

***Section 4: Integrating different perspectives:
New insights from production, perception and acquisition***

Chapter 15: Perception and comprehension

How perceptual and cognitive constraints affect learning of speech categories

Lori L. Holt
Department of Psychology
Carnegie Mellon University

1 Introduction

Categorization is an important facet of speech communication. However, we do not yet have a complete understanding of how speech categories are learned in infancy or adulthood. At least part of the reason for this is that it is not feasible to entirely control and manipulate speech to observe the consequences of different patterns of experience. Converging methods of cross-language observation, laboratory-based training of speech and nonspeech categories, and animal models of learning can provide a means of balancing the competing demands of ecological validity and experimental control to reveal how auditory and cognitive constraints affect speech category learning. The present half-chapter describes evidence from these approaches and explains how this evidence informs us about how general perceptual and cognitive constraints affect learning speech categories.

2 Speech Categorization

The notorious acoustic complexity of speech presents a challenge for listeners. Some of the acoustic variability of speech is linguistically-significant, but some is unrelated to the message. Even clear speech in a quiet room varies with talker affect, phonetic context, and room acoustics. Adding to the complexity, what counts as linguistically-significant is language

dependent. Thus, to extract a message from speech, listeners must accomplish two complementary perceptual feats. They must discriminate linguistically-relevant acoustic variability and generalize across irrelevant variability. Said another way, listeners must *categorize* speech in a manner specific to their language (see Holt and Lotto, 2008). Since the mapping of acoustic variability is language-specific, these categories must be learned from experience with speech.

Young infants from different language communities respond to speech in a way that is more similar than different, discriminating sounds without respect to whether they are phonemically distinctive in the ambient language (Jusczyk, 1997 for review; **Demuth and Song**, **Munson et al. **, **Maye**, this volume). In stark contrast, adults have difficulty discriminating even highly acoustically-distinct differences between some non-native sounds (**Iverson**, **Smiljanic** this volume). Native Japanese adults, for example, are poor at discriminating English /r/ versus /l/ (Miyawaki et al., 1975), although 6- to 8-month-old Japanese-learning infants discriminate the sounds as well as English-learning infants (Kuhl et al., 2006). Experience with a native language shapes perception of speech.

Already mid-way through their first year, (Kuhl et al., 1992) infants' behavior is beginning to be influenced by speech experience. Older infants discriminate acoustic differences between native sounds even more effectively than early in development (Kuhl et al., 2006), but no longer very accurately distinguish many non-native sounds (see Werker and Tees, 1999). Decreases in speech discrimination are most evident among non-native speech contrasts similar to those of the native language (Best, 1995; Flege, 1995) whereas very

dissimilar non-native contrasts (like clicks in Zulu for English listeners, Best et al., 1988) continue to be discriminated in adulthood.

The experience-dependent change in non-native speech perception is thought to reflect the influence of native-language speech category learning and has been described as a “warping” of perceptual space (see Kuhl et al., 2008). Imagining perceptual space as a multidimensional topography, the landscape appears to be relatively flat in early infancy with any discontinuities arising from general auditory processing. At this point, the mapping from acoustics to the relatively flat perceptual space is closely related to the raw acoustic differences among speech sounds and infants’ speech discrimination is mostly independent of the native language environment. Speech category learning warps perceptual space in ways that reflect regularities of the native speech input and infants begin to perceive speech relative to native-language categories rather than solely according to psychoacoustic differences. Perceptual space expands and shrinks as a function of the diagnosticity of dimensions in contributing to categorization (Nosofsky, 1986; Francis and Nusbaum, 2002) and there are regions of increased within-category similarity contrasted with regions of reduced between-category similarity (Lieberman et al., 1957; Iverson et al. 2003) that serve to exaggerate between-category differences and shrink within-category differences.¹ In terms of the topographic metaphor, categories can be envisioned as basins, or attractors (Spivey, 2007), in perceptual space in which there is limited perceptual discriminability flanked by peaks between category basins representing regions of exaggerated perceptual discriminability.

This perceptual warping aligns with native language regularities, promoting efficiency in native speech categorization. However, once the perceptual system has committed to a native-

language parse of the perceptual space, it can be quite difficult for adults to learn categories of a second language that do not align with native-language regularities (Best, 1995; Flege, 1995). For example, learning Italian requires learning a single category /i/ in a region of perceptual space that accommodates both /i/ and /I/ in English. Perceptual warping for Italian minimizes within-/i/-category differences, promoting efficient Italian speech categorization. However, the shrinking of perceptual discriminability within this Italian category comes at some detriment to later learning English (which requires perceptual discrimination between /i/ and /I/ in the very same region of perceptual space; Flege et al., 1999). Of note, sounds (like Zulu clicks for English listeners) that fall in regions of perceptual space not inhabited by native sounds are relatively spared interactions with native categories and continue to be well-distinguished, presumably as a result of their acoustic differences (see **Smiljanic** this volume for further discussion).

It is remarkable that infants must begin to form speech categories without an indication of how many categories exist in the native language and without significant exposure to these sounds in isolation (see Vallabha et al., 2007). Nonetheless, the dual tasks of speech sound discrimination and generalization appear to be well underway well before infants speak the first word or develop a significant lexicon (Jusczyk, 1997 for review). Although the groundwork for speech category learning begins in infancy, there is a lengthy developmental course of speech category refinement whereby even 12-year-olds have not reached adult levels of speech categorization for some native sounds (Hazan and Barrett, 2000). Lexical development, learning to read, and continued development of perceptual expertise with speech are all likely contributors along this protracted developmental course (see **Munson et al.**, **Maye** this volume).

