UC San Diego
UC San Diego Previously Published Works

Title
Chapter Three Protracted perceptual learning of auditory pattern structure in spoken language

Permalink
https://escholarship.org/uc/item/7t15k2g5

ISBN
978-0-12-817175-2

Author
Creel, Sarah C

Publication Date
2019

DOI
10.1016/bs.plm.2019.07.003

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Running title: Protracted perceptual learning

Protracted perceptual learning of auditory pattern structure in spoken language

Sarah C. Creel (screel@ucsd.edu)

UC San Diego

Department of Cognitive Science

Abstract

The chapter addresses the developmental time course of the perceptual learning underlying spoken language. Many researchers suggest that most of this learning has saturated around the end of the first year of life—during infancy. My own work with young children suggests a different perspective: protracted perceptual learning. I lay out the evidence for perceptual precocity and explain why I think the perceptual precocity story is wrong. Following this, I present my findings on young children's perceptual processing in speech and nonspeech contexts (word learning, voice learning, melody learning), using visual fixations and pointing accuracy as response measures. Across all measures, children undershoot adult performance, particularly when perceptual attributes are challenging to tell apart. I then describe alternative, non-perceptual explanations for these findings, as well as ways to address said explanations. Last, I outline possible paths toward a unified picture of auditory perceptual learning across development.

Keywords: eye tracking; speech perception; spoken language development; music perception; talker recognition; perceptual development; word recognition

This chapter addresses the question of how, and when, children come to discover the sound patterns of their native language or languages. While early theorists proposed that there might be innate sound categories, most researchers now agree that some sort of perceptual learning takes place that allows the child to specialize toward their native language(s). What they don't agree on is how long this learning takes, or what the end product is.

Rapidly? Perceptual precocity

One view might be termed the perceptual precocity view: children acquire the speech sounds of their native language over roughly the first year of life. The end product is a set of symbol-like speech sound representations that allow fast, efficient encoding of sound sequences for use during word learning. This is the view that is portrayed in most language acquisition textbooks and is the party line that most of us deliver to our undergraduate classes. Most researchers would probably acknowledge that this is an oversimplified interpretation of a more nuanced picture. Of course things are learned after a year of age! My concern, though, is that by promoting a simplified story, that of perceptual precocity, we have started to believe this story too deeply.

Slowly? Protracted perceptual learning

A different view is one of protracted perceptual learning. On this view, perceptual learning of language continues for years, without any clear endpoint or end product, but with incremental gains facilitating processing of other levels of language structure. This is a less popular perspective, maybe because it's less sexy: the brilliant baby narrative is one that appeals to the public and to undergraduate classes and fits neatly into a hierarchical view of language acquisition and language structure. Nonetheless, I hope to convince the reader that protracted perceptual learning is at least as interesting as, and more believable than, perceptual precocity.

Goals of chapter

Before jumping into the evidence associated with each of these views, here's a brief overview of the structure of this chapter. First, I will describe why it's consequential, for society and for science, to understand the time course of perceptual learning in spoken language. Next, I will outline evidence for early learning and the perceptual precocity view. I will then discuss why, even though the evidence is right, the perceptual precocity story itself is wrong. I will then offer a different perspective, the protracted perceptual learning perspective, that has emerged from my own research on word learning and nonspeech sound pattern acquisition. Following this, I consider several alternative, non-perceptual explanations for age-related changes in performance and offer suggestions for fruitful approaches to studying protracted perceptual learning. I close by outlining open questions in this area and summarizing the material presented here.

Slow or fast learning: consequences for society

Should the person on the street care about how long it takes to learn speech sound patterns? Yes! Many potential language learners seem to have internalized a precocity view, that language is best learned (as in, most rapidly and most completely) when young. Thus, if I am an adult, I'm pretty much out of luck, and it will be difficult. Yet there is ample evidence that non-infant learners are excellent at acquiring some aspects of language. There is also ample evidence that children take a long time to learn aspects of a new language. If we as a field fail to communicate this, the person on the street is left with some very defeatist expectations about their capacity to learn languages as adults. Further, the public's expectations of young children as language learners may be too high, imposing overly strict timelines for second-language-learning children in the school system. Again, more realistic expectations would benefit these learners. Finally, we need a full understanding of typical development of speech sound processing across childhood if we are going to adequately characterize and intervene upon developmental disorders.

Slow or fast learning: consequences for science

Why should scientists care about the time course of speech sound pattern learning? For one thing, it's a scientific question in and of itself. For another thing, it is closely linked to questions about the nature of learning and cognitive processes more broadly. Early learning is associated with theories of knowledge as feature hierarchies (features make up speech sounds, which make up words, which make up sentences, which make up discourses): if you hold a strict hierarchical view, the lower levels need to be learned (or at least need to exist) before the higher levels. Protracted learning, on the other hand, is associated more with holistic or configural representation, the notion that learners are mentally representing complex structure without composing it out of subparts. (Note that I'm not arguing for holistic in the sense of "vague," but rather in the sense of not-necessarily-analyzed.) On a configural account, it doesn't matter so much that you have completely encoded the lower-level units before encoding higher-level units. Thus, word learning does not have to wait until speech sound category learning is complete. Holistic representation figures strