ABSTRACT: It is clear that the ability to learn new speech contrasts changes over development, such that learning to categorize speech sounds as native speakers of a language do is more difficult in adulthood than it is earlier in development. There is also a wealth of data concerning changes in the perception of speech sounds during infancy, such that infants quite rapidly progress from language-general to more language-specific perceptual biases. It is often suggested that the perceptual narrowing observed during infancy plays a causal role in the loss of plasticity observed in adulthood, but the relationship between these two phenomena is complicated. Here I consider the relationship between changes in sensitivity to speech sound categorization over the first 2 years of life, when they appear to reorganize quite rapidly, to the “long tail” of development throughout childhood, in the context of understanding the sensitive period for speech perception.

And the Gileadites took the fords of the Jordan against the Ephraimites; and it was so, that when any of the fugitives of Ephraim said: “Let me go over,” the men of Gilead said unto him: “Art thou an Ephraimite?” If he said: “Nay”; then said they unto him: “Say now Shibboleth”; and he said “Sibboleth”; for he could not frame to pronounce it right; then they laid hold on him, and slew him at the fords of the Jordan; and there fell at that time of Ephraim forty and two thousand. (Judges 12:5-6)

The consequences of failing to learn a new speech contrast later in life are rarely as dire as depicted in this biblical account, and yet there is an extensive scientific literature concerning the mechanisms underlying the loss of plasticity for speech contrasts over development. It is clear from this research that learning to categorize speech sounds in a second language (L2) is more difficult in adulthood than it is earlier in development (e.g., Bradlow & Pisoni, 1999; Flege & MacKay, 2010). There is also a wealth of data concerning changes in the perception of speech sounds during infancy, such that infants quite rapidly progress from language-general to more language-specific perceptual biases (Kuhl, 2004; Werker & Tees, 1999). This process is commonly described as an example of “perceptual narrowing” in which an initially broadly tuned perceptual system becomes increasingly specialized over the course of development, so that it is eventually finely tuned to support species-specific (or in the case of speech, language-specific) responses to the environment (e.g., Lewkowicz & Ghazanfar, 2009; Scherf & Scott, this issue). It is often suggested that the perceptual narrowing observed during infancy plays a causal role in the...
loss of plasticity observed in adulthood. The relationship between these two phenomena seems to be more complex than this, however.

In this review, I will first lay out the potential problem that this disconnection between the rapid changes in perception over infancy and the relatively gradual, long-term changes in plasticity for speech sound categorization pose for theoretical positions that propose a relationship between learning of native-language categorization skills and loss of plasticity (“entrenchment” or “neural commitment” hypotheses, e.g., Kuhl, 2004; Seidenberg & Zevin, 2006; Werker & Tees, 2005). Next, I consider the literature on the development of speech sound categorization abilities over the course of infancy, and into later childhood in light of this view. One way to reconcile this literature with the entrenchment hypothesis is to consider that the process by which learning contributes to a loss of plasticity, while it begins in infancy, is actually continuous, so that the later, more gradual refinements during the “long tail” of this developmental process may be more important in understanding the sensitive period for speech perception than the well-documented changes in categorization that occur over infancy. Finally, I discuss some of the difficulties inherent in measuring speech sound categorization across the lifespan, and how these may reflect more than just methodological inconveniences. Some of these difficulties arise because of developmental differences in the goals and interests of the learner, and these are important factors to keep in mind when considering developmental differences in the rate and success of learning.

