Phonological Generalization from Distributional Evidence

Bożena Pająk & Roger Levy
{bpajak, rlevy}@ucsd.edu
Department of Linguistics, UC San Diego
9500 Gilman Drive #108, La Jolla, CA 92093 USA

Abstract

We propose a model of L2 phonological learning in which the acquisition of novel phonological category inventories proceeds not by mapping L2 inputs onto existing category inventories available in L1 and other already known languages, but rather through general categorization processes in which L1 and other language knowledge serves as an inductive bias. This approach views linguistic knowledge as hierarchically organized such that the outcome of acquisition of a language—L1 or otherwise—includes not only knowledge of the specific language in question, but also beliefs about *how any language is likely to be structured*. In this paper we test a set of predictions regarding how two key types of information can come together to drive L2 learning: distributional information within a single phonetic dimension and generalization bias derived from existing knowledge of language. We tested these predictions by training adult monolingual English speakers in a distributional learning paradigm (Maye & Gerken, 2000; Maye, Werker, & Gerken, 2002) on a novel contrast, *segmental length*, and testing them on categorization of short and long segments for both trained and untrained items. Results show both learning and generalization from one class of segments (sonorants) to another class (obstruents), broadening the empirical range of phonetic contrasts for which distributional learning has been shown to be effective and providing evidence for our approach to L2 learning as one of inductive inference and generalization rather than of mapping.

Keywords: L2 phonological acquisition; distributional learning; speech perception; categorization; generalization.

Introduction

Language learning in adulthood has traditionally been regarded as an inherently difficult process due to first language (L1) interference. One reason for this view is the common assumption that second language (L2) begins as parasitic on L1, and only gradually separates itself as an independent language in the course of learning (e.g., MacWhinney, 1987). We propose a model in which L2 learning (and, more specifically, phonological acquisition) is instead viewed as a process of inductive inference, where learners make implicit predictions about the possible underlying structures of the novel language by combining two sources of information: (1) the statistical properties of the L2 input, and (2) previous language knowledge (including both experience and any innate biases), which serves as an inductive bias guiding learners in their inferences about novel phonological structures. The proposed model assumes that the structure of language knowledge is represented at multiple levels with one level for knowledge of specific languages, and a higher level representing more abstract knowledge of the structure of languages in general. This model fits within the general approach to learning as a process of rational hypothesis construction and testing, in which learners infer the underlying structure of their input by generalizing beyond the specific surface properties that they are exposed to (e.g., Tenenbaum & Griffiths, 2001; Xu & Tenenbaum, 2007; Gerken, 2010). At the same time, the proposed model is radically different from traditional views on L2 phonological acquisition, where perception and learning of novel sounds have been assumed to rely on the process of mapping of L2 sounds onto L1 phonological categories (Best, 1995; Flege, 1995; Hancin-Bhatt, 1994; Kuhl & Iverson, 1995). Under these views, L2 learners—instead of making implicit rational predictions about the L2 phonological categories—try to establish conceptual links between L2 sounds and their most similar L1 counterparts, so as to process the unfamiliar sounds directly through their L1 phonological system. We propose, in contrast, that learners do not directly filter the L2 speech input through their L1 phonological categories, but rather that they make the best possible guesses about how individual novel sounds are grouped into categories by relying on the same mechanisms that are used in general categorization processes for many types of perceptual stimuli.

In order to define the details of the proposed model we follow the general categorization literature in that any perceptual stimulus can be represented as a point in a multidimensional psychological space. People are able to categorize the stimuli by abstracting information about stimulus dimensions (e.g., color, shape, size, etc.) from single instances of the input (Posner & Keele, 1968; Kruschke, 1992). Within Kruschke's model, learning categories occurs by computing and attaching weights (or attention strength) to each of the stimulus dimensions. The attention strength reflects the relevance of any given dimension for a particular categorization task. That is, high strength will be associated with dimensions hypothesized as the most informative in distinguishing between categories. This way, people are able to perform categorization tasks by selectively attending to dimensions that are relevant, while at the same time ignoring other dimensions (Nosofsky, 1986). For instance, with stimuli varying along three dimensions such as color, shape, and size (Fig. 1a), people are good at categorizing by just one dimension, for example color. In this situation, the psychological space gets stretched along the color dimension—due to high attention strength assigned to this dimension (Kruschke, 1992)—and shrunk along the size and shape dimensions (Fig. 1b). This strategy is effective in categorization tasks because by attending selectively to the relevant dimension, people maximize within-category similarity and between-category discriminability, thus avoiding between-category confusion due to variation along irrelevant dimensions.