Despite an appreciation for the profound influence of categorization on speech processing, we do not yet have a complete understanding of how speech categories are learned in infancy or adulthood. At least part of the reason for this is because it is not feasible to entirely control and manipulate speech to observe the consequences of different patterns of experience. Natural cross-language comparisons for adults and infants, like those described above, are the standard upon which the majority of our understanding is based and they have provided an understanding of the range of behaviors to be accounted for by any theory. But, without controlled manipulation of experience models of speech category learning are, by necessity, more descriptive than predictive.

3 Converging Methods

Given the difficulty in manipulating and controlling speech experience, it is useful to take a converging methods approach, investigating auditory and cognitive constraints on speech category learning from multiple, coordinated perspectives that vary in the extent to which they emphasize ecological validity, or naturalness, of experience versus experimental control of experience. Ideally, the coordinated approaches converge such that it is possible to develop predictions from tightly controlled laboratory experiments that may be tested in more natural speech communication. The sections below describe representative research findings that exemplify multiple converging methods for investigating speech category learning.

3.1 Cross-language comparisons

Cross-language comparisons provide a natural experiment in differing histories of speech experience (see **Smiljanic**, **Iverson**, and **Escudero** this volume). Clever investigations of naturally-occurring differences in listeners' histories of speech experience

have defined questions of significance for understanding speech category learning. However, a drawback is that experimental control of experience is limited making it difficult to infer learning mechanisms.

The review above highlighted some of what is known about infant speech category learning from this approach. Among adult learners, plasticity for learning non-native speech categories appears to be maintained, although its expression is critically dependent on the amount and quality of the second-language input and its interaction with the speech categories learned for the first language (Flege, 1995). Flege and MacKay (2004), for example, report that native speakers' ability to discriminate non-native vowels is best predicted by self-estimated amount of first-language usage, with lower usage predicting better second-language performance. In fact, non-native perception among adults that arrived in the second-language environment earlier and used their first language less often was not statistically distinguishable from native listeners. Flege suggests that the learning mechanisms that guide first-language speech category learning remain intact through adulthood but that first and second language processing share common resources and therefore mutually influence one another. Non-native speech categories are perceived through the lens of the perceptual space warped by learning native categories. On the whole, naturalistic cross-language studies indicate plasticity in adult non-native speech category learning but, as is the case for infant learning, the mechanistic details of this learning remain unclear.

3.2 Laboratory-based speech category training studies

Wrestling with the issue of control over experience, some studies have taken the approach of manipulating short-term speech experience in the laboratory. Artificial "languages" comprised

of speech tokens manipulated to have special characteristics have been used widely as a tool in understanding infant language acquisition (e.g., Saffran et al., 1996; Thiessen, 2007), including speech category learning (Maye et al., 2002). In these studies, listeners hear speech possessing particular well-controlled regularities and perception is measured thereafter to ascertain the influence of experience with these short-term regularities. Using this method, Maye et al. exposed infants to each exemplar of a speech stimulus series, with some exemplars presented more often than others. Some infants heard sounds sampled with a unimodal distribution whereas other infants heard the same stimuli sampled with a bimodal distribution. Subsequent speech discrimination accuracy was exaggerated for the bimodal group, suggesting a categorization-like warping of perceptual space as a result of short-term exposure to different distributions of speech.

A limitation of these approaches, to date, is that they leave open important issues about how category learning proceeds with more natural speech. Whereas infants experience a rather continuous stream of speech, laboratory-based experiments typically solve the segmentation problem for the listeners by presenting isolated instances (e.g., Maye et al., 2002). Although input regularities can guide segmentation (Saffran et al., 1996), the extent to which distributional regularities support speech category learning in unsegmented speech remains an open question (see Pierrehumbert, 2003). In addition, artificial languages tend to be rather simple with a single acoustic dimension (or just a few dimensions) defining categories. In natural speech, infants must contend with highly multidimensional speech acoustics. It will be important for future research to determine the extent to which distributional learning scales to more natural speech category learning challenges.

Kuhl and colleagues (2003) have taken a step in this direction by exposing 9-month-old English-learning infants to Mandarin Chinese across 12 play sessions with a Mandarin-speaking experimenter. This exposure was sufficient to reverse the decline in Mandarin speech discrimination observed among infants exposed instead to English-language play sessions. Perhaps telling of the mechanisms involved in infant speech category learning, the preservation of Mandarin speech discrimination was observed only among infants experiencing Mandarin with a real adult and not among infants exposed to the same speech via audiovisual or audio recordings. Thus, mere exposure to distributional regularities may not be enough to direct learning in more natural circumstances. It seems likely that a combination of factors, including distributional regularity in speech input (Holt et al., 1998) and the potential for socially-driven feedback (see Goldstein and Schwade, in press; 2008) influence early speech category learning, but details of these mechanisms remain to be discovered.