CATEGORICAL PERCEPTION, PERCEPTUAL ASSIMILATION, AND THE FAUSTIAN BARGAIN OF EXPERT SPEECH SOUND CATEGORIZATION

When perceptual abilities for speech are measured very early in infancy, little evidence for language-specificity is observed. This is sometimes taken to mean that infants are innately equally capable of distinguishing all of the speech contrasts used in all of the world’s languages, and the development of speech perception proceeds by a largely selective mechanism by which those contrasts that are not heard in the ambient language are deleted (see, Eimas, 1975, for an early formulation of this view, which still appears in many general psychology and linguistics texts). In fact some of the initial perceptual biases infants evince are inconsistent with the languages they will eventually learn (Lasky, Syrdal-Lasky, & Klein, 1975), and not all phonetic contrasts are equally easy for infants to perceive (Eilers, Wilson, & Moore, 1979). It has thus long been suspected that the development of speech sound categorization is best characterized by a heterogenous set of learning mechanisms including enhancement of some perceptual boundaries and De novo formations of others in addition to selection (Aslin & Pisoni, 1980; Aslin, Werker, & Morgan, 2002).

Nonetheless, it is clear that within the first year of life perceptual abilities become tuned to the ambient language, and that one consequences of this is that infants become less sensitive to distinctions that are not used in their ambient language (e.g., Eimas, Miller, & Jusczyk, 1987; Kuhl et al., 1997; Werker & Tees, 1984; among many others). This is such a striking finding because it occurs in the context of an overall developmental pattern in which perceptual abilities at other levels of analysis are enhanced (e.g., Lippé, Kovacevic, & McIntosh, 2009; Norcia & Tyler, 1985). The loss of sensitivity to differences that are not meaningful within the context of the target language are related to the notion of categorical perception, which has long been considered a key feature of speech perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967): that is, relative to their physical differences, members of different categories are perceived to be more different from one another, and members of the same category are perceived to be more similar to one another.

The fact that patterns of categorical perception appear to change so rapidly to accommodate the language environment in infancy, and yet can remain stubbornly fixed in adulthood is in itself evidence that plasticity for speech perception changes drastically over the life span. There is also some consensus that this early plasticity itself sets the stage for failures of learning later in life. The notion is that at some point in development, native-language (L1) categories are so well-learned that second language (L2) categories are largely perceived in terms of how they assimilate to L1 categories (Best, 1995; Flege, 1995). The shift from language-general to language-specific perception documented in infancy may be the first step in the process of neural commitment (Kuhl, 2004) or entrenchment (McClelland, Thomas, McCandliss, & Fiez, 1999; Seidenberg & Zevin, 2006) in a perceptual system that emphasizes some stimulus contrasts and diminishes others.

A Paradox for Entrenchment Models?
The principle of entrenchment—that early learning to support language processing creates internal structure in neural networks, that can then in turn interfere with
later learning—can be modeled in simple computational simulations with learning dynamics that have some interesting properties of learning in neural systems (McClelland et al., 1999; Seidenberg & Zevin, 2006; Thomas & Johnson, 2006; Vallabha, McClelland, Pons, Werker, & Amano, 2007). In the McClelland et al. model, the logic by which this could occur is demonstrated very clearly: “perception” and “learning” are both served by the same set of weighted connections between an input layer (standing in for perceptual information from more general levels of auditory processing) and a “representation” layer (standing in for cortical regions that represent categorical information). A competitive Hebbian learning rule is used, so that early biases in the model’s perceptual abilities are sharpened over time into a very strong attractor states for a small set of categories. Categorical perception emerges in the model because competition to arrive at a stable output representation results in a set of weights that cause many overlapping input patterns to generate an identical pattern on the representation layer. Once this perceptual effect is in place, it becomes impossible to undo because the network dynamics of learning depend the existence of some residual ambiguity in the representation layer. If only one category is active (and the rest of the representation layer is inactive) the learning rule cannot induce any change in the network. At some point in development, the perceptual system chooses rapid, unambiguous categorization at the cost of weights that cause many overlapping input patterns to generate an identical pattern on the representation layer. Once this perceptual effect is in place, it becomes impossible to undo because the network dynamics of learning depend the existence of some residual ambiguity in the representation layer. If only one category is active (and the rest of the representation layer is inactive) the learning rule cannot induce any change in the network. At some point in development, the perceptual system chooses rapid, unambiguous categorization at the cost of the ability to learn new categories. The key feature of this and related models is that nothing changes about the learning rule itself, or the parameters of the network that have a direct effect on plasticity. That is, the modelers made no attempt to simulate broad changes in the neural substrate for learning, such as synaptic pruning, changes in NMDA receptor subtype or distribution, formation of perineuronal nets, or any other biological changes aside from the adjustment of synaptic weights that result directly from learning. Changes in plasticity arise from the fact that learning and perception are intimately linked.