We pursue a similar idea to account for phonological

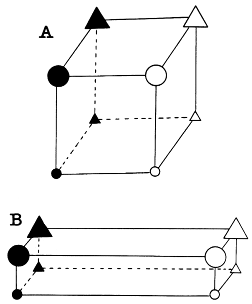

Figure 1: (a) Eight stimuli that vary along three binary-valued dimensions: color (black/white), shape (circle/triangle), and size (big/small). (b) Attending selectively to the color dimension. (Figure from Nosofsky, 1986, p. 4.)

categorization. Sound segments, as other perceptual stimuli, can be represented as points in a multidimensional perceptual space. Languages group segments into phonological categories, or phonemes, by partitioning this perceptual space along different phonetic dimensions (Maddieson, 1984). Therefore, a large part of the problem facing children in L1 phonological learning is to differentiate between phonologically relevant (i.e., informative for assigning meaning) and irrelevant phonetic dimensions. That is, children learn to selectively attend to certain phonetic dimensions (e.g., formant frequencies), while disregarding—at least for purposes of categorization—others (e.g., amplitude) (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Jusczyk, 1992).

Now, let us turn our attention to L2. Accurate phonological learning in L2 requires repartitioning of the perceptual space in accordance with which phonetic dimensions are informative to categorize L2 sounds (Strange & Shafer, 2008). On our proposal this is difficult because learners' L1 knowledge has created strong inductive bias in inferences about what any L2 structure may be like, which leads them to selectively attend to only those phonetic dimensions that are phonologically relevant in their L1. This might be implemented in Kruschke's model by readjusting weights, which would increase attention to dimensions phonologically relevant in the L2 and suppress attention to irrelevant dimensions.

If our proposal is correct and learners categorize L2 sounds based on their inferences about which phonetic dimensions are likely to be relevant in that language, then we expect listeners' perception and categorization of novel speech sounds to be guided by their experience with *phonetic dimensions*, and not individual segments. That is, in contrast to previous approaches, we predict that novel distinctions within a given phonetic dimension should be perceived more accurately by listeners who know a language in which that dimension is contrastive for *some* set of segments than by listeners for whom that dimension is never contrastive, even when the novel distinctions are used within segment classes for which the dimension is never contrastive for either group of listeners. This means that listeners are predicted to generalize the relevance of phonetic dimensions from known segments to novel segments. Pająk (2010) confirmed this prediction for the length dimension: for example, speakers of Cantonese, who are familiar with *vowel* length contrasts, are better at discriminating short from long *consonants* ([kasa]/[kassa]) than speakers of Mandarin, who are not familiar with any length contrasts. We took this result to suggest that Cantonese speakers generalized length across segment classes. This result is problematic for theories assuming L2-to-L1 segment mappings, under which familiarity with *vowel* length contrasts should not have any effect on perception and learning of *consonant* length contrasts: novel long consonants would be assumed to map onto L1 short consonants for both Cantonese and Mandarin speakers, thus making their discrimination equally difficult for *both* groups.