Laboratory-based speech training among adults learning a second language also informs our understanding of speech category learning (e.g., Jamieson and Morosan, 1989; Logan et al., 1991; Pisoni et al., 1994; Bradlow et al., 1999; McCandliss et al., 2002; Iverson et al., 2005; Goudbeek et al., 2008). Some early attempts to train adults on non-native categories included discrimination training with little acoustic variance in the speech training set. Although listeners learned to discriminate among training stimuli, they typically could not transfer this learning to natural speech or to different contexts (Strange and Dittmann, 1984). Recent research has underscored the importance of acoustic variability. Including multiple speakers and phonetic contexts in training seems to aid learning and generalization (Jamieson and Morosan, 1989; Lively et al., 1993; Bradlow et al., 1999; McCandliss et al., 2002; Iverson et al., 2005). In such

studies, participants tend to improve in their ability to reliably categorize non-native speech over the course of training with learning persisting across months and generalizing to speech production in some studies (Bradlow et al., 1999). However, extensive training is necessary to evidence learning and the final level of achievement typically has not been equal to that of native listeners (Bradlow et al., 1999; Lively et al., 1993; Logan et al., 1991). Thus, training studies provide evidence of plasticity in the adult system to support category learning, although the system is clearly not as flexible as in infancy.

Such studies also have begun to make mechanistic predictions about adult learning. For example, McCandliss et al. (2002) hypothesized that the perceptual warping apparent in native speech category learning produces neural circuits committed to processing native-language speech categories. Hearing similar non-native sounds activates these native circuits, thereby further reinforcing them. Thus, somewhat counterintuitively, training listeners with non-native sounds may reinforce existing *native* categories because perceptually-similar non-native sounds activate neural circuits supporting native categories. By this logic, McCandliss et al. (2002) predicted that beginning training with highly exaggerated instances of non-native speech falling outside native perceptual space and then incrementally adjusting training stimuli to be more representative instances of the non-native categories may facilitate learning. Their results support this prediction, but also indicate an additional role for explicit feedback in learning (Tricomi et al., 2006) suggesting a more complex set of learning mechanisms (see also Goudbeek et al., 2005; 2008 for discussion of the role of explicit feedback).

Many studies, including most of those cited above, have investigated Japanese adults learning English /r/ versus /l/, an adult speech category learning problem that is notoriously

challenging. Other speech categories appear to be more easily learned by non-native listeners (Pisoni et al., 1982; Jamieson and Morosan 1986; Polka, 1992) and this may be predicted by the relationship between first and second language categories and their interaction (Best, 1995; Flege, 1995). For example, it may be difficult for native Japanese adults to learn English /r/-/l/ because Japanese possesses only a single category in this region of phonetic space. The warping of perceptual space in learning Japanese causes heightened within-category similarity for the Japanese category in nearly the same region of phonetic space that must be differentiated to distinguish the two categories /r/ and /l/ in English. Adults learning new speech categories must contend with a perceptual space already committed to a native language and perceptually warped according to its regularities (Flege, 1995).

Even among more easily-learned categories, there are enormous individual differences in adult speech category learning (Golestani and Zatorre, in press), making it difficult to draw sweeping conclusions about the degree of plasticity in adult speech category learning. Although it is not yet the norm for studies to investigate individual differences in detail, it seems likely that research can capitalize on individual differences to understand more about the auditory and cognitive constraints on speech category learning (e.g., Slevc and Miyake, 2006).

3.3 Laboratory-based nonspeech category learning studies

One way to gain experimental control over listeners' histories of experience is to create novel sound stimuli with which listeners have no experience and for which listeners possess no *a priori* categories. The major benefit of training listeners to categorize such artificial nonspeech sounds is that it is possible to exert control over and have knowledge of listeners' entire history of experience with the sounds, thus providing the opportunity to investigate explicitly the

general perceptual and cognitive constraints on auditory processing that might influence speech categorization.

The literature in this area is not yet large, but already insights have been gained about auditory category learning relevant to speech categorization. It has long been observed that vowels and consonants exhibit different patterns of categorization and discrimination, with vowels tending to be perceived more continuously, with less abrupt categorization boundaries and less sharp discrimination peaks than stop consonants (Repp, 1984; Schouten and Van Hessen, 1992). Mirman et al. (2004) examined whether general auditory constraints on processing the differing spectrotemporal acoustics of vowels and consonants might play a role in this pattern by training listeners to categorize nonspeech sounds modeling rapidly-changing acoustic dimensions of consonants or steady-state acoustic dimensions of simplified vowels. Patterns of nonspeech discrimination and categorization mirrored those of the speech stimuli they modeled. General characteristics of auditory sensory memory may play a role: more quickly-decaying perceptual memory traces for rapidly-changing sounds relative to steady-state sounds could account for this pattern for both speech and nonspeech.

Many accounts have suggested that infants' initial parse of the perceptual space may rely upon natural "boundaries" in auditory processing that arise from discontinuities in the mapping from acoustics to audition (Pisoni, 1977). The most compelling case is a proposed discontinuity in auditory temporal processing that may influence perception of voicing (Holt et al., 2004 for review). Examining the question of how discontinuities would interact with experience, Holt et al. (2004) trained listeners on nonspeech sounds that varied along this perceptually discontinuous acoustic dimension. When the sound input distribution boundary

aligned with the perceptual discontinuity, learning was facilitated relative to when listeners were forced to categorize across the perceptual discontinuity by an input distribution requiring listeners to treat stimuli on either side of the discontinuity as members of the same category. However, listeners did eventually learn in this latter situation. Thus, basic auditory constraints on perceptual processing may provide an initial parse facilitating learning, but learning is flexible enough to overcome perceptual biases.