It cannot be the case, however, that changes in the responses to speech contrasts that are measured within the first year of life are sufficient to complete this process. Limits on when new speech contrasts are naturally learned from exposure to a second language emerge much later in childhood. For the entrenchment hypothesis—or any of the related theoretical positions—to hold, it must be that some residual sensitivity to non-native contrasts remains, even after the large, initial reorganization that appears to occur in infancy. This later fine-tuning of perceptual categorization would then play a causal role in limiting plasticity for speech sound categorization.

**RECONSIDERING THE DEVELOPMENT OF SPEECH SOUND CATEGORIZATION**

Two questions about the development of speech sound categorization are critical to assessing the viability of the entrenchment hypothesis. First, what is the nature of the change in perceptual categorization of speech sounds during the initial, rapid phase that occurs during infancy? If the shift from “language-general” to “language specific” perception were better characterized as qualitative than quantitative, this would pose a serious problem for any account of changes in plasticity after infancy that depend on learning mechanisms that underlie this initial reorganization. Second, is there any evidence of continued fine-tuning of speech sound categorization after infancy? If not, it would be difficult to maintain that learning plays any role in changing plasticity during childhood. Thus, both of these questions are vital to understanding the role of learning in the sensitive period for speech perception.

**Measures of Speech Perception Through the Ages—Infancy**

Some of the most important data regarding the development of language-specific perceptual abilities for speech comes from studies comparing infants in different language environments. These are heroic studies, involving cleverly contrived measures of perception: you cannot, as it happens, ask an infant to push one key if a pair of stimuli sound identical, and another if they differ. Thus, measures of what infants can perceive are necessarily indirect, and most often involve some form of short-term learning (either habituation, operant conditioning, or some combination of the two Houston-Price & Nakai, 2004; Werker, Polka, & Pegg, 1997; Werker et al., 1999). Evidence that infants are capable of perceiving a contrast most often come in the form of slightly (but significantly) greater than chance responding to changes in a stream of repeated stimuli. Although they can provide a wealth of information about the relative difficulty of speech contrasts, and how this is influenced by language experience, they may vastly underestimate infants’ perceptual abilities.

Ambiguities in the infant literature are thrown into relief when we consider how adults perform in similar tasks. Eilers et al. (1979) used a reinforced looking paradigm to measure infants’ and adults’ sensitivity to different voice onset time (VOT) boundaries, and found striking differences between adults and infants. VOT is one of the primary cues that distinguishes so-called “voiced” consonants, such as /b/, /d/, and /g/ from “voiceless” consonants, such as /p/, /t/, and /k/. Languages differ with respect to where the boundary
between voiced and voiceless is set. In English, for example, voiced consonants tend to have a VOT near or slightly greater than zero (i.e., the vowel begins almost immediately after the consonant is released), whereas for voiceless consonants there is a period of 50–100 ms of aspiration between the consonant release and the vowel.

Many previous studies had found evidence for categorical perception of VOT boundaries in infants (Eimas, 1974), but when they used the same task with infants and adults, Eilers et al. (1979) found that 6 months old, native English-speaking infants were only consistently sensitive to the English boundary, whereas adults were sensitive to a wide range of boundaries; in particular, adults with some formal training in phonetics performed significantly above chance in discriminating every contrast presented. This result is hard to reconcile with a model of the development of speech perception in which perceptual abilities narrow from language-universal to language-specific over the first year of life. Methodological problems with this study have been discussed at length (see, e.g., Aslin & Pisoni, 1980), and there is something unsportsmanlike about pitting trained phoneticians against infants, to be sure, but the study raises an important issue about what is being measured in infant studies. Although they are clearly sufficient to reveal differences between groups with respect to sensitivity to particular contrasts, the methods used to study speech perception in infants are not well suited to any claims about what infants cannot hear. The vital issue here is what can be concluded about the state of the learner with respect to what they will be able to learn.

