

A model of generalization in distributional learning of phonetic categories

Bozena Pajak

Brain & Cognitive Sciences
University of Rochester
Rochester, NY 14627-0268
bpajak@bcs.rochester.edu

Klinton Bicknell

Psychology
UC San Diego
La Jolla, CA 92093-0109
kbicknell@ucsd.edu

Roger Levy

Linguistics
UC San Diego
La Jolla, CA 92093-0108
rlevy@ucsd.edu

Abstract

Computational work in the past decade has produced several models accounting for phonetic category learning from distributional and lexical cues. However, there have been no computational proposals for how people might use another powerful learning mechanism: generalization from learned to analogous distinctions (e.g., from /b/–/p/ to /g/–/k/). Here, we present a new simple model of generalization in phonetic category learning, formalized in a hierarchical Bayesian framework. The model captures our proposal that linguistic knowledge includes the possibility that category types in a language (such as voiced and voiceless) can be shared across sound classes (such as labial and velar), thus naturally leading to generalization. We present two sets of simulations that reproduce key features of human performance in behavioral experiments, and we discuss the model’s implications and directions for future research.

1 Introduction

One of the central problems in language acquisition is how phonetic categories are learned, an unsupervised learning problem involving mapping phonetic tokens that vary along continuous dimensions onto discrete categories. This task may be facilitated by languages’ extensive re-use of a set of phonetic dimensions (Clements 2003), because learning one distinction (e.g., /b/–/p/ varying along the voice onset time (VOT) dimension) might help learn analogous distinctions (e.g., /d/–/t/, /g/–/k/). Existing experimental evidence supports this view: both infants and adults generalize newly learned phonetic category distinctions to untrained sounds along the same dimension (McClaskey et al. 1983, Maye et al. 2008, Perfors

& Dunbar 2010, Pajak & Levy 2011a). However, while many models have been proposed to account for learning of phonetic categories (de Boer & Kuhl 2003, Vallabha et al. 2007, McMurray et al. 2009, Feldman et al. 2009, Toscano & McMurray 2010, Dillon et al. 2013), there have been no computational proposals for how generalization to analogous distinctions may be accomplished. Here, we present a new simple model of generalization in phonetic category learning, formalized in a hierarchical Bayesian framework. The model captures our proposal that linguistic knowledge includes the possibility that category types in a language (such as voiced and voiceless) can be shared across sound classes (defined as previously learned category groupings, such as vowels, consonants, nasals, fricatives, etc.), thus naturally leading to generalization.

One difficulty for the view that learning one distinction might help learn analogous distinctions is that there is variability in how the same distinction type is implemented phonetically for different sound classes. For example, VOT values are consistently lower for labials (/b/–/p/) than for velars (/g/–/k/) (Lisker & Abramson 1970), and the durations of singleton and geminate consonants are shorter for nasals (such as /n/–/nn/) than for voiceless fricatives (such as /s/–/ss/) (Giovanardi & Di Benedetto 1998, Mattei & Di Benedetto 2000). Improving on our basic model, we implement a modification that deals with this difficulty by explicitly building in the possibility for analogous categories along the same dimension to have different absolute phonetic values along that dimension (e.g., shorter overall durations for nasals than for fricatives).

In Section 2 we discuss the relevant background on phonetic category learning, including previous modeling work. Section 3 describes our basic computational model, and Section 4 presents simulations demonstrating that the model can re-

produce the qualitative patterns shown by adult learners in cases when there is no phonetic variability between sound classes. In Section 5 we describe the extended model that accommodates phonetic variability across sound classes, and in Section 6 we show that the improved model qualitatively matches adult learner performance both when the sound classes implement analogous distinction types in identical ways, and when they differ in the exact phonetic implementation. Section 7 concludes with discussion of future research.