Nonspeech category learning studies also highlight how task influences category learning. Discrimination training (explicit comparison of stimuli) and categorization training (responding to acoustically-variable exemplars as category members) warp listeners' perception of nonspeech stimuli, but in different ways. Discrimination training increases listeners' sensitivity to small distinctions among stimuli thereby working against categorization, which requires that discriminably different acoustic exemplars be treated as functionally equivalent (Guenther et al., 1999). This insight from nonspeech learning is important because it is common in studies of speech perception to use discrimination tasks as indices of categorization, taking heightened discrimination between pairs as an indication of a category boundary. Guenther et al.'s nonspeech auditory training study indicates that discrimination training is not equivalent to category learning and it has implications for interpreting the fact that Japanese listeners trained to discriminate English /r-/l/ do not generalize well to natural speech categories (Strange and Ditman, 1984).

Another characteristic of speech categories is their multidimensionality; typically, numerous acoustic dimensions covary with speech categories. The dimensions defining perceptual category space are not equivalent and some acoustic dimensions play a greater role

in determining the category of a sound than do others. As described above, the perceptual space expands and shrinks as a function of the diagnosticity of a dimension in contributing to categorization (Nosofsky, 1986) and this is reflected in the perceptual weight listeners give to different acoustic dimensions in their categorization responses. As an example, both formant frequency and vowel duration co-vary with English /i/ and /I/, but native listeners rely much more on formant frequency than vowel duration (Hillenbrand et al., 2000).

Relative perceptual cue weight changes across development (Nittrouer, 2004) and is native-language specific. Whereas native English listeners rely primarily on formant frequency for /i/-/I/, non-native listeners often rely more on duration (Flege et al., 1997). In fact, differences in perceptual cue weighting appear to account for some of the difficulties among adults learning non-native speech categories. Native English listeners rely primarily upon third formant (F3) onset frequency in categorizing English /r/ and /l/ (the dimension that is most diagnostic of /r/ vs. /l/ in English speech productions; Lotto et al., 2004), but native Japanese listeners categorize English /r/ and /l/ using the less diagnostic second formant (F2) onset frequency. Since F2 is less diagnostic of English /r/-/l/ speech productions, this perceptual weighting results in less-than-native categorization (Iverson et al., 2003). These findings and others make clear the importance of learning in perceptual cue weighting for speech categories, but exactly how perceptual weighting relates to details of speech experience remains unclear.

The control over experience afforded by nonspeech categories allowed Holt and Lotto (2006) to test the kinds of input distributions that affect perceptual cue weighting in category learning. They trained adults with explicit feedback to learn to categorize two novel nonspeech

categories drawn from a two-dimensional acoustic space defined by the rate at which sine wave tones repeatedly increased and decreased in frequency (Modulation Frequency, MF) around a particular base frequency (Center Frequency, CF). Although the dimensions were psychoacoustically equated and the stimuli defining the categories were sampled such that each acoustic dimension was equally informative to the categorization task, listeners relied much more upon CF than MF in their categorization responses. This bias allowed Holt and Lotto to investigate how different stimulus training sets influence perceptual cue weighting. Moving the distributions closer along the preferred, CF, dimension thereby making CF a less reliable categorization cue had no effect. However, making CF more variable within each category distribution caused listeners to rely on MF instead of CF in categorization responses. It appears that the variance experienced across an acoustic dimension is significant to perceptual cue weighting. An implication of this finding from nonspeech is that use of an inefficient acoustic dimension in non-native speech categorization (e.g., F2 in English /r/-/l/) may be lessened by experience with substantial variability along this dimension. Experiencing multiple talkers produce speech categories appears to facilitate non-native speech training (Lively et al., 1993), which may provide more variability along less diagnostic acoustic dimensions and serve to modify listeners' perceptual cue weights for the non-native speech categories.

An issue in the above studies is the role of feedback. Laboratory-based nonspeech category training tends to rely on explicit feedback atypical of natural speech experience, which does not seem to involve explicit category labels, or explicit feedback (e.g., Jusczyk, 1997). Goudbeek and colleagues (2005; 2008) have used nonspeech categories to investigate the role of feedback in auditory category learning, reporting that without explicit feedback listeners find

it very difficult to categories defined by multiple acoustic dimensions. This is curious considering that highly multidimensional speech categories appear to be learned by infants without explicit feedback. Whereas social signals may be considered to be a form of feedback guiding learning (Kuhl et al., 2003; Goldstein and Schwade, 2008), it is not the sort of explicit feedback that appeared to be necessary in this study.

It seems likely that speech category instances during first language acquisition involve complex relationships among acoustic speech and various simultaneous events in the environment. Of course, these naturally complex interactions are difficult to control, making it challenging to infer mechanism. Video games are an immersive environment in which researchers can maintain control over auditory experience while manipulating complex multimodal relationships that the sounds have with other perceptual events, thus involving participants in the functional use of sound without explicitly training them in auditory categorization or giving them explicit feedback for category learning.

Wade and Holt (2005) developed a space-invaders-style videogame with visual creatures, each associated with a category of sounds. The sounds were designed to model some of the multidimensional complexity of speech categories, without sounding like speech. To succeed in the game, participants had to learn the relationship between each creature and the corresponding sound category, although this was never made explicit. Similar to the process of learning to treat acoustically distinct speech signals as members of the same speech category, listeners gradually learned that perceptually discriminable creatures' sounds were functionally equivalent in the game. There was no explicit feedback and participants were not aware of the category learning task, but sounds served a function. After 30 minutes of game

play, listeners' responses indicated significant category learning and generalization to novel sounds. Though there was no explicit feedback, participants were able to learn the complex auditory categories incidentally suggesting that functional use of sound and multimodal relationships between sound and other perceptual dimensions may be significant in complex, multidimensional category learning. Of interest to understanding how sound categories are represented by the brain (**Nguyen** this volume), neuroimaging methods reveal that learning to categorize nonspeech sounds in this way recruits brain regions typically associated with speech processing (Leech, Holt, Devlin and Dick, 2009) and warps the perceptual space in a manner similar to that observed among infants learning native-language speech categories (Liu and Holt, in press).