If failure to discriminate in a conditioned head turn task were a reflection of what infants can do “in the limit,” we could not explain de novo category formation, or category boundary shifts, but there is evidence for both patterns in the literature. Hoonhorst et al. (2009) showed that French-hearing infants go from showing evidence for categorical perception at the “language-general” boundaries of a VOT continuum (including the English contrast observed in previous studies, Eilers et al., 1979; Eimas, 1974; Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Lasky et al., 1975) at 4 months of age, to boundaries that more closely approximate the distribution for French at 8 months. This pattern takes the form of an age by contrast interaction, such that heart rate changes to within-category contrasts are diminished, whereas heart-rate changes to between-category changes are enhanced over this period. Critically, the younger infants showed no evidence of discriminating the French contrast, distinguishing this study from previous work demonstrating increased sensitivity to native contrasts (e.g., Polka, Colantonio, & Sundara, 2001; Sundara, Polka, & Genesee, 2006; Tsao, Liu, & Kuhl, 2006).

Narayan, Werker, and Beddor (2010) found evidence for a similar developmental pattern in a study of older infants using two nasal contrasts (/ma/-/na/ and /ma/-/Na/). At 6–8 months, neither English-learning infants nor Filipino learning infants showed evidence of discriminating the less acoustically salient Filipino /na/-/Na/ contrast; Filipino-learning infants (but not English-learning infants) showed sensitivity to this contrast at 10–12 months. They discuss their results in terms of different factors that influence the salience of a contrast. Contrasts that are not very salient acoustically may depend on support from the ambient language to be treated as behaviorally relevant, whereas highly salient contrasts may continue to elicit responses indicating discrimination despite a total lack of evidence in the language (Best, McRoberts, & Goodell, 2001).

Thus, contrasts that are used consistently in the ambient language have enhanced salience relative to contrasts that are not—and contrasts that are unattested in the ambient language may also have diminished salience, but this is not the same as saying that they have been “lost.” This point is critical for the entrenchment hypothesis. If perceptual abilities were shaped so dramatically at this stage that infants assimilate non-native contrasts to their native system, as adults appear to do, changes in plasticity that occur after infancy would be difficult to explain in terms of continuous learning mechanisms.

### Measures of Speech Perception Through the Ages—Childhood

The evidence that perceptual abilities for speech are shaped by native language experience in the first years of life is compelling and abundant. Curiously, there are relatively few studies examining whether or how these perceptual abilities continue to develop later into childhood. Studies that have looked at the development of phonetic processing suggest that it is surprisingly protracted. For example, Hazan and Barrett (2000) studied children ranging in age from 6 to 12 in a series of discrimination tasks with synthetic stimuli from four different continua. They found subtle differences between adults and even the oldest children tested. Similarly, monolingual English-speaking children run...
as part of a second language speech perception experiment by Tsukada et al. (2004) differed significantly from adults in their ability to perceive some difficult English contrasts, but not others.

Much of the data on the development of speech perception during childhood comes from studies comparing atypically developing populations to matched controls. For example, Serniclaes, Van Heghe, Mousty, Carre, and Sprenger-Charolles (2004) found significant differences between dyslexic and typically developing children at a mean age of 9 years: dyslexics were more accurate than controls at making within-category discriminations, suggesting that their categorization of these sounds is less finely tuned. Serniclaes et al. (2004) refer to this as an “allophonic mode of perception,” because differences between members of the same phonological category are enhanced (see also Bogliotti, Serniclaes, Messaoud-Galusi, & Sprenger-Charolles, 2008). Interestingly, their results also revealed subtle differences between the typically developing children and adults in the same direction as the differences between dyslexics and controls.