In this paper we investigate another type of evidence for the proposed model. One of the hallmark phenomena in the human categorization literature is the ability to learn category distinctions on the basis of purely distributional evidence—a bimodal distribution on some perceptual dimension, for example, generally supports the inference of a category distinction more strongly than a unimodal distribution—as predicted by rational accounts such as that underlying the *Size Principle* of Tenenbaum (1999) & Tenenbaum and Griffiths (2001). Since L2 phonological learning in our approach is simply a special case of the general problem of categorization, then distributional evidence may be able to overcome L1-derived bias against a category distinction in a phonetic dimension which is never distinctive in L1 (although the amount of exposure needed might differ depending on the cue's perceptual salience and on its L1 distribution). Crucially, generalization to a novel set of segments should straightforwardly follow from learning the relevance of a dimension for just one set of segments, exactly as in the previous study with Cantonese speakers. The perceptual learning literature provides mixed evidence on whether limited laboratory training can be sufficient to induce generalization in adults. Previous research focused on novel voicing distinctions (e.g., prevoiced vs. voiceless unaspirated stops for native speakers of English), and only tested limited types of generalization: for stops from one place of articulation to another (e.g., from alveolar [d]-[t] to velar [g]-[k]). Early studies with explicit category training showed that this type of generalization is possible (McClaskey, Pisoni, & Carrell, 1983; Tremblay, Kraus, Carrell, & McGee, 1997). On the other hand, training adults on a novel voicing distinction in the distributional learning paradigm (Maye & Gerken, 2000; Maye et al., 2002) was inconclusive regarding the ability of participants to generalize to a different place of articulation: Maye and Gerken (2001) reported no generalization, but Perfors and Dunbar (2010) found some evidence of generalization by increasing the duration of training and using natural stimuli.

In this study we tested the predictions of our model using the distributional learning paradigm, in which listeners (here, monolingual English speakers) are exposed to a new language (L2) through listening to stimuli sampled from a continuum of sounds that vary along some phonetic dimension (here, segmental length). The stimuli are sampled from either a bimodal frequency distribution, suggesting that there are two categories along the continuum (here, short and long segments), or a unimodal distribution, suggesting only one category (and, thus, no contrast between short and long segments). Crucially, all participants are exposed to the same inventory of stimuli, differing only in relative frequency of occurrence among stimuli within the inventory. Thus, any differences between bimodal and unimodal conditions in subsequent testing must be due to participants' interpretation of the novel sounds as influenced by training and not just to auditory sensitization. Beyond its relevance for testing the distinctive predictions made by our model (as compared with the mapping models described earlier), this work contributes to the perceptual learning literature by investigating distributional learning on a previously unstudied phonetic dimension—*segmental length*—and generalization across segment classes (sonorants and obstruents). Unlike the voicing dimension, length cross-cuts a wide range of possible segments, and is not in any form contrastive in the participants' native language (English).[1]

Experiment

We trained monolingual English speakers on a novel phonological contrast, segmental length, using the distributional learning paradigm, as applied by Maye and Gerken (2001) in a study with adult participants. Subsequently, we tested their categorization of short and long segments for trained and untrained segment classes (sonorants and obstruents). We predicted that participants would generalize the relevance of length in sound categorization from a trained class to an untrained class.

Method

Participants 48 undergraduate students at UC San Diego participated in the experiment for course credit. They were all monolingual speakers of English, in most cases with some limited high school and/or college exposure to Spanish or French. Crucially, none of them had any exposure to any language that uses length contrastively. All participants reported no history of speech or hearing problems.

Materials The materials consisted of nonce words recorded in a soundproof booth by a phonetically-trained native speaker of Polish. The critical length items included segments from two classes: sonorants ([j], [l], [m], [n]), and obstruents ([s], [f], [θ], [ʃ]). They were recorded as words with long consonants: [ajja], [illa], [amma], [inna], [assa], [iffa], [aθθa], [iʃʃa]. Subsequently, the consonant length in each word was manipulated to create length continua, each with eight tokens. There are several ways in which such continua could be created. One way would be to maintain natural between-segment duration differences (e.g., sonorant consonants are generally shorter than fricatives[2]), but manipulate relative durations so that for each continuum the endpoints are always in the same duration ratio (cross-linguistically, the long-to-short consonant ratio varies between 1.5 to 3; Ladefoged & Maddieson, 1996). Another way, which we adopted, is to use the same distribution on absolute durations for all segments (see the discussion section for more on the consequences of this choice). In the continua we created, durations of all consonants ranged from 100msec (short) to 205msec (long), and each adjacent token differed by 15msec. The fillers resembled the critical items, but different consonants were used: [iɾa], [iʔa], [aʤa], [aʧa], [iʥa], [iʦa], [aba], [apa], [ida], [ita], [aga], [aka], [ixa], [iχa], [aʁa], [aʕa].

