

***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

**Author Biography:** Lori L. Holt is an Associate Professor of Psychology at Carnegie Mellon University. She investigates the perceptual and cognitive mechanisms contributing to speech perception and how speech can be used to understand auditory cognition using human perceptual and learning paradigms, animal behavioral experiments and computational models. Holt received a B.S. in psychology from the University of Wisconsin-Madison in 1995 and a Ph.D. in cognitive psychology with a minor in neurophysiology from UW-Madison in 1999.

**Abstract:** 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 experience to observe the consequences of different patterns of experience. Converging methods of cross-language observation, laboratory-based training of speech and non-speech 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.

**Keywords:** speech, categorization, auditory, learning, cognitive

## Categorization & Speech

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 & Lotto, in press). 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). In stark contrast, adults have difficulty discriminating even highly salient differences between some non-native sounds. Native Japanese adults are poor at discriminating English /r/ versus /l/ (Strange & Jenkins, 1978), 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 non-native sounds (see Werker & Tees, 1999). Decreases in speech discrimination occur among speech contrasts similar to those of the native language (Best, 1995; Flege, 1995) but not for very dissimilar sounds (like Zulu clicks for English listeners, Best et al., 1988), which continue to be discriminated.

This experience-dependent change 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 auditory processing. At this point, infants’ speech discrimination is driven by psychoacoustic differences among speech sounds. Speech category learning warps perceptual space in ways that reflect regularities of the native speech input and infants begin to perceive speech relative to the characteristics of the native language rather than solely according to psychoacoustic differences. Perceptual space expands and shrinks as a function of the salience of a dimension in contributing to categorization (Nosofsky, 1986) and there are regions of increased within-category similarity contrasted with regions of reduced between-category similarity (Liberman et al., 1957). Ultimately, this serves to exaggerate between-category differences (promoting discrimination) and shrink within-category differences (promoting generalization)<sup>1</sup>. Sounds (like Zulu clicks) that fall in regions of perceptual space not inhabited by native sounds are relatively spared the effects of native categories and continue to be distinguished according to their

---

<sup>1</sup> 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 & 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 & Volaitis, 1989; Johnson et al., 1993; Iverson & Kuhl, 1995; Iverson et al., 2003).

acoustic differences. Thus, with category learning, the perceptual space is organized to accommodate structure of the native language with some cost to perceiving acoustic differences among similar non-native sounds that do not align with native-language regularities (Best, 1995; Flege, 1995).

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. 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). Moreover, although the groundwork for speech category learning begins in infancy, there is a lengthy developmental course of category refinement whereby even 12-year-olds have not reached adult levels of speech categorization for some native sounds (Hazan & Barrett, 2000).

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 (Burnham et al., 1991). In infancy, the diminution of discrimination for particular non-native speech contrasts is well-predicted by the relationship of the non-native sounds relative to the developing native speech categories (Best, 1995; Flege, 1995; Kuhl, 2000a) and these relationships appear to determine, at least in part, the ease with which speech categories of a second language are learned.

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. At least part of the reason for this is because it is not feasible to entirely control and

manipulate speech experience to observe the consequences of different patterns of experience. Natural cross-language comparisons, 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 they are, by necessity, more descriptive than predictive.

### **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 perspectives. These approaches emphasize ecological validity via correspondence to natural speech processing versus experimental control over experience to greater and lesser degrees. Ideally, the approaches converge such that it is possible to develop predictions from tightly controlled laboratory experiments that may be tested in more natural speech communication. Representative research findings are described below to exemplify multiple methods of understanding speech category learning.

### ***Cross-language comparisons***

Cross-language comparisons provide a natural experiment in differing histories of speech experience. These studies emphasize how profound an influence experience has on speech perception and have defined questions of significance in understanding speech category learning. However, a drawback is that experimental control of experience is limited.

Nevertheless, clever investigations of naturally-occurring differences in listeners' histories of speech experience provide important insights. 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 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 was not statistically distinguishable from native listeners. Flege suggests that mechanisms that guide first-language speech category learning remain intact but that first and second language processing share common resources and therefore mutually influence one another. Naturalistic studies indicate plasticity in adult non-native speech category learning but, as is the case for infant learning, mechanistic details remain unclear.

### ***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 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 regularities and perception is measured thereafter to ascertain the influence of experience. 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 distributional exposure to 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. 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 (or just a few) acoustic 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 an 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 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 & 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 & 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 training sets of computer-synthesized speech. Although listeners learned to discriminate among training stimuli, they typically could not transfer this learning to natural speech or to different contexts (Strange & 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 & 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 hearing non-native sounds may reinforce existing native categories (via Hebbian synaptic plasticity; Hebb, 1949), interfering with learning. By this logic, beginning training with exaggerated instances of non-native speech falling outside native perceptual space and adaptively adjusting training stimuli to be more representative instances of the non-native categories may facilitate learning. McCandliss et al.'s results support this prediction, but also indicate an additional role for explicit feedback in learning (Tricoli et al., 2006) suggesting a more complex set of mechanisms.