Studies comparing children’s performance to adults’ can be difficult to evaluate, because differences in perceptual abilities may be exaggerated (or mimicked) by differences in the ancillary cognitive and decision-making aspects of the task. Data concerning qualitative shifts in performance over development are thus of particular interest, because they cannot easily be accounted for by these factors (Nittrouer, Crowther, & Miller, 1998). For example, Nittrouer (2002) examined the interaction of frication noise—the noise generated by air turbulence in the mouth when producing sounds like the first sound in “phoneme”—and formant transitions—changes in the spectral structure of the sound during the transition between consonants and vowels—in distinguishing different fricative contrasts. For a contrast in which these two sources of information do not interact (i.e., /θ/-/θ/), she found no differences between children and adults in the use of the two cues, whereas for a contrast in which they are known to interact (i.e., /s/-/s/), a clear developmental pattern of greater reliance on spectral properties of frication noise over development was observed. Critically, these developmental changes do not depend on quantitatively increasing accuracy or categoricity of a perceptual contrast, and are difficult to ascribe to general changes in competence with the particular tasks employed.

Thus, while the data on the development of speech sound categorization between infancy and adulthood are sparse, they provide substantial evidence that the process of fine-tuning perceptual categorization abilities continue quite late into development. It seems plausible, then, that this protracted process of fine-tuning plays a role in the loss of plasticity for new speech sound categories, as predicted by the entrenchment model. Taken together with the infant data, these findings also open up new questions. For example, it is unclear whether the process that underlies the protracted development of expertise throughout childhood is driven by the same mechanisms that underlie the initial reorganization of perception during infancy. Further, the relationship between changes in perceptual categorization over the course of childhood and loss of plasticity for L2 speech contrasts will require much more data concerning both phenomena to establish more firmly.

RELEVANCE, SALIENCE, INTERESTINGNESS, AND THEIR CONSEQUENCES FOR MEASURES OF PERCEPTION

Methodological considerations create considerable difficulties for understanding the relationship between changes in perception over infancy and the loss of plasticity for speech sound categorization that manifest in adulthood. When interrogating the representations that infants bring to perceptual tasks, researchers are at the mercy of their participants with respect to the available response modalities—changes in heart rate, high-amplitude sucking, conditioned head-turn responses, etc. These methods share, first and foremost, a dependence on measures of change detection, expressed either as release from adaptation or conditioned responses to a salient change in the stimulus environment. Salience—perhaps a better word would be “interestingness”—is jointly determined by some combination of the infant’s ability to represent the input, her experience (the ambient language) and her goals with respect to speech stimuli, making it difficult to claim that these are “pure” measures of perception. There is, of course, a long and vital tradition of perspectives that treat the boundary among perception, attention and action as fluid, or even epiphenomenal (e.g., Gibson, 1986; Noé’s, 2004). Here I consider paradigms that use oblique, or passive tasks in adults, and conclude that it may not be possible to achieve the goal of disentangling behavioral relevance from perceptual processes, even in the absence of an overt task. I then consider a set of findings in the developmental literature that were initially understood as reflecting changes in perceptual abilities, but can be interpreted as driven by developmental shifts in the participants’ goals with respect to speech stimuli in general. Both themes demonstrate the key role of developmental shifts in the goals listeners bring to any task involving speech
stimuli in determining responses that are often interpreted as indices of perceptual ability.

**Oblique and “Passive” Measures of Speech Sound Categorization**

We have seen that differences of task pose problems of interpretation when comparing adults to infants and children. Applying the same task across ages is sometimes impossible. Consider the Eilers et al. (1979) study: infants and phonetically trained adults have very different goals, not to mention perceptual and strategic processes available to them. Even when the same task is appropriate, as in comparisons between adults and older children, care must be taken that group differences are not driven by developmental changes in abilities that are more related to laboratory tasks than the processes they are designed to measure. I have described this situation thus far as being a particular difficulty of developmental research, but in fact part of the problem is that adults are notoriously wily participants. Late learners of a second language are rarely completely at chance when discriminating even the most difficult L2 contrasts, and even participants who are naive to a language can do quite well with some contrasts, based on general perceptual abilities (e.g., Best et al., 2001). In the limit trained phoneticians can perceive and produce dozens of late-learned contrasts well enough to document and record them (e.g., Ladefoged & Maddieson, 1996), but this is a very different ability from native-like perception or production. One strategy is to try and tap adult perceptual abilities *obliquely*—by making the target contrast incidental to task goals—or to take advantage of neuroimaging techniques that permit measurement of perceptually driven responses in the absence of any overt task.