Procedure The experiment adhered as closely as possible to the procedure used by Maye and Gerken (2001), and consisted of two main parts: training and testing.

Training: In training, participants listened to single words presented over headphones that were of one of two STIMULUS TYPES: *critical* or *filler*. Each participant was trained on critical items from one TRAINED SEGMENT CLASS (either *sonorants* or *obstruents*), and in one of two CONDITIONS: (1) *bimodal*, imitating a language with phonemic contrasts between short and long consonants, and (2) *unimodal*, imitating a language with no phonemic length contrasts (see Fig. 2). All participants were trained on the same filler items: the words [iɾa], [iʔa], [aʤa], [aʧa]. To maintain participants' attention on the experimental items, they were instructed to push a button after they heard each word. The response to a given stimulus triggered the presentation of the following stimulus with a delay of 1sec. Training consisted of a total of 384 words and lasted for about 10min. This included four repetitions of a training block, where each block had 64 critical items (16 tokens from each length continuum) and 32 filler items (8 different recordings of each item). Stimulus order was randomized for each participant, and there was a self-terminated break after each block.

Testing: The testing was identical for all participants, and consisted of an AX discrimination task. Participants listened to pairs of words, and were asked to judge whether these were two different words or two repetitions of the same word. For critical pairs, these were endpoints of each continuum, either 'different' (100msec–205msec, 205msec–100msec) or 'same' (100msec–100msec, 205msec–205msec). For filler 'different' pairs, these were two words that differed by one

[1] Although English vowels do vary in length, and length is used by native speakers as an auxiliary cue for voicing in word-final stops, vowel length alone is never used to distinguish between two vowel categories. This is reflected in how English native speakers process length: by 18 months of age English-learning infants show differences in their sensitivity to the length cue compared to infants learning a language that has phonemic length contrasts, such as Dutch or Japanese (Dietrich, Swingley, & Werker, 2007; Mugitani, Pons, Fais, Werker, & Amano, 2008).

[2] The ranges of duration for English consonants that are equivalent to those used in the experiment are roughly the following (in msec): [j] 39-100, [l] 42-85, [m] 50-89, [n] 38-83, [s] 61-126, [f] 88-138, [θ] 46-90, [ʃ] 88-138 (based on the phonetically annotated portion of the Switchboard corpus, as described in 'The Switchboard Transcription Project' report by Steven Greenberg, 1996.)

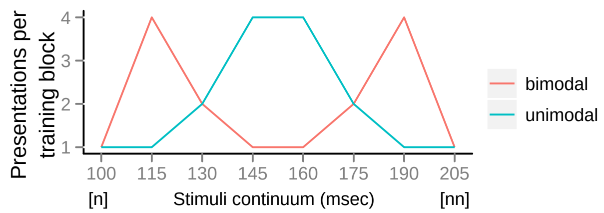

Figure 2: Critical training stimuli.

segment: the contrasts were either in voicing ([ʤ]–[ʧ], [ʥ]–[ʦ], [b]–[p], [d]–[t], [g]–[k]), in place of articulation ([x]–[χ], [ʁ]–[ʕ]), or in both ([ɾ]–[ʔ]). The 'same' pairs were always physically identical. The TESTED WORDS were of one of two types: *trained* (i.e., heard in training) or *untrained* (i.e., heard for the first time in testing). There was a total of 384 word pairs, which included 6 repetitions of a testing block. One block consisted of 32 critical pairs (16 'same' and 16 'different') and 32 filler pairs (16 'same' and 16 'different'). The words in each pair were separated by an interstimulus interval of 750ms. As with training, stimulus order was randomized for each participant, and there was a self-terminated break after each block. Participants responded by pushing a button on a gamepad. They were instructed to respond according to their intuition based on what they learned during the training period, and were assured that there were no strictly right or wrong answers. The instructions included a short practice with 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. Testing lasted about 20min.