2 Background

One important source of information for unsupervised learning of phonetic categories is the shape of the distribution of acoustic-phonetic cues. For example, under the assumption that each phonetic category has a unimodal distribution on a particular cue, the number of modes in the distribution of phonetic cues can provide information about the number of categories: a unimodal distribution along some continuous acoustic dimension, such as VOT, may indicate a single category (e.g., /p/, as in Hawaiian); a bimodal distribution may suggest a two-category distinction (e.g., /b/ vs. /p/, as in English); and a trimodal distribution implies a three-category distinction (e.g., /b/, /p/, and /p^h/, as in Thai). Infants extract this distributional information from the speech signal (Maye et al. 2002, 2008) and form category representations focused around the modal values of categories (Kuhl 1991, Kuhl et al. 1992, Lacerda 1995). Furthermore, information about some categories bootstraps learning of others: infants exposed to a novel bimodal distribution along the VOT dimension for one place of articulation (e.g., alveolar) not only learn that novel distinction, but also generalize it to an analogous contrast for another (e.g., velar) place of articulation (Maye et al. 2008). This ability is preserved beyond infancy, and is potentially used during second language learning, as adults are also able to both learn from distributional cues and use this information when making category judgments about untrained sounds along the same dimensions (Maye & Gerken 2000, 2001, Perfors & Dunbar 2010, Pajak & Levy 2011a,b).

The phonetic variability in how different sound classes implement the same distinction type might in principle hinder generalization across classes. However, there is evidence of generalization even in cases when sound classes differ in the exact

phonetic implementation of a shared distinction type. For example, learning a singleton/geminate length contrast for the class of voiceless fricatives (e.g., /s/–/ss/, /f/–/ff/) generalizes to the class of sonorants (e.g., /n/–/nn/, /j/–/jj/) even when the absolute durations of sounds in the two classes are different – overall longer for fricatives than for sonorants (Pajak & Levy 2011a) – indicating that learners are able to accommodate the variability of phonetic cues across different sound classes.

Phonetic categorization from distributional cues has been modeled using Gaussian mixture models, where each category is represented as a Gaussian distribution with a mean and covariance matrix, and category learning involves estimating the parameters of each mixture component and – for some models – the number of components (de Boer & Kuhl 2003, Vallabha et al. 2007, McMurray et al. 2009, Feldman et al. 2009, Toscano & McMurray 2010, Dillon et al. 2013).¹ These models are successful at accounting for distributional learning, but do not model generalization. We build on this previous work (specifically, the model in Feldman et al. 2009) and implement generalization of phonetic distinctions across different sound classes.

3 Basic generalization model

The main question we are addressing here concerns the mechanisms underlying generalization. How do learners make use of information about some phonetic categories when learning other categories? Our proposal is that learners expect category types (such as singleton and geminate, or voiced and voiceless) to be shared among sound classes (such as sonorants and fricatives). We implement this proposal with a hierarchical Dirichlet process (Teh et al. 2006), which allows for sharing categories across data groups (here, sound classes). We build on previous computational work in this area that models phonetic categories as Gaussian distributions. Furthermore, we follow Feldman et al. (2009) in using Dirichlet processes (Ferguson 1973), which allow the model to learn the number of categories from the data, and implementing the process of learning from distributional cues via nonparametric Bayesian inference.

¹In Dillon et al. (2013) each phoneme is modeled as a mixture of Gaussians, where each component is an allophone.

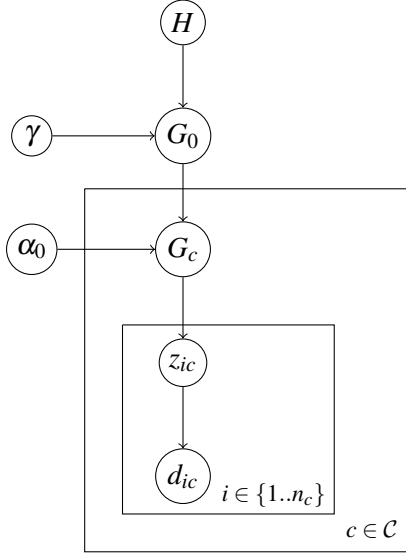


Figure 1: The graphical representation of the basic model.

$$\begin{aligned}
 H: \quad \mu &\sim \mathcal{N}(\mu_0, \frac{\sigma^2}{\kappa_0}) \\
 &\sigma^2 \sim \text{InvChiSq}(v_0, \sigma_0^2) \\
 G_0 &\sim \text{DP}(\gamma, H) \\
 G_c &\sim \text{DP}(\alpha_0, G_0) \\
 z_{ic} &\sim G_c \\
 d_{ic} &\sim \mathcal{N}(\mu_{z_{ic}}, \sigma_{z_{ic}}^2) \\
 \hline
 f_c &\sim \mathcal{N}(0, \sigma_f^2) \\
 d_{ic} &\sim \mathcal{N}(\mu_{z_{ic}}, \sigma_{z_{ic}}^2) + f_c
 \end{aligned}$$

Figure 2: Mathematical description of the model. The variables below the dotted line refer to the extended model in Figure 6.

3.1 Model details

As a first approach, we consider a simplified scenario of a language with a set of sound classes, each of which contains an unknown number of phonetic categories, with perceptual token defined as a value along a single phonetic dimension. The model learns the set of phonetic categories in each sound class, and the number of categories inferred for one class can inform the inferences about the other class. Here, we make the simplifying assumption that learners acquire a context-independent distribution over sounds, although the model could be extended to use linguistic context (such as coarticulatory or lexical information; Feldman et al. 2009).

Figure 1 provides the graphical representation of the model, and Figure 2 gives its mathematical

Variable	Explanation
H	base distribution over means and variances of categories
G_0	distribution over possible categories
G_c	distribution over categories in class c
γ, α_0	concentration parameters
z_{ic}	category for datapoint d_{ic}
d_{ic}	datapoint (perceptual token)
n_c	number of datapoints in class c
\mathcal{C}	set of classes
f_c	offset parameter
σ_f	standard deviation of prior on f_c

Table 1: Key for the variables in Figures 1, 2, and 6. The variables below the dotted line refer to the extended model in Figure 6.

description. Table 1 provides the key to the model variables. In the model, speech sounds are produced by selecting a phonetic category z_{ic} , which is defined as a mean $\mu_{z_{ic}}$ and variance $\sigma_{z_{ic}}^2$ along a single phonetic dimension,² and then sampling a phonetic value from a Gaussian with that mean and variance. We assume a weak prior over categories that does not reflect learners’ prior language knowledge (but we return to the possible role of prior language knowledge in the discussion). Learners’ beliefs about the sound inventory (distribution over categories and mean and variance of each category) are encoded through a hierarchical Dirichlet process. Each category is sampled from the distribution G_c , which is the distribution over categories in a single sound class. In order to allow sharing of categories across classes, the G_c distribution for each class is sampled from a Dirichlet process with base distribution G_0 , which is shared across classes, and concentration parameter α_0 (which determines the sparsity of the distribution over categories). G_0 , then, stores the full set of categories realized in any class, and it is sampled from a Dirichlet process with concentration parameter γ and base distribution H , which is a normal inverse chi-squared prior on category

²Although we are modeling phonetic categories as having values along a single dimension, the model can be straightforwardly extended to multiple dimensions, in which case the variance would be replaced by a covariance matrix $\Sigma_{z_{ic}}$.

means and variances.³ The parameters of the normal inverse chi-squared distribution are: ν_0 and κ_0 , which can be thought of as pseudo-observations, as well as μ_0 and σ_0^2 , which determine the prior distribution over means and variances, as in Figure 2.

3.2 Inference

The model takes as input the parameters of the base distribution H , the concentration parameters α_0 and γ , and the data, which is composed of a list of phonetic values. The model infers a posterior distribution over category labels for each datapoint via Gibbs sampling. Each iteration of Gibbs sampling resamples the assignments of each datapoint to a lower-level category (in G_c) and also resamples the assignments of lower-level categories to higher-level categories (in G_0). We marginalize over the category means and variances.

4 Simulations: basic model

The first set of simulations has three goals: first, to establish that our model can successfully perform distributional learning and second, to show that it can use information about one type of class to influence judgements about another, in the case that there is no variability in category structure between classes. Finally, these simulations reveal a limitation of this basic model, showing that it cannot generalize in the presence of substantial between-class variability in category realizations. We address this limitation in Section 5.

4.1 The data

The data we use to evaluate the model come from the behavioral experiments in Pajak & Levy (2011a). Adult native English speakers were exposed to novel words, where the middle consonant varied along the length dimension from short (e.g., [ama]) to long (e.g., [amma]). The distributional information suggested either one category along the length dimension (unimodal distribution) or two categories (bimodal distribution), as illustrated in Figure 3. In Experiment 1, the training included sounds in the sonorant class (4 continua: [n]-...-[nn], [m]-...-[mm], [j]-...-[jj], [l]-...-[ll]) with the duration range of 100–205msec. In Experiment 2 the training included sounds in

³In the case of categories defined along multiple dimensions, the base distribution would be a normal inverse-Wishart.