Oblique measures are collected in tasks where stimulus variability along the dimension of interest is nominally irrelevant to the task demands (e.g., Garner, 1974). For example, Navarra, Sebastian-Galles, and Soto-Faraco (2005) asked participants to identify the first syllable of four repeated two-syllable nonwords—/tike/, /tikE/, /pukE/, and /pukE/*—in two conditions: one in which the final (task-irrelevant) vowel was variable, and one in which it was held constant. Although the final vowel was task-irrelevant, and the distinction between /tik/ and /puk/ easily perceived by speakers of any language, a large influence of final vowel variability was observed on performance participants whose dominant language is Catalan, which contrasts /e/ and /E/. Native Spanish speakers, who assimilate both vowels to their own /e/ category did not show this effect, nor did Catalan-Spanish bilinguals whose dominant language is Spanish. Here again the notion of salience is key; the bilinguals do well enough with this contrast when tested directly on it, but it is not salient enough to influence responses when it is not task-relevant. The strong influence of language-dominance in this and other studies of early bilinguals (e.g., Sebastian-Galles & Soto-Faraco, 1999; Sebastián-Gallés, Vera-Constán, Larsson, Costa, & Deco, 2009; Werker & Byers-Heinlein, 2008) suggests that quality and quantity of experience with an L2 can play a crucial role in determining whether particular contrasts are learned, even among otherwise proficient speakers of a language, who have been using it for most of their lives.

Another common approach used to address the challenge of comparing tasks across developmental stages is to depend on more or less passive measures of perception that depend on habituation and change detection responses, and can thus be observed in the absence of explicit task performance. This is enabled by brain imaging, which permits the collection of data thought to reflect perceptual processes in the absence of an overt task. In a typical EEG experiment of this type, a stream of repeated stimuli is interrupted at intervals by a rare stimulus, and characteristic brain responses to the stimulus change are taken as an index of perceptual sensitivity (the mismatch negativity, or MMN, Naatanen, 2001) that can be used across development, beginning in infancy (Cheour, Leppanen, & Kraus, 2000; Dehaene-Lambertz, Dupoux, & Gout, 2000). In adults, the MMN is weaker for novel non-native speech contrasts, which is often understood to reflect weaker representations for those stimulus categories (Dehaene-Lambertz, 1997; Naatanen et al., 1997; Peltola et al., 2003). Interestingly, Winkler et al. (1999) found that relatively short periods of immersion in a second language could enhance the amplitude of the MMN to a second-language contrast. More recently, however, Zevin, Datta, Maurer, Rosania, and McCandliss (2010) found differences in topography between native listeners and relatively proficient non-native listeners, with a more left-lateralized response for natives. This agrees with locations found in fMRI using adaptations of the MMN paradigm (Celsis et al., 1999; Jacquemot, Pallier, LeBihan, Dehaene, & Dupoux, 2003; Joanisse, Zevin, & McCandliss, 2007; Liu, Skipper, McCandliss, & Zevin, 2009; Zevin & McCandliss, 2005).