Results

We predicted that successful distributional training should lead to a difference between the bimodal and the unimodal conditions on critical length trials: bimodal training resulting in more 'different' responses (since the training should suggest that short and long consonants are contrastive in this language), while unimodal training leading to fewer 'different' responses (because the training provided no evidence that short and long consonants belong to different categories). Furthermore, we predicted that participants would generalize the relevance of length from trained to untrained words (reflected in no difference in performance on trained and untrained items), and that this generalization would be bidirectional (i.e., from sonorants to obstruents, and vice versa).

Since performance was at ceiling on 'same' trials (>95% correct for each TYPE, CONDITION, TRAINED SEGMENT CLASS, and TESTED WORDS type), we only analyzed the responses from 'different' trials, using mixed-effects logit models with random slopes and intercepts for participant and item.[3]

[3]We also performed ANOVA analyses and found no major differences in results. Minor discrepancies are reported in footnotes.

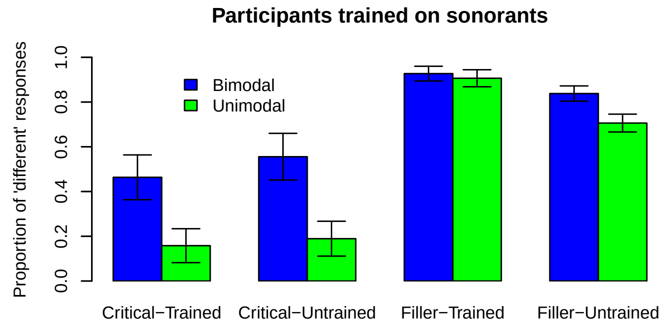

Figure 3: Performance by participants trained on the <u>sonorant class</u>. (Error bars are standard errors.)

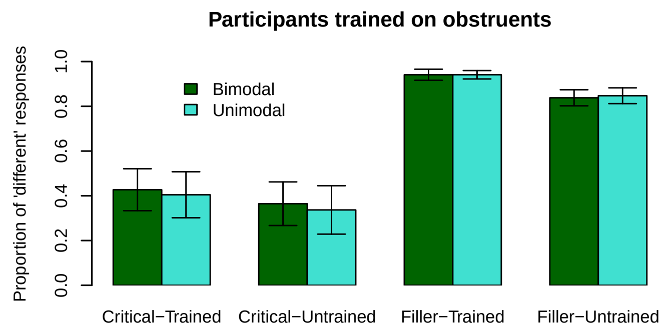

Figure 4: Performance by participants trained on the <u>obstruent class</u>. (Error bars are standard errors.)

First, we examined the critical trials for the fixed effects of CONDITION (*bimodal*, *unimodal*), TESTED WORD (*trained*, *untrained*), and TRAINED SEGMENT CLASS (*sonorant*, *obstruent*). There was a main effect of CONDITION ($p < .05$): as predicted, participants in the bimodal condition responded 'different' more often than in the unimodal condition. However, there was also a significant interaction between CONDITION and TRAINED SEGMENT CLASS ($p < .05$): the difference between the bimodal and the unimodal conditions was driven mainly by the participants trained on the sonorant class.[4] That is, as can be seen in the left part of Fig. 3, participants trained on sonorants responded 'different' more often in the bimodal than in the unimodal condition. However, as illustrated in the left part of Fig. 4, all participants trained on obstruents performed similarly regardless of the condition, even on the trained items. These results suggest that the distributional training was successful when it was done on sonorant length continua, but not when the training continua involved obstruents, in which case there was no difference between the bimodal and the unimodal conditions on any tested words: whether critical or filler, or trained and untrained.