To date, there are relatively few nonspeech auditory category learning studies that address the category learning challenges most relevant to speech category learning. Assumptions that distributional learning drives speech category learning abound, but it is not yet well understood to what distribution statistics listeners are sensitive, how feedback of various forms may influence speech category learning, how acoustic dimensions are perceptually weighted and how task affects the warping of perceptual space. The opportunity to carefully manipulate experience with nonspeech categories provides an opportunity to investigate these issues in greater depth to discover constraints on auditory learning relevant to speech categorization.

3.4 Non-human animal speech category training studies

Speech category training studies with non-human animals offer some of the same benefits of experimental control over experience present for nonspeech learning studies with humans.

Animals as diverse as birds, macaques, and chinchillas can discriminate speech (Dewson, 1964; Burdick and Miller, 1975; Kuhl and Miller, 1975; Morse and Snowdon, 1975; Dooling and Brown, 1990) and there is a long history of using animals to probe speech perception absent speech experience. These studies have defined general auditory constraints on speech perception (Kuhl and Miller, 1975, 1978; Kluender and Lotto, 1994; Dooling, 1995; Lotto et al., 1997; Sinnott et al., 1998). For example, Lotto and colleagues (1997) found that Japanese quail trained to peck in response to /ga/ peck more heavily to a perceptually ambiguous sound between /ga/ and /da/ when it is preceded by /a/. This context-dependent response pattern exactly mirrors the pattern of context-dependent speech categorization among human listeners that is taken as evidence of perceptual compensation for coarticulation (Mann, 1980). The existence of parallel perceptual responses to speech in humans and nonhumans suggests general auditory constraints on speech processing may contribute to speech perception challenges like compensation for coarticulation (Lotto et al., 1997), trading relations (Kluender and Lotto, 1994), categorical perception (Kuhl and Miller, 1975), and discrimination of prosodic qualities of speech (Ramus et al., 2000).

Animal models also allow controlled investigation of the effects of experience on speech processing. For example, Kuhl (1991) reported that monkeys do not show the patterns of graded internal vowel category responses indicative of perceptual warping that are observed for human adults and infants (Grieser and Kuhl, 1989), perhaps indicating species-specificity in this aspect of speech categorization. However, the monkeys had no experience with speech. When Kluender and colleagues (1998) provided birds experience with vowel input distributions, birds' subsequent responses were graded and highly correlated with human listeners' graded

categorization responses to the same sounds. Experience with the distributional characteristics of speech categories is essential in producing graded responses to speech indicative of perceptual warping, but without the possibility of control over experience afforded by animal models this hypothesis would not have been testable.

Control over animals' speech experience allowed Holt, Lotto, and Kluender (2001) to determine that the relationship of fundamental frequency (F0) and voicing (with higher F0 associated with voiceless categories in English and other languages, see Kingston and Diehl, 1994) is not an obligatory influence of F0 on voicing arising from perceptual constraints, but rather is more likely due to the learnability of covariation between these acoustic dimensions. It arises only when animals experience correlation between the acoustic dimensions during training. Kluender et al. (1987) found that Japanese quail learn the complex mapping among multiple acoustic dimensions defining English alveolar stop consonants and generalize to speech never heard in training. This category learning was impressive because there were no invariant acoustic cues among the stimuli that could define category membership. Thus, the high multidimensionality of speech categories can be accommodated by rather simple learning processes such as those available to quail.

The issue of how feedback influences speech category learning is important in interpreting evidence from animal studies because most methods rely on explicit feedback in training animals to respond to speech. However, even with animal training paradigms that require explicit feedback, it is possible to learn about characteristics of unsupervised learning. In the Kluender et al. (1998) study mentioned above, birds responded to tokens from one of two vowel categories. All vowels were equivalent in training in that response to each vowel

elicited the same feedback. Nonetheless, birds' responses mirrored distributional characteristics of the vowel input distributions such that the birds responded to some vowel exemplars more robustly than others. This aspect of animal learning cannot arise from the feedback and appears to reflect something general about distributional learning.

In sum, studying animal learning can serve as a means of understanding how general auditory capacities and general learning mechanisms may solve some of the challenges of speech category learning. Prototype effects (Kuhl, 1991; Kluender et al., 1998), lack of acoustic invariance and multidimensional learning (Kluender et al., 1987), perceptual warping by categorization (Kluender et al., 1998), perceptual segmentation (Hauser et al., 2001), and the effects of correlation among acoustic dimensions (Holt et al., 2001) are characteristics of speech category learning that have been illuminated by animal learning models.

4 Conclusion

Experience alters speech processing, but by what means? There are important unresolved questions in speech category learning, ripe for research. Accounts ultimately must explain how experience alters the perceptual space among infants learning speech categories and, in doing so, shapes the learning challenges encountered by adults learning non-native speech categories. There is a need to better define distributional learning and to delineate its mechanisms, including the role of feedback. We must understand exactly what it means to “warp” a perceptual space and discover the representations that inhabit the space. Moreover, we must interpret individual differences, where they exist, and attend to the role higher-level cognitive constraints like attention, working memory and decisional processes in guiding first- and second-language speech category learning.