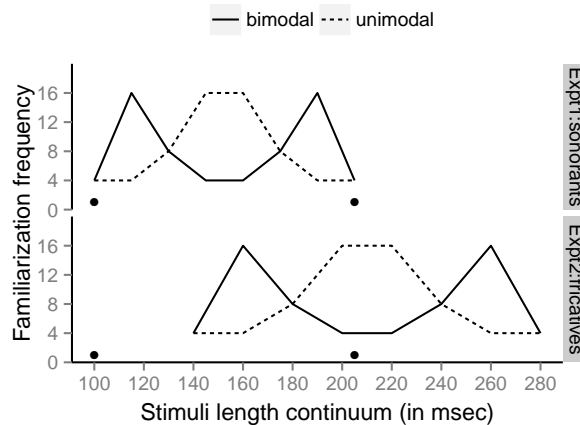


Figure 3: Experiment 1 & 2 training (Pajak and Levy 2011a). The y axis reflects the frequency of tokens from each training continuum. The four points indicate the values of the untrained datapoints.

the voiceless fricative class (4 continua: [s]-...-[ss], [f]-...-[ff], [θ]-...-[θθ], [ʃ]-...-[ʃʃ]) with the duration range of 140–280msec. The difference in duration ranges between the two classes reflected the natural duration distributions of sounds in these classes: generally shorter for sonorants and longer for fricatives (Greenberg 1996, Giovanardi & Di Benedetto 1998, Mattei & Di Benedetto 2000).

Subsequently, participants’ expectations about the number of categories in the trained class and another untrained class were probed by asking for judgments about tokens at the endpoints of the continua: participants were presented with pairs of words (e.g., sonorant [ama]–[amma] or fricative [asa]–[assa]) and asked whether these were two different words in this language or two repetitions of the same word. As illustrated in Table 2, in the test phase of Experiment 1 the durations of both the trained and the untrained class were identical (100msec for any short consonant and 205msec for any long consonant), whereas in the test phase of Experiment 2 the durations were class-specific: longer for trained fricatives (140msec for a short fricative and 280msec for a long fricative) and shorter for untrained sonorants (100msec for a short sonorant and 205msec for a long sonorant).

The experiment results are illustrated in Figure 4. The data from the ‘trained’ condition shows that learners were able to infer the number of categories from distributional cues: they were more

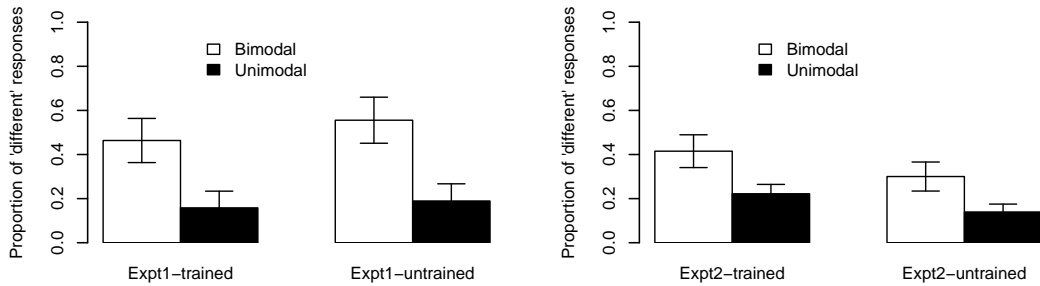


Figure 4: Experiment 1 & 2 results: proportion of ‘different’ responses on ‘different’ trials (Pajak and Levy, 2011a).

	Expt. 1	Expt. 2
trained	(sonorants) 100ms – 205ms	(fricatives) 140ms – 280ms
untrained	(fricatives) 100ms – 205ms	(sonorants) 100ms – 205ms

Table 2: Experiment 1 & 2 test (Pajak and Levy, 2011a).

likely to posit two categories (i.e., respond ‘different’ on ‘different’ trials) when the distribution was bimodal than when the distribution was unimodal. In addition, as demonstrated by the ‘untrained’ condition, learners used the information about the trained class to make inferences about the untrained class: they were more likely to accept length-based category distinctions for fricatives after learning the distinction for sonorants (Expt. 1), and vice versa (Expt. 2). This generalization occurred both (a) when each class implemented the distinction in exactly the same way (with the same absolute durations; Expt. 1), and (b) when the classes differed in how the shared distinction type was implemented (the absolute durations of the untrained class were shifted relative to the trained class; Expt. 2).