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 & Morosan 1986) and this may be predicted by the relationship between first and second language categories and their interaction (Best, 1995; Flege, 1995). Even among more easily-learned categories, there are enormous individual differences in adult speech category learning (Golestani & Zatorre, in press). Often, this is a limitation since it can be difficult to draw general conclusions. Although it is not yet the norm for studies to investigate individual differences in detail, research can capitalize on individual differences to understand more about the auditory and cognitive constraints on speech category learning.

### ***Laboratory-based non-speech 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 possess no *a priori* categories. The major benefit of training listeners to categorize such non-speech sounds is that it is possible to

exert control over and have knowledge of listeners' experience with the sounds, thus developing an understanding of the general perceptual and cognitive constraints on auditory processing that might influence speech categorization.

The literature in this area is not yet large, but there are already insights about category learning relevant to speech. 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 & Van Hesse, 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 non-speech sounds modeling rapidly-changing acoustic dimensions of consonants or steady-state acoustic dimensions of simplified vowels. Patterns of non-speech discrimination and categorization mirrored those of the speech stimuli they modeled. A general auditory propensity for more quickly-decaying perceptual memory traces for rapidly-changing sounds relative to steady-state sounds could account for this pattern for both speech and non-speech.

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 (e.g., Kuhl, 1993). 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 non-speech sounds that varied along the 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. This study sheds light on why it seems to be easier to train adult listeners to categorize non-native voicing categories compared to other non-native speech categories (Pisoni et al., 1982; Jamieson & Morosan, 1989).

Non-speech category learning studies also highlight how task influences category learning. Discrimination training (explicit comparison of stimuli) and categorization training (responding to exemplars as category members) warp listeners' perception of non-speech stimuli, but in different ways. Discrimination training increases listeners' sensitivity to small distinctions among stimuli thereby working against categorization (Guenther et al., 1999). This insight from non-speech category learning has implications for interpreting the fact that Japanese listeners trained to discriminate English /r-/l/ (Strange & Ditman, 1984) do not generalize well to natural speech categories.

Another characteristic of speech categories is their multidimensionality. Multiple acoustic dimensions covary with speech categories and they are not equally diagnostic of category membership. English listeners, for example, perceptually weight formant frequency much more than vowel duration in categorizing /i/ and /ɪ/ although both dimensions covary with the categories (Hillenbrand et al., 2000). The control over experience afforded by non-speech categories allowed Holt and Lotto (2006) to investigate what kinds of input distributions

affect perceptual cue weighting in category learning. In brief, changes in relative input distribution variance are effective in shifting listeners' perceptual weighting. Since adults' non-native speech categorization is characterized by non-native perceptual weighting (Iverson et al., 2003), this study has implications for more effective training of adult non-native speech categories.

An issue in the above studies is that laboratory category training involved explicit feedback atypical of natural speech experience. Goudbeek and colleagues (2005; 2008) have used non-speech categories to investigate the role of feedback in category learning, reporting that without explicit feedback listeners find it very difficult to learn to categorize multidimensional sounds. 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 & Schwade, 2008), it is not the sort of explicit feedback that appeared to be necessary in this study.

Bridging this gap, Wade and Holt (2005) developed a space-invaders-style videogame involving visual creatures, each associated with a category of sounds. The sounds were designed to model some of the 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, but sounds served a

function. After 30 minutes, listeners' responses indicate 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 may be significant in complex, multidimensional category learning.

To date, there are relatively few non-speech 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 distributional learning, how acoustic dimensions are perceptually weighted and how task affects the warping of perceptual space. The opportunity to carefully manipulate experience with non-speech categories provides an opportunity to investigate these issues in greater depth to discover constraints on auditory learning relevant to speech categorization.

### ***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. Animals as diverse as birds, macaques, and chinchillas can discriminate speech (Dewson, 1964; Burdick & Miller, 1975; Kuhl & Miller, 1975; Morse & Snowdon, 1975; Dooling & Brown, 1990) and here 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 & Miller, 1975, 1978; Kluender & Lotto, 1994; Dooling, 1995; Lotto et al., 1997; Sinnott et al., 1998). For example, Lotto and colleagues (1997) found that Japanese quail trained to respond to /ga/-/da/ exhibit shifts in behavior as a function of

preceding speech context. Thus, an animal species exhibits the same pattern of response taken as evidence of perceptual compensation for coarticulation among human listeners (Mann, 1980), indicating perceptual constraints on speech processing may play a role in context-dependent speech categorization. Such studies serve as an existence proof that general perceptual processing that can meet some speech perception challenges (Kluender et al., 2005).

Animal models also serve as explicit models of the effects of experience on speech processing. For example, Kuhl (1991) reported that monkeys do not show the pattern of graded internal vowel category response observed for human adults and infants (Grieser & Kuhl, 1989), perhaps indicating species-specificity in this aspect of categorization. However, the monkeys had no experience with speech. When Kluender and colleagues (1998) provided birds experience with vowel input distributions, birds' responses were graded and highly correlated with human listeners' categorization responses to the same sounds. Experience with the distributional characteristics of speech categories is essential in producing graded responses to speech, but without the possibility of control over experience this prospect would not have been testable.