Although there is much work ahead to understand speech categorization, much has been learned in recent years. This brief overview above highlights representative research approaches that extend a bit beyond the traditional borders of laboratory phonology to bring converging methods to bear on the question of how speech categories are learned. These approaches share the aim of understanding the mechanisms of speech category learning by investigating the cognitive and perceptual constraints listeners bring to the task. The research that exists indicates the promise of a converging methods approach in gaining control over experience to move our models from descriptive to predictive.

5 References

- Best, C. T. (1995). 'A direct realist perspective on cross-language speech perception'. In W. Strange (Ed.) *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*, pp. 167-200. Timonium MD: York Press.
- Best, C. T., McRoberts, G. W., and Sithole, N. M. (1988). 'Examination of the perceptual reorganization for speech contrasts: Zulu click discrimination by English-speaking adults and infants', *Journal of Experimental Psychology: Human Perception and Performance*, 14: 345-360.
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B. and Tohkura, Y. (1999). 'Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production', *Perception and Psychophysics*, 61: 977-985.

- Burdick, C. K., and Miller, J. D. (1975). 'Speech perception by the chinchilla: Discrimination of sustained /a/ and /i/'. *Journal of the Acoustical Society of America*, 58: 961-970.
- Dewson, J. H. (1964). 'Speech sound discrimination by cats', *Science*, 144: 555-556.
- Dooling, R. J., and Brown, S. D. (1990). 'Speech perception by budgerigars (*Melopsittacus undulatus*): Spoken vowels', *Perception and Psychophysics*, 47: 568-574.
- Dooling, R. J., Best, C. T., and Brown, S. D. (1995). 'Discrimination of synthetic full-formant and sinewave /ra-la/ continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*)', *Journal of the Acoustical Society of America*, 97: 1839-1846.
- Flege, J. E. (1995). 'Second-language speech learning: Theory, findings, and problems'. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonim, MD: York Press.
- Flege, J. E., Bohn, O. -S. and Jang, S. (1997). 'Effects of experience on nonnative speakers' production and perception of English vowels', *Journal of Phonetics*, 25: 437-470.
- Flege, J. and MacKay, I. (2004). 'Perceiving vowels in a second language', *Studies in Second Language Acquisition*, 26: 1-34.
- Flege, J. E., MacKay, I. R. A., and Meador, D. (1999). 'Native Italian speakers' production and perception of English vowels', *Journal of the Acoustical Society of America*, 106: 2973-2987.
- Francis, A. L., and Nusbaum, H.C., (2002). 'Selective attention and the acquisition of new phonetic categories', *Journal of Experimental Psychology: Human Perception and Performance*, 28: 349- 366.

- Goldstein, M. H., and Schwade, J. A. (in press). 'From birds to words: Perception of structure in social interactions guides vocal development and language learning'. In M. S. Blumberg, J. H. Freeman, and S.R. Robinson (Eds.), *The Oxford Handbook of Developmental and Comparative Neuroscience*. Oxford University Press.
- Goldstein, M. H., and Schwade, J. A. (2008). 'Social feedback to infants' babbling facilitates rapid phonological learning', *Psychological Science*, 19: 515-522.
- Golestani, N. and Zatorre, R. J. (in press). 'Individual differences in the acquisition of second language phonology. *Brain and Language*. doi:10.1016/j.bandl.2008.01.005.
- Goudbeek, M., Smits, R., Swingley, D., and Cutler, A. (2005). 'Acquiring auditory and phonetic categories'. In H. Cohen, and C. Lefebvre (Eds.), *Categorization in Cognitive Science*, pp. 497-513. Amsterdam: Elsevier.
- Goudbeek, M., Cutler, A., and Smits, R. (2008). 'Supervised and unsupervised learning of multidimensionally varying non-native speech categories', *Speech Communication*, 50: 109-125.
- Grieser, D., and Kuhl, P. K. (1989). 'Categorization of speech by infants: Support for speech-sound prototypes', *Developmental Psychology*, 25: 577-588.
- Guenther, F. H., Husain, F. T., Cohen, M. A., and Shinn-Cunningham, B. G. (1999). 'Effects of categorization and discrimination training on auditory perceptual space', *Journal of the Acoustical Society of America*, 106: 2900-2912.
- Harnad, S. R. (1990). *Categorical perception: The groundwork of cognition*. Cambridge University Press.

- Hauser, M.D., Newport, E.L., and Aslin, R.N. (2001). 'Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton top tamarins', *Cognition*, 78: B53-B64.
- Hazan, V., and Barrett, S. (2000). 'The development of phonemic categorization in children aged 6 to 12', *Journal of Phonetics*, 28: 377-396.
- Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. Lawrence Erlbaum Associates.
- Hillenbrand, J. M., Clark, M. J., and Houde, R. A. (2000). 'Some effects of duration on vowel recognition', *Journal of the Acoustical Society of America*, 108: 3013–3022.
- Holt, L. L. and Lotto, A. J. (2006). 'Cue weighting in auditory categorization: Implications for first and second language acquisition', *Journal of the Acoustical Society of America*, 119: 3059-3071.
- Holt, L.L., Lotto, A.J., and Diehl, R.L. (2004). 'Auditory discontinuities interact with categorization: Implications for speech perception', *Journal of the Acoustical Society of America*, 116: 1763-1773.
- Holt, L. L., Lotto, A. J., and Kluender, K. R. (1998). 'Incorporating principles of general learning in theories of language acquisition'. In M. Gruber, C. Derrick Higgins, K. S. Olson and T. Wysocki (Eds.), *Chicago Linguistic Society, Volume 34: The Panels*. Chicago: Chicago Linguistic Society, 253-268.
- Holt, L. L., Lotto, A. J., and Kluender, K. R. (2001). 'Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?' *Journal of the Acoustical Society of America*, 109: 764-774.