Recent data suggest, however, that responses in left temporoparietal regions that seem to distinguish native from non-native speech perception may not directly reflect the representations that underlie speech sound categorization, but are more likely domain-general change detection mechanisms (Zevin, Yang, Skipper, & McCandliss, 2010). Differences in the strength of this response can be attributed to differential salience of the
stimuli. This is clearly influenced by whether the repeated and rare stimuli are contrastive in one’s native language, but also depends on other aspects of the experimental context. When a change between speakers is introduced in the same experiment, change detection responses to native-language phoneme categories are abolished. This could be taken to suggest that these responses are epiphenomenal, although differences between native- and non-native speakers in such paradigms still demand some explanation. One way to understand these findings is in terms of differences in the behavioral relevance to the native and non-native participants. This interpretation assumes that even nominally passive, perceptual processes are penetrated to some degree by the listener’s goals with respect to the stimulus class, and the dimensions over which they vary in the experimental context.

Changes in the Goals of Language Learning Over Development

In adults, even techniques that permit us to observe brain responses in the absence of an overt task do not license conclusions about raw perceptual abilities without reference to the difficult-to-define notion of “salience.” A similar problem emerges when considering the developmental literature—how do we distinguish between phenomena driven by changes in perceptual abilities and changes driven by developmental changes in the goals listeners bring to any task involving speech stimuli? One key example of developmental shifts in salience comes from apparent changes in phonemic representation that coincide with the earliest stages of word learning.

Stager and Werker (1997) habituated infants to repeated pairings of novel objects with novel nonsense words in a Switch task. Critically, in this task, habituation occurs to two pairing simultaneously, so that when the “switch” occurs, both the word and the object are familiar. If the infant provides evidence of detecting the switch—with renewed attention to the display—it is likely because she has learned about the co-occurrence of the particular word-object pairing (Werker, Cohen, Lloyd, Casasola, & Stager, 1998). When habituated to stimuli that differ as “minimal pairs,” that is, “bih” and “dih,” 14-month-old infants failed to demonstrate evidence of associative learning. This was surprising, first because infants at this age—and indeed as young as 8 months—were shown to be capable of perceiving this contrast when no word-learning task was present. Second, the younger infants also showed evidence of perceiving this contrast in the context of the Switch paradigm. One interpretation of this (see also Pater, Stager, & Werker, 2004) is that infants “have” the appropriate categorical information in place at this age, but that the task demands of word learning impose a cognitive bottleneck that prevents them from using this information for the task. These data were also seen as consistent with the view that word learning is at least initially holistic, and not dependent on knowledge of speech sound categories (Beckman & Edwards, 2000; Walley, 1993).

More recently, Yoshida, Fennell, Swingley, and Werker (2009) demonstrated that infants’ apparent failures to learn fine phonetic detail may be related to testing modality. When a visual choice task was used instead of the Switch task, 14-month olds showed some evidence for learning. Similarly, Fennell and Waxman (2010) found evidence that 14-month olds could learn new words with a relatively high degree of phonetic detail, when the task was designed so that they were felicitous as instances of naming. That is, when the to-be-learned word was presented in a series of sentences, or in the context of naming familiar objects, infants were sensitive to mispronunciations, in contrast to a control condition in which familiar items were paired with exclamations (Namy & Waxman, 2000).

The role of pragmatics here is key: any experimental situation that indicates to the infant that they are now “playing a naming game” will support them treating instances of word-object pairings as naming events, rather than a confusing confluence of two independent streams of information. The window during which infants fail in the Switch task for novel minimal pairs seems to close around 17 months of age, and individual differences in performance during this transitional phase is strongly influenced by vocabulary size (Werker, Fennell, Corcoran, & Stager, 2002). It may be that when infants are further along in the word learning process, they are more inclined to treat word-object pairings as instances of naming, and require less support from social or pragmatic cues or from enhanced salience arising from distributional information from recent experience. The difficulty of establishing the use of segment-like information for unfamiliar words early on in the process of word learning is striking in contrast to the abundant evidence for fine phonetic detail for familiar words much earlier in development, but it may be limited to particular nonecological contexts in which there is little context to establish the salience of particular stimulus features.