The model simulations described below attempt to replicate the key features of human performance: distributional learning and generalization. We model both experiments of Pajak & Levy (2011a): (a) ‘same durations’ across classes (Expt. 1), and (b) ‘different durations’ across classes (Expt. 2). Thus, the datasets we used were closely modeled after their experimental design: (1) Expt. 1 bimodal, (2) Expt. 1 unimodal, (3) Expt. 2 bimodal, and (4) Expt. 2 unimodal. In each dataset, the data consisted of a list of phonetic values (duration in msec), where each data-

point was tagged as belonging to either the sonorant or the fricative class. The frequencies of the ‘trained’ class were as listed in Figure 3 (simulating a single training continuum). In addition to the ‘trained’ class, each dataset included two datapoints from the ‘untrained’ class with the values as listed in Table 2 in the ‘untrained’ condition. These two datapoints were included in order to evaluate the model’s categorization of sounds for which no distributional evidence is available, thus assessing the extent of generalization. We simulated weak perceptual noise by adding to each datapoint normally-distributed error with standard deviation of 0.3 times the distance between adjacent continuum steps.

4.2 Methodology

We ran the basic model on each of the four datasets. For each, we performed 1,000,000 iterations of Gibbs sampling, and analyzed the results for the second half. To assess convergence, we ran four Markov chains for each dataset, using two overdispersed initializations: (1) assigning one category label to all datapoints, and (2) assigning a different label to each datapoint. We used a weak prior base distribution H ($\kappa_0 = .001$; $\nu_0 = .001$; $\sigma_0^2 = 1$; μ_0 was set to the overall mean of the data), and set the concentration parameters $\gamma = \alpha_0 = 1$.

4.3 Results and discussion

The simulation results are illustrated in Figure 5,⁴ plotting the proportion of samples on which the model assigned the datapoints to two different categories, as opposed to a single category.⁵ Note that

⁴All variables we report in all simulations appear to have converged to the posterior, as assessed by \hat{R} values of 1.1 or less, calculated across the 4 chains (Gelman & Rubin 1992).

⁵No models we report assign the trained category datapoints to more than two categories more than 1% of the time.

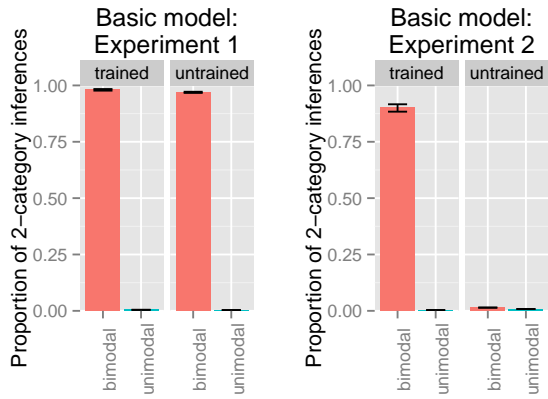


Figure 5: Simulation results for the basic model. Error bars give 95% binomial confidence intervals, computed using the estimated number of effectively independent samples in the Markov chains.

in the ‘trained’ condition, this means categorization of all datapoints along the continuum. In the ‘untrained’ condition, on the other hand, it is categorization of two datapoints: one from each endpoint of the continuum.

The results in the ‘trained’ conditions demonstrate that the model was able to learn from the distributional cues, thus replicating the success of previous phonetic category learning models.

Of most interest here are the results in the ‘untrained’ condition. The figure on the left shows the results modeling the ‘same-durations’ experiment (Expt. 1), demonstrating that the model categorizes the two datapoints in the untrained sound class in exactly the same way as it did for the trained sound class: two categories in the bimodal condition, and one category in the unimodal condition. Thus, these results suggest that we can successfully model generalization of distinction types across sound classes in phonetic category learning by assuming that learners have an expectation that category types (such as short and long, or voiceless and voiced) may be shared across classes.

The figure on the right shows the results modeling the ‘different-durations’ experiment (Expt. 2), revealing a limitation of the model: failure to generalize when the untrained class has the same category structure but different absolute phonetic values (overall shorter in the untrained class than in the trained class). Instead, the model categorizes both untrained datapoints as belonging to a single category. This result diverges from the experimental results, where learners generalize the learned distinction type in both cases, whether the abso-

lute phonetic values of the analogous categories are identical or not. We address this problem in the next section by implementing a modification to the model that allows more flexibility in how each class implements the same category types.