- Iverson, P., Hazan, V., and Bannister, K. (2005). 'Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults', *Journal of the Acoustical Society of America*, 118: 3267-3278.
- Iverson, P., and Kuhl, P. (1995). 'Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling', *Journal of the Acoustical Society of America*, 97: 553-562.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A. and Siebert, C. (2003). 'A perceptual interference account of acquisition difficulties for non-native phonemes', *Cognition*, 87: B47-57.
- Jamieson, D. G., and Morosan, D. E. (1989). 'Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques', *Canadian Journal of Psychology*, 43: 88-96.
- Johnson, K., Flemming, E. and Wright, R. (1993). 'The hyperspace effect: Phonetic targets are hyperarticulated', *Language*, 69: 505-528.
- Jusczyk, P. W. (1997). *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.
- Kingston, J., and Diehl, R. L. (1994). 'Phonetic knowledge'. *Language*, 70: 419-454.
- Kluender, K. R., Diehl, R. L., and Killeen, P. R. (1987). 'Japanese quail can learn phonetic categories', *Science*, 237: 1195-1197.
- Kluender, K. R., and Lotto, A. J. (1994). 'Effects of first formant onset frequency on [-voice] judgments result from general auditory processes not specific to humans', *Journal of the Acoustical Society of America*, 95: 1044-1052.

- Kluender, K. R., Lotto, A. J., and Holt, L. L. (2005). 'Contributions of nonhuman animal models to understanding human speech perception'. In S. Greenberg and W. Ainsworth (Eds.) *Listening to Speech: An Auditory Perspective*, Oxford University Press: New York, NY.
- Kluender, K.R., and Lotto, A.J. (1994). 'Effects of first formant onset frequency on [-voice] judgments result from general auditory processes not specific to humans', *Journal of the Acoustical Society of America*, 95: 1044-1052.
- Kluender, K. R., Lotto, A. J., Holt, L. L., and Bloedel, S. B. (1998). 'Role of experience for language-specific functional mappings for vowel sounds', *Journal of the Acoustical Society of America*, 104: 3568-3582.
- Kuhl, P. K. (1991). 'Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not', *Perception and Psychophysics*, 50: 93-107.
- Kuhl, P. K. (1993). 'Innate predispositions and the effects of experience in speech perception: The native language magnet theory', In B. deBoysseon-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, and J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259-274). Dordrecht, Netherlands: Kluwer Academic Publishers.
- Kuhl, P.K. (2000a). 'A new view of language acquisition', *Proceedings of the National Academy of Science*, 97: 11850-11857.
- Kuhl, P. K. (2000b). 'Language, mind, and brain: Experience alters perception', In M. S. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (2nd ed.) (pp. 99-115). Cambridge, MA: MIT Press.

- Kuhl, P.K., and Miller, J.D. (1975). 'Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants', *Science*, 190: 69–72.
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). 'Native Language Magnet Theory Expanded (NLM-e)'. *Philosophical Transactions of the Royal Society B*, 363: 979-1000.
- Kuhl, P. K., and Miller, J. D. (1975). 'Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants', *Science*, 190: 69-72.
- Kuhl, P. K., and Miller, J. D. (1978). 'Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli', *Journal of the Acoustical Society of America*, 63: 905-917.
- Kuhl, P K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S. and Iverson, P. (2006). 'Infants show a facilitation effect for native language phonetic perception between 6 and 12 months', *Developmental Science*, 9: F13-F21.
- Kuhl, P. K., Tsao, F.-M., and Liu, H.-M. (2003). 'Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning', *Proceedings of the National Academy of Sciences*, 100: 9096-9101.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). 'Linguistic experience alters phonetic perception in infants by 6 months of age', *Science*, 255: 606-608.
- Leech, R, Holt, L. L., Devlin, J. T. and Dick, F. (2009). 'Expertise with nonspeech sounds recruits speech-sensitive cortical regions', *Journal of Neuroscience*, 29: 5234-52389.

- Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). 'The discrimination of speech sounds within and across phoneme boundaries', *Journal of Experimental Psychology*, 54: 358-368.
- Lively, S. E., Logan, J. S., and Pisoni, D. B. (1993). 'Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories', *Journal of the Acoustical Society of America*, 94: 1242-1255.
- Liu, R. and Holt, L. L. (in press). 'Changes in mismatch negativity reflecting the acquisition of complex, nonspeech auditory categories', *Journal of Cognitive Neuroscience*.
- Logan, J., Lively, S. and Pisoni, D. (1991). 'Training Japanese listeners to identify English /r/ and /l/: a first report', *Journal of the Acoustical Society of America*, 89: 874-886.
- Lotto, A.J. (2000). 'Language acquisition as complex category formation', *Phonetica*, 57: 189-196.
- Lotto, A. J., Kluender, K. R., and Holt, L. L. (1997). 'Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*)'. *Journal of the Acoustical Society of America*, 102: 1134-1140.
- Lotto, A. J., Kluender, K. R., and Holt, L. L. (1998). 'Depolarizing the perceptual magnet effect', *The Journal of the Acoustical Society of America*, 103: 3648-3655.
- Lotto, A.J., Sato, M., and Diehl, R.L. (2004). 'Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/', J. Slifka, S. Manuel, and M. Matthies (Eds.) *From Sound to Sense: 50+ Years of Discoveries in Speech Communication*.
- Mann, V. A. (1980). 'Influence of preceding liquid on stop-consonant perception', *Perception and Psychophysics*, 28: 407-412.