Since the training was only successful for the sonorant-trained participants, we examined the critical trials for the effect of generalization for this group alone. We used a mixed

[4]Both of these effects were only marginal in ANOVAs with $p = .06$ and $p = .08$, respectively.

model with fixed effects of CONDITION (*bimodal*, *unimodal*) and TESTED WORD (*trained*, *untrained*). As expected by previous main effect, there was a significant main effect of CONDITION ($p < .01$) with participants in the bimodal condition responding 'different' more often than in the unimodal condition. Furthermore, as predicted by the generalization hypothesis, there was no significant main effect of TESTED WORD, meaning that participants in both bimodal and unimodal conditions performed similarly on trained and untrained items. Separate pairwise comparisons revealed that the difference between bimodal and unimodal conditions was significant for both trained and untrained critical items (ps$< .01$). These results suggest that participants generalized length to the novel segment class.

This effect was not due to a simple bias of bimodally-trained participants to respond 'different' on any trial, as reflected by a significant interaction between CONDITION and STIMULUS TYPE (*critical*, *filler*) ($p < .05$), as well as the same interaction for only untrained items ($p < .05$)[5]: the difference between the bimodal and the unimodal conditions was significantly larger for the critical than for the filler trials, even when just the untrained items were considered.

The fact that testing was identical for all participants, but the distributional training was only successful for the sonorant-trained group, and not for the obstruent-trained one, allows us to make a direct comparison between the two groups. By treating the performance of the obstruent-trained group as a baseline (38% 'different' responses), we can see the net effect of bimodal vs. unimodal training by comparing the performance of sonorant-trained participants to the baseline. This comparison reveals that successful bimodal training increased 'different' responses by 13%, whereas successful unimodal training decreased 'different' responses by 21%.

Discussion

This study yielded two key results. First, monolingual speakers of English can be trained through purely distributional learning to recognize a phonological category distinction on a phonetic dimension (segmental length) which is never contrastive in their native language. After only one ten-minute training session of 256 critical items, participants exposed to sonorants sampled from a bimodally distributed length continuum categorized words differing only in sonorant length as being distinctive more often than did participants exposed to sonorants of unimodally distributed length. Second—and even more crucially to our model's predictions—speakers generalized the relevance of length for sound categorization to a different set of consonants, obstruents. This generalization was quite aggressive, with the effect on obstruent categorization during testing just as strong as the effect on sonorant categorization. This result seems not to be reducible to greater general sensitization to any phonetic distinctions for the bimodally trained group, since the effect on performance for fillers—even those to which participants received no exposure during training—was smaller (though this comparison must be taken with caution since performance for fillers was higher across the board than for critical trials). This result contrasts with Maye and Gerken's (2001) study of distributional learning of a novel voicing distinction, where no evidence of generalization was found.[6] Since Maye and Gerken only used one segment continuum for training, our results suggest that training on a wider range of segments might yield stronger generalization.

For participants trained on obstruents, in contrast, the choice of bimodal versus unimodal distribution of segment length had no discernible effect on word categorization. We believe the most likely reason for this is related to the differences in duration between these two classes of consonants in naturally spoken English: obstruents (or at least all fricatives that we used in the experiment) are generally longer than sonorants. Since we created uniform length continua for both segment classes, this meant that all the tokens from the sonorant continua were longer than their usual duration range in English, while for obstruents these ranges partially overlapped. We believe that this might have been the reason why the obstruent-trained participants did not pick up on the distributional information: they may have heard the fricatives of around 200msec as unusually long, but still interpreted them as within reasonable English-like duration range, which consequently was not sufficient for bimodally-trained participants to infer contrastiveness of the length dimension. If this is correct, then modifying the obstruent continua (by including longer durations) should be more effective in guiding participants' inferences. Preliminary data from a follow-up experiment (N=11) suggest that this is indeed the case: when the obstruent continua range from 140msec to 280msec, the results for obstruent-trained participants look similar to those for sonorant-trained participants in the experiment reported in this paper. In the face of the learning failure observed in the present experiment for obstruent-trained participants, the generalization by sonorant-trained participants to obstruents is all the more impressive: distributional evidence as to whether length is contrastive for sonorants informs participants' perception of obstruent length contrastiveness even within a range of the continuum which would not itself drive learning through exposure to obstruents themselves.