5 Extended generalization model

The goal of the extended model is to explicitly allow for phonetic variability across sound classes. As a general approach, we could imagine functions that transform categories across classes so that the same categories can be “reused” by being translated around to different parts of the phonetic space. These functions would be specific operations representing any intrinsic differences between sound classes. Here, we use a very simple function that can account for one widely attested type of transformation: different absolute phonetic values for analogous categories in distinct sound classes (Ladefoged & Maddieson 1996), such as longer overall durations for voiceless fricatives than for sonorants. This type of transformation has been successfully used in prior modeling work to account for learning allophones of a single phoneme that systematically vary in phonetic values along certain dimensions (Dillon et al. 2013).

5.1 Model details

We implement the possibility for between-class variability by allowing for one specific type of idiosyncratic implementation of categories across classes: learnable class-specific ‘offsets’ by which the data in a class are shifted along the phonetic dimension, as illustrated in Figure 6 (the key for the variables is in Table 1).

5.2 Inference

Each iteration of MCMC now includes a Metropolis-Hastings step to resample the offset parameters f_c , which uses a zero-mean Gaussian proposal, with standard deviation $\sigma_p = \frac{\text{range of data}}{5}$.

6 Simulations: extended model

This second set of simulations has two goals: (1) to establish that the extended model can successfully replicate the performance of the basic model in both distributional learning and generalization in the no-variability case, and (2) to show that explicitly allowing for variability across classes lets the model generalize when there is between-class variability in category realizations.

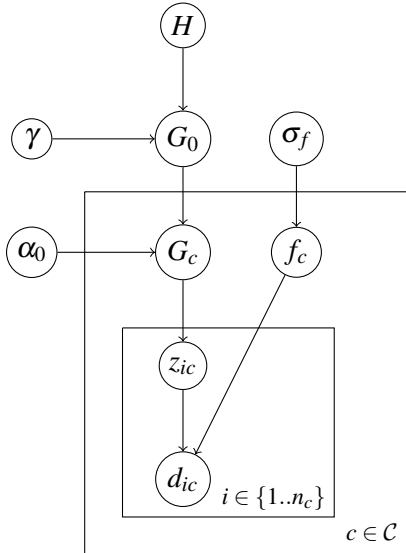


Figure 6: The graphical representation of the extended model.

6.1 Methodology

We used the same prior as in the first set of simulations, and used a Gaussian prior on the offset parameter with standard deviation $\sigma_f = 1000$. Because only the relative values of offset parameters are important for category sharing across classes, we set the offset parameter for one of the classes to zero. The four Markov chains now crossed category initialization with two different initial values of the offset parameter.

6.2 Results and discussion

The simulation results are illustrated in Figure 7. The figure on the left demonstrates that the extended model performs similarly to the basic model in the case of no variability between classes. The figure on the right, on the other hand, shows that – unlike the basic model – the extended model succeeds in generalizing the learned distinction type to an untrained sound class when there is phonetic variability between classes. These results suggest that allowing for variability in category implementations across sound classes may be necessary to account for human learning. Taken together, these results are consistent with our proposal that language learners have an expectation that category types can be shared across sound classes. Furthermore, learners appear to have implicit knowledge of the ways that sound classes can vary in their exact phonetic implementations of different category types. This

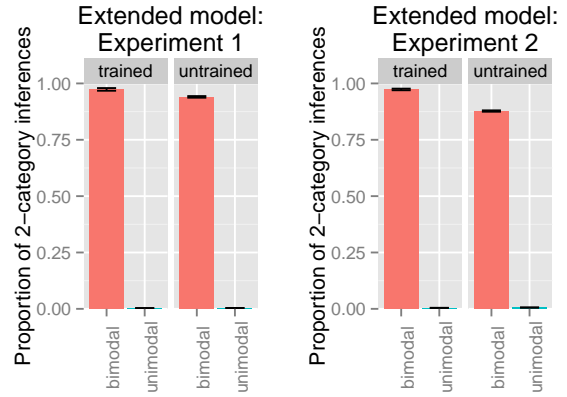


Figure 7: Simulation results for the extended model. Error bars give 95% binomial confidence intervals, computed using the estimated number of effectively independent samples in the Markov chains.

type of knowledge may include – as in our extended generalization model – the possibility that phonetic values of categories in one class can be systematically shifted relative to another.

