses for unimodal-length were considerably lower. Overall, this suggests that Korean speakers were sensitive to subtle distributional cues present on length in training. For Mandarin speakers, on the other hand, the response pattern was quite different. There were marginal ($p=.076$) and close to marginal ($p=.13$) effects of CONDITION when comparing filler vs. unimodal-length, and filler vs. bimodal-length conditions, respectively. Furthermore, contrary to what we found for speakers of Korean, Mandarin speakers' responses on length trials were numerically higher after *both* bimodal-length training and unimodal-length training compared to the filler-training condition, and this difference was close to significant ($p=.11$) when the length-trials data from bimodal-length and unimodal-length conditions were pooled together. This result suggests that Mandarin speakers were not attending to the distributional length cues in training, but rather increased their 'different' responses after any exposure to variability in length. (Note that the smaller number of participants in the filler-training condition may place some limits on statistical power in this last analysis.)

Taken together, the type of training did not have a clear effect on Mandarin speakers, but – if anything – training on our length/place materials sensitized them overall to subtle differences and increased their proclivity to infer that tokens differing either in place or length are different words. Korean speakers, on the other hand, started out overall fairly sensitive to both distinctions (insofar as they could perceive them), and the main effect of training on the length/place stimuli was to *desensitize* them to differences when the length distribution was unimodal. The question then remains why Korean speakers had greater proclivity to answer 'different' for both contrasts in bimodal-length than in unimodal-length training. At this point, we can speculate that Korean speakers may have tended to infer four categories in the bimodal-length training, not just two. The reasons for this are unclear, but it may be that despite our initial assumptions the evidence for place contrasts in training was overall more salient than the evidence for length. This interpretation would be consistent with the discrepancy between our perceptual-discrimination results in the present study, where Korean speakers were nearly as accurate in identifying place distinctions as they were in identifying length distinctions, and those in Pająk (2010a) and Pająk and Levy (in prep.), where Korean speakers were far more accurate on length contrasts than on place contrasts.

3. Conclusion

The results presented in this paper replicate our previous finding that non-native speech perception is guided by L1-derived perceptual biases (Pająk 2010a, Pająk and Levy in prep.). The results also provide new evidence that perceptual abilities do not fully predetermine how novel sounds are categorized in a new language, but instead – as predicted by our model – that categorization of L2 sounds is a result of combining L1-shaped perceptual biases with the distributional information from L2 input. Crucially, the interpretation of L2 statistics varies depending on the expectations that learners have about L2 given their previous language background. The results reported here are, however, far from conclusive. There are many unanswered questions regarding how exactly distributional information and prior L1 biases interact, and what factors might play a role in how each piece of information is weighed. We believe that the study presented here constitutes the first step in understanding the complex nature of how previous linguistic experience affects the use of distributional evidence in the acquisition of phonetic categories in a new language.

References

Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In *Speech perception and linguistic experience: issues in cross-language research,* ed. by Winifred Strange, 171–204. Timonium, MD: York Press.

Best, & Michael D. Tyler. 2007. Nonnative and second-language speech perception. In *Language experience in second language speech learning: in honor of James Emil Flege,* ed. by Ocke-Schwen Bohn & Murray J. Munro, 13–34. Amsterdam/ Philadelphia: John Benjamins.

Flege, James. 1995. Second-language speech learning: theory, findings and problems. In *Speech perception and linguistic experience: issues in cross-language research,* ed. by Winifred Strange, 229–273. Timonium, MD: York Press.

Goudbeek, Martijn, Anne Cutler, & Roel Smits. 2008. Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication* 50. 109–125.

Hall, Deborah A., Ingrid S. Johnsrude, Mark P. Haggard, Alan R. Palmer, Michael A. Akeroyd, & A. Quentin Summerfield. 2002. Spectral and temporal processing in human auditory cortex. *Cerebral Cortex* 12. 140–149.

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

Kuhl, Patricia K., & Paul Iverson. 1995. Linguistic experience and the "perceptual magnet effect". In *Speech perception and linguistic experience: issues in cross-language research,* ed. by Winifred Strange, 121–154. Timonium, MD: York Press.

Ladefoged, Peter, & Ian Maddieson. 1996. *The sounds of the world's languages.* Oxford, UK; Cambridge, MA: Blackwell.

Lim, Sung-joo, & Lori Holt. 2011. Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science* 35. 1390–1405.

Lin, Hua. 2001. *A grammar of Mandarin Chinese.* München: Lincom Europa.

Logan, John S., Scott E. Lively, & David B. Pisoni. 1991. Training Japanese listeners to identify English /r/ and /l/: a first report. *Journal of the Acoustical Society of America* 89. 874–886.

Maye, Jessica, & LouAnn Gerken. 2000. Learning phonemes without minimal pairs. In *Proceedings of the 24th Annual Boston University Conference on Language Development,* ed. by S. Catherine Howell, Sarah A. Fish, & Thea Keith-Lucas, 522–533. Somerville, MA: Cascadilla Press.

Maye, Jessica, & LouAnn Gerken. 2001. Learning phonemes: how far can the input take us? In *Proceedings of the 25th Annual Boston University Conference on Language Development,* ed. by A. H-J. Do, L. Domínguez, & A. Johansen, 480–490. Somerville, MA: Cascadilla Press.

Maye, Jessica, Janet F. Werker, & LouAnn Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82. B101–B111.

McClaskey, Cynthia L., David B. Pisoni, & Thomas D. Carrell. 1983. Transfer of training of a new linguistic contrast in voicing. *Perception and Psychophysics* 34. 323–330.

Miyawaki, Kuniko, Winifred Strange, Robert R. Verbrugge, Alvin M. Liberman, Jennifer J. Jenkins, & Osamu Fujimura. 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.

Nowak, Paweł M. 2006. The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of Phonetics* 34. 139–152.

Pająk, Bożena. 2010a. Bilinguals generalize from known phonological contrasts in perception of a novel language. Poster at the 51st Annual Meeting of the Psychonomic Society. St. Louis, MO.

Pająk, Bożena. 2010b. Perceptual advantage from generalized linguistic knowledge. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society,* ed. by Stellan Ohlsson & Richard Catrambone, 369–374. Austin, TX: Cognitive Science Society.

Pająk, Bożena, & Roger Levy. 2011. Phonological generalization from distributional evidence. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society,* ed. by Laura Carlson, Christoph Hölscher, & Thomas Shipley, 2673–2678. Austin, TX: Cognitive Science Society.

Pająk, Bożena, & Roger Levy, in prep. Adult language acquisition: The role of generalization from known languages. Manuscript.

Perfors, Amy, & David Dunbar. 2010. Phonetic training makes word learning easier. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society,* ed. by Stellan Ohlsson & Richard Catrambone, 1613–1618. Austin, TX: Cognitive Science Society.

Pisoni, David B., Richard N. Aslin, Alan J. Perey, & Beth L. Hennessy. 1982. Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance* 8.297– 314.

Sohn, Ho-Min. 2001. *The Korean language*. Cambridge, UK: Cambridge University Press.

Trubetzkoy, Nikolai. 1939/1969. *Principles of phonology [Grundzüge der Phonologie]*. Berkeley, CA: University of California Press.

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