Further evidence for the role of experimental context in establishing the salience of a phonetic contrast comes from Thiessen (2007). They found that 15- to 16-month olds could be encouraged to use a high level of phonetic detail in encoding novel words when the habituation materials contained additional evidence for the contrast. Importantly, direct evidence for minimal
pairs (“dawgoo” and “tawgoo”) did not have this effect; only when the children heard the contrast of interest in highly distinct contexts (“dawbo” and “tawgoo”) did exposure to the contrast improve performance in the Switch task. Thiessen and Yee (2010) then asked how general this facilitation for the contrast was, by replicating their earlier design, but manipulating the similarity of the critical habituation stimuli to the target item. In this case, participants only distinguished the novel word “yad” from a distractor (“yat”) when the habituation trials included stimuli in which the contrast occurred in the same linguistic structure (i.e., as the coda of the same syllable in “boeyat” and “gooyad”), and not in other conditions (e.g., in the onset of the first syllable as in “dawbo” and “tawgoo”). The failure to enhance salience via presentation of minimal pairs is an important detail. Minimal pairs are perhaps not sufficiently distinct to support tracking of distributional information in a continuous stream (Emberson et al., submitted), despite being discriminable in the limit. Thus, one effective way to enhance the salience of a contrast is to present different phonemes in very different contexts.

Considered together, the methodological difficulties that arise in understanding oblique and passive tasks, and the phenomena discussed here suggest caution in interpreting any finding as reflecting purely perceptual abilities. For adults, a lifetime of experience with particular speech contrasts, and a local stimulus environment in which variability is limited to specific features can conspire to produce brain responses that appear to be specific to the perception of speech contrasts, but are more accurately described as being tuned to behaviorally relevant stimulus variability. In infants, a similar phenomenon arises in tasks that rely upon habituation: responses to stimulus variability are more readily observed when task or stimulus properties are properly aligned with the participants’ interests and goals with respect to speech sounds. The expression of perceptual abilities, and their exploitation with respect to particular tasks is dependent to a large extent on the state of the listener—both in terms of their developmental state, and in the more immediate sense of their orientation toward the task.

CONCLUSIONS

We have seen that changes in sensitivity to speech contrasts over the course of development are likely to reflect changes in the salience of change from one stimulus to another, and that throughout the life-span salience is task-and context-dependent, so that, for example, local enhancements of salience can increase responding, whereas distractions of various kinds can diminish the relative salience of contrasts, with a concomitant decrease in the relevant response. These interactions among task, context, and the perceptual abilities are further complicated by developmental changes in the level of description that best suits the learner’s goals, such that, for example, sound-object pairings under the same presentation parameters that are ignored early on in development, can become distracting later on, and are eventually treated as an opportunity for word learning. All of this complicates our understanding of the relationship between the acquisition of expertise in speech sound categorization and the loss of plasticity for this ability later in life.

Similar complications arise in examining the outcomes of this learning process. Whereas there is little doubt that natural immersion in an L2 environment is not sufficient for learning of speech sound categorization later in life, when training regimes are crafted specifically so that appropriate feedback is provided, adults show evidence of substantial plasticity (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002). One possibility is that adults fail to demonstrate plasticity for non-native speech sound categorization in part because they do not need this ability to attain proficiency with this aspect of the language to achieve their communicative goals. Spoken communication rarely depends on the ability to categorize or produce individual phonemes unambiguously, and thus the kind of feedback required to overcome existing biases based on native-language experience may not be available in the context of naturalistic exposure to a second language. Understanding the sensitive period for speech sound categorization may thus require a shift in focus away from the fine-grained discrimination abilities that are the primary indices of native-like speech perception and toward a more integrated perspective, taking into account how these perceptual abilities interact with developmental changes in how speech is used for communication that play out across all stages of development.

REFERENCES


Emerson, L. L., Liu, R., & Zevin, J. D. (submitted). Finding functional units amidst perceptual variability: or how is statistical learning accomplished using varying exemplars of complex, novel sound categories?


Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. (1975). VOT discrimination by four and six and a half month old...


Scherf, K. S., & Scott, L. S. Connecting developmental trajectories: Biases in face processing from infancy to adulthood. (in press).