7 General discussion

In this paper we presented the first model of generalization in phonetic category learning, in which learning a distinction type for one set of sounds (e.g., /m/–/mm/) immediately generalizes to another set of sounds (e.g., /s/–/ss/), thus reproducing the key features of adult learner performance in behavioral experiments. This extends previous computational work in phonetic category learning, which focused on modeling the process of learning from distributional cues, and did not address the question of generalization. The basic premise of the proposed model is that learners’ knowledge of phonetic categories is represented hierarchically: individual sounds are grouped into categories, and individual categories are grouped into sound classes. Crucially, the category structure established for one sound class can be directly shared with another class, although different classes can implement the categories in idiosyncratic ways, thus mimicking natural variability in how analogous categories (e.g., short /m/ and /s/, or long /mm/ and /ss/) are phonetically implemented for different sound classes.

The simulation results we presented succeed in reproducing the human pattern of generalization performance, in which the proportion of two-category inferences about the untrained class is

very similar to that for the trained class. Note, however, that there are clear quantitative differences between the two in learning performance: the model learns almost perfectly from the available distributional cues ('trained' condition), while adult learners are overall very conservative in accepting two categories along the length dimension, as indicated by the overall low number of 'different' responses. There are two main reasons why the model might be showing more extreme categorization preferences than humans in this particular task. First, humans have cognitive limitations that the current model does not, such as those related to memory or attention. In particular, imperfect memory makes it harder for humans to integrate the distributional information from all the trials in the exposure, and longer training would presumably improve performance. Second, adults have strong native-language biases that affect learning of a second language (Flege 1995). The population tested by Pajak & Levy (2011a) consisted of adult native speakers of American English, a language in which length is not used contrastively. Thus, the low number of 'different' responses in the experiments can be attributed to participants' prior bias against category distinctions based on length. The model, on the other hand, has only a weak prior that was meant to be easily overridden by data.

This last point is of direct relevance for the area of second language (L2) acquisition, where one of the main research foci is to investigate the effects of native-language knowledge on L2 learning. The model we proposed here can potentially be used to systematically investigate the role of native-language biases when learning category distinctions in a new language. In particular, an L2 learner, whose linguistic representations include two languages, could be implemented by adding a language-level node to the model's hierarchical structure (through an additional Dirichlet process). This extension will allow for category structures to be shared not just within a language for different sound classes, but also across languages, thus effectively acting as a native-language bias.

As a final note, we briefly discuss alternative ways of modeling generalization in phonetic category learning. In the model we described in this paper, whole categories are generalized from one class to another. However, one might imagine another approach to this problem where generaliza-

tion is a byproduct of learners' attending more to the dimension that they find to be relevant for distinguishing between some categories in a language. That is, learners' knowledge would not include the expectation that whole categories may be shared across classes, as we argued here, but rather that a given phonetic dimension is likely to be reused to distinguish between categories in multiple sound classes.

This intuition could be implemented in different ways. In a Dirichlet process model of category learning, the concentration parameter α might be learned, and shared for all classes along a given phonetic dimension, thus producing a bias toward having a similar number of categories across classes. Alternatively, the variance of categories along a given dimension might be learned, and also shared for all classes. Under this scenario, learning category variance along a given dimension would help categorize novel sounds along that dimension. That is, two novel datapoints would be likely categorized into separate categories if the inferred variance along the relevant dimension was smaller than the distance between the datapoints, but into a single category if the inferred variance was comparable to that distance.

Finally, this model assumes that sound classes are given in advance, and that only the categories within each class are learned. While this assumption may seem warranted for some types of perceptually dissimilar sound classes (e.g., consonants and vowels), and also may be appropriate for L2 acquisition, it is not clear that it is true for all sound classes that allow for generalization in infancy. It remains for future work to determine how learners may generalize while simultaneously learning the sound classes.

We plan to pursue all these directions in future work with the ultimate goal of improving our understanding how human learners represent their linguistic knowledge and how they use it when acquiring a new language.

Acknowledgments

We thank Gabriel Doyle and three anonymous CMCL reviewers for useful feedback. This research was supported by NIH Training Grant T32-DC000041 from the Center for Research in Language at UC San Diego to B.P. and NIH Training Grant T32-DC000035 from the Center for Language Sciences at University of Rochester to B.P.