The results reported in this paper are problematic for traditional mapping approaches to L2 sound perception and learning because these approaches have no straightforward explanation of distributional learning, much less of generalization. If we assume that phonological categorization of novel sounds proceeds through mapping of these sounds onto the most similar L1 categories, then frequency of exposure to sounds from a given phonetic continuum (as in distributional learning) should not have any effect on how the end-

[5] For these cases the models with the full random effects structure failed to converge. Thus, we iteratively removed random effects with the smallest variance until convergence was successful.

[6] Perfors and Dunbar (2010) did obtain both learning and generalization for a voicing distinction similar to Maye and Gerken's, but they used much more training and no fillers.

points of that continuum are mapped. Our results show, however, a clear difference in responses between bimodally- and unimodally-trained participants. Furthermore, under mapping approaches there is no reason why exposure to novel stimuli from one segment class should affect perception and categorization of stimuli from another class. Yet our results show this exact kind of dependency.

Conclusion

In this paper we described a model of L2 phonological acquisition, in which learners are assumed to use their previous language knowledge, combined with statistical properties of the novel language, to make implicit predictions about the underlying structure of the phonological system of that language. We predicted that learners should be able to infer, during a short period of exposure, that if a given phonetic dimension is contrastive for some set of segments, then it is also possibly contrastive for a different set of segments in that language. Consequently, listeners trained to attend to a given dimension for some segments should also be able to attend to this dimension for novel segments. We tested this prediction of the model by training monolingual speakers of English on a novel phonological contrast (segmental length), and then testing them on categorization of the contrasted segments for both a trained segment class and an untrained class. We showed that participants were able to infer a phonological contrast on this dimension even though the dimension is never contrastive in their native language, and that they generalized length from one class to another (from sonorant to obstruent consonants), suggesting that they were able to infer that length cross-cuts a wide range of segments. These results support our approach to understanding L2 learning as a process of inductive inference.

Acknowledgments

For their helpful feedback the authors thank Eric Baković, Klinton Bicknell, Sarah Creel, Vic Ferreira, Tamar Gollan, Sharon Rose, and four anonymous CogSci reviewers. K. Michael Brooks, Annabelle Cadang, 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.

References

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.

Dietrich, C., Swingley, D., & Werker, J. F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences*, *104*, 16027-16031.

Flege, J. (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.

Gerken, L. (2010). Infants use rational decision criteria for choosing among models of their input. *Cognition*, *115*, 362–366.

Hancin-Bhatt, B. J. (1994). Segment transfer: A consequence of a dynamic system. *Second Language Research*, *10*, 241–269.

Jusczyk, P. W. (1992). Developing phonological categories from the speech signal. In C. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: models, research, implications* (pp. 17–64). Timonium, MD: York Press.

Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22–44.

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.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*, 606–608.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, UK; Cambridge, MA: Blackwell.

MacWhinney, B. (1987). The Competition Model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Hillsdale, NJ: Lawrence Erlbaum.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development* (pp. 522–533). Somerville, MA: Cascadilla Press.

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

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111.

McClaskey, C. L., Pisoni, D. B., & Carrell, T. D. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception and Psychophysics*, *34*(4), 323–330.

Mugitani, R., Pons, F., Fais, L., Werker, J. F., & Amano, S. (2008). Perception of vowel length by Japanese- and English-learning infants. *Developmental Psychology*, *45*(1), 236–247.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57.

Pająk, B. (2010). Perceptual advantage from generalized linguistic knowledge. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 369–374). Austin, TX: Cognitive Science Society.

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

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3), 353–363.

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

Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. In M. S. Kearns, S. A. Solla, & D. A. Cohn (Eds.), *Advances in neural information processing systems* (Vol. 11). Cambridge, MA: MIT Press.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and bayesian inference. *Behavioral and Brain Sciences*, *24*, 629–640.

Tremblay, K., Kraus, N., Carrell, T. D., & McGee, T. (1997). Central auditory system plasticity: generalization to novel stimuli following listening training. *Journal of the Acoustical Society of America*, *102*(6), 3762–3773.

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