- Maye, J., Werker, J. F., and Gerken, L. (2002). 'Infant sensitivity to distributional information can affect phonetic discrimination', *Cognition*, 82: B101-B111.
- McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., and McClelland, J. L. (2002). 'Success and failure in teaching the r-l contrast to Japanese adults: predictions of a hebbian model of plasticity and stabilization in spoken language perception'. *Cognitive, Affective, and Behavioral Neuroscience*, 2: 89-108.
- McMurray, B. and Aslin, R.N. (2005). 'Infants are sensitive to within-category variation in speech perception', *Cognition*, 95: B15-B26.
- McMurray, B., Aslin, R., Tanenhaus, M., Spivey, M., and Subik, D. (2008). 'Gradient sensitivity to within-category variation in speech: Implications for categorical perception', *Journal of Experimental Psychology: Human Perception and Performance*, 34: 1609-1631.
- Miller, J. L., and Volaitis, L. E. (1989). 'Effect of speaking rate on the perceptual structure of a phonetic category', *Perception and Psychophysics*, 46: 505-512.
- Mirman, D., Holt, L.L., and McClelland, J.L. (2004). 'Categorization and discrimination of nonspeech sounds: differences between steady-state and rapidly-changing acoustic cues', *Journal of the Acoustical Society of America*, 116: 1198-1207.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. L., Jenkins, J. J. & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics*, 18: 331-340.
- Morse, P. A., and Snowdon, C. T. (1975). 'An investigation of categorical speech discrimination by rhesus monkeys', *Perception and Psychophysics*, 17: 9-16.

- Nittrouer S. (2004). 'The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults', *Journal of the Acoustical Society of America*, 115: 1777-1790.
- Nosofsky, R. M. (1986). 'Attention, similarity, and the identification-categorization relationship', *Journal of Experimental Psychology: General*, 115: 39-57.
- Pierrehumbert, J. (2003). 'Phonetic diversity, statistical learning, and acquisition of phonology', *Language and Speech*, 46: 115-154.
- Pisoni, D. B. (1977). 'Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops', *Journal of the Acoustical Society of America*, 61: 1352-1361.
- Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (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.
- Pisoni, D.B, Lively, S. E., and Logan, J. S. (1994). 'Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception', In: Goodman, Judith C.; Ed; Nusbaum, Howard C.; Ed; *The development of speech perception: The transition from speech sounds to spoken words*, pp. 121-166; Cambridge, MA, US : The MIT Press.
- Polka, L. (1992). 'Characterizing the influence of native experience on adult speech perception', *Perception and Psychophysics*, 52: 37-52.
- Ramus, F., Hauser, M. D., Miller, C., Morris, D., and Mehler, J. (2000). 'Language discrimination by human newborns and by cotton-top tamarin monkeys', *Science*, 288: 349-351.

- Repp, B. H. (1984). 'Categorical perception: Issues, methods, findings', *Speech and Language*, 10: 243–335.
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). 'Statistical learning by 8-month-old infants', *Science*, 274: 1926-1928.
- Schouten, M. E., and Van Hessen, A. J. (1992). 'Modeling phoneme perception: Categorical perception', *Journal of the Acoustical Society of America*, 92: 1841–1855.
- Slevc, L.R. and Miyake, A. (2006). 'Individual differences in second language proficiency: Does musical ability matter?' *Psychological Science*, 17: 675-681.
- Sinnott, J. M., Brown, C. H., and Borneman, M. A. (1998). 'Effects of syllable duration on stop-glide identification in syllable-initial and syllable-final position by humans and monkeys', *Perception and Psychophysics*, 60: 1032-1043.
- Spivey, M. J. (2007). *The continuity of mind*. New York: Oxford University Press.
- Strange, W., and Dittmann, S. (1984). 'Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English', *Perception and Psychophysics*, 36: 131-145.
- Strange, W., and Jenkins, J. J. (1978). 'Role of linguistic experience in the perception of speech'. In R. D. Walk, and H. L. Pick (Eds.), *Perception and Experience*, pp. 125–69. New York: Plenum.
- Thiessen, E. D. (2007). 'The effect of distributional information on children's use of phonemic contrasts', *Journal of Memory and Language*, 56: 16-34.
- Tricomi, E., Delgado, M.R., McCandliss, B.D., McClelland, J.L., and Fiez, J.A. (2006). 'Performance feedback drives caudate activation in a phonological learning task', *Journal of Cognitive Neuroscience*, 18: 1029-1043.

- Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. and Amano, S. (2007). 'Unsupervised learning of vowel categories from infant-directed speech', *Proceedings of the National Academy of Science*, 104: 13273-13278
- Wade, T. and Holt, L. L. (2005). 'Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task', *Journal of the Acoustical Society of America*, 118: 2618-2633.
- Werker, J. F., and Tees, R. C. (1999). 'Influences on infant speech processing: Toward a new synthesis', *Annual Review of Psychology*, 50: 509-535.

Endnote

¹ Although within-category differences are diminished perceptually, it is now understood that learning speech categories does not produce entirely “categorical” perception (Liberman et al., 1957; Harnad, 1990). Infants (McMurray and Aslin, 2005) and adults (Lotto et al., 1998; McMurray et al., 2008) remain sensitive to within-category acoustic variation. Speech categories exhibit graded internal structure such that instances of a speech sound are treated as relatively better or worse exemplars of the category (Miller and Volaitis, 1989; Johnson et al., 1993; Iverson and Kuhl, 1995; Iverson et al., 2003).