References

- de Boer, Bart & Patricia K. Kuhl. 2003. Investigating the role of infant-directed speech with a computer model. *Acoustic Research Letters Online* 4(4). 129–134.
- Clements, George N. 2003. Feature economy in sound systems. *Phonology* 20. 287–333.
- Dillon, Brian, Ewan Dunbar & William Idsardi. 2013. A single-stage approach to learning phonological categories: Insights from Inuktitut. *Cognitive Science* 37. 344–377.
- Feldman, Naomi H., Thomas L. Griffiths & James L. Morgan. 2009. Learning phonetic categories by learning a lexicon. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 2208–2213. Austin, TX: Cognitive Science Society.
- Ferguson, Thomas S. 1973. A Bayesian analysis of some nonparametric problems. *Annals of Statistics* 1. 209–230.
- Flege, James E. 1995. Second-language speech learning: theory, findings and problems. In Winifred Strange (ed.), *Speech perception and linguistic experience: issues in cross-language research*, 229–273. Timonium, MD: York Press.
- Gelman, Andrew & Donald B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7. 457–511.
- Giovanardi, Maurizio & Maria-Gabriella Di Benedetto. 1998. Acoustic analysis of singleton and geminate fricatives in Italian. *The European Journal of Language and Speech (EACL/ESCA/ELSNET)* 1998. 1–13.
- Greenberg, Steven. 1996. The Switchboard transcription project. Report prepared for the 1996 CLSP/JHU Workshop on Innovative Techniques in Continuous Large Vocabulary Speech Recognition.
- Kuhl, Patricia K. 1991. Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics* 50(2). 93–107.
- Kuhl, Patricia K., Karen A. Williams, Francisco Lacerda, Kenneth N. Stevens & Björn Lindblom. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255. 606–608.
- Lacerda, Francisco. 1995. The perceptual magnet-effect: An emergent consequence of exemplar-based phonetic memory. In K. Ellenius & P. Branderud (eds.), *Proceedings of the 13th International Congress of Phonetic Sciences*, 140–147. Stockholm: KTH and Stockholm University.
- Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world’s languages*. Oxford, UK; Cambridge, MA: Blackwell.
- Lisker, Leigh & Arthur S. Abramson. 1970. The voicing dimensions: Some experiments in comparative phonetics. In *Proceedings of the Sixth International Congress of Phonetic Sciences*, Prague: Academia.
- Mattei, Marco & Maria-Gabriella Di Benedetto. 2000. Acoustic analysis of singleton and geminate nasals in Italian. *The European Journal of Language and Speech (EACL/ESCA/ELSNET)* 2000. 1–11.
- Maye, Jessica & LouAnn Gerken. 2000. Learning phonemes without minimal pairs. In S. Catherine Howell, Sarah A. Fish & Thea Keith-Lucas (eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development*, 522–533. Somerville, MA: Cascadilla Press.
- Maye, Jessica & LouAnn Gerken. 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*, 480–490. Somerville, MA: Cascadilla Press.
- Maye, Jessica, Daniel J. Weiss & Richard N. Aslin. 2008. Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science* 11(1). 122–134.
- 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(4). 323–330.
- McMurray, Bob, Richard N. Aslin & Joseph C. Toscano. 2009. Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science* 12(3). 369–378.
- Pajak, Bozena & Roger Levy. 2011a. How abstract are phonological representations? Evidence from distributional perceptual learning.

- In *Proceedings of the 47th Annual Meeting of the Chicago Linguistic Society*, Chicago, IL: University of Chicago.
- Pajak, Bozena & Roger Levy. 2011b. Phonological generalization from distributional evidence. In L. Carlson, C. Hölscher & T. Shipley (eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, 2673–2678. Austin, TX: Cognitive Science Society.
- Perfors, Amy & David Dunbar. 2010. Phonetic training makes word learning easier. In S. Ohlsson & R. Catrambone (eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, 1613–1618. Austin, TX: Cognitive Science Society.
- Teh, Yee Whye, Michael I. Jordan, Matthew J. Beal & David M. Blei. 2006. Hierarchical Dirichlet processes. *Journal of the American Statistical Association* 101(476). 1566–1581.
- Toscano, Joseph C. & Bob McMurray. 2010. Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science* 34. 434–464.
- Vallabha, Gautam K., James L. McClelland, Ferran Pons, Janet F. Werker & Shigeaki Amano. 2007. Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences* 104(33). 13273–13278.