Control over animals' speech experience allowed Holt, Lotto, and Kluender (2001) to conclude that the relationship of fundamental frequency and voicing (as observed in English and other languages, see Kingston & Diehl, 1994) is not an obligatory influence of  $f_0$  on voicing arising from perceptual constraints, but rather is more likely due to the learnability of covariation between acoustic dimensions. It arises only when animals experience correlation between the acoustic dimensions in 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 available to a quail.

The issue of how feedback influences speech category learning is relevant to animal models of speech categorization 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 birds responded to some 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 can 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 correlation among acoustic dimensions (Holt et al., 2001) are characteristics of speech category learning that have been illuminated by animal learning models.

## 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 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.

The brief overview above highlights representative research approaches that extend a bit beyond 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. As yet, this goal has not been fully met. However, 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.



## 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., & Sithole, N. M. (1988). Examination of the perceptual re-organization for speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception & 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 & Psychophysics*, 61, 977-985.
- Burdick, C. K., & 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., & Brown, S. D. (1990). Speech perception by budgerigars (*Melopsittacus undulatus*): Spoken vowels. *Perception & Psychophysics*, 47, 568-574.
- Dooling, R. J., Best, C. T., & 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. & MacKay, I. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 26, 1-34.
- Goldstein, M. H., & 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, & S.R. Robinson (Eds.), The Oxford Handbook of Developmental and Comparative Neuroscience. Oxford University Press.
- Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, 19, 515-522.
- Golestani, N. & 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., Swingle, D., & Cutler, A. (2005). Acquiring auditory and phonetic categories. In H. Cohen, & C. Lefebvre (Eds.), *Categorization in Cognitive Science*, pp. 497-513. Amsterdam: Elsevier.
- Goudbeek, M., Cutler, A., & Smits, R. (2008). Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication*, 50, 109-125.
- Grieser, D., & 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., & 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., & 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., & 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., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America*, 108, 3013–3022.
- Holt, L. L., & Lotto, A. J. (accepted pending minor revisions). Speech perception as categorization. *Perception & Psychophysics*.
- Holt, L. L. & 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., & 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., & Kluender, K. R. (1998). Incorporating principles of general learning in theories of language acquisition. In M. Gruber, C. Derrick Higgins, K. S. Olson & T. Wysocki (Eds.), *Chicago Linguistic Society, Volume 34: The Panels*. Chicago: Chicago Linguistic Society, 253-268.
- Holt, L. L., Lotto, A. J., & 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., & 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., & 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. & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47-57.
- Jamieson, D. G., & 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. & 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., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, 70, 419–454.
- Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science*, 237(4819), 1195-1197.
- Kluender, K. R., & 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. (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., Lotto, A. J., Holt, L. L., & 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 & 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. deBoysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & 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., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Native Language Magnet Theory Expanded (NLM-e). *Philosophical Transactions of the Royal Society B*, 363, 979-1000.
- Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190, 69-72.
- Kuhl, P. K., & 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. & 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., & 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., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606-608.
- Lieberman, A. M., Harris, K. S., Hoffman, H. S., & 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., & 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.
- Logan, J., Lively, S. & 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., & 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., & Holt, L. L. (1998). Depolarizing the perceptual magnet effect. *The Journal of the Acoustical Society of America*, 103, 3648-3655.
- Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, 28, 407–412.
- Maye, J., Werker, J. F., & 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., & 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. & 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., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, 46, 505–512.
- Mirman, D., Holt, L.L., & McClelland, J.L. (2004). Categorization and discrimination of non-speech sounds: differences between steady-state and rapidly-changing acoustic cues. *Journal of the Acoustical Society of America*, 116, 1198-1207.
- Morse, P. A., & Snowdon, C. T. (1975). An investigation of categorical speech discrimination by rhesus monkeys. *Perception and Psychophysics*, 17, 9-16.
- 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., Aslin, R. N., Perey, A. J., & 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., & 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.



- Repp, B. H. (1984). Categorical perception: Issues, methods, findings. *Speech and Language*, 10, 243–335.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.
- Schouten, M. E., & Van Hesson, A. J. (1992). Modeling phoneme perception: Categorical perception. *Journal of the Acoustical Society of America*, 92, 1841–1855.
- Sinnott, J. M., Brown, C. H., & Borneman, M. A. (1998). Effects of syllable duration on stop-glide identification in syllable-initial and syllable-final position by humans and monkeys. *Perception & Psychophysics*, 60, 1032-1043.
- Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, 36, 131-145.
- Strange, W., & Jenkins, J. J. (1978). Role of linguistic experience in the perception of speech. In R. D. Walk, & 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., & Fiez, J.A. (2006). Performance feedback drives caudate activation in a phonological learning task. *Journal of Cognitive Neuroscience*, 18, 1029-1043.
- Wade, T. & 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., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, 50, 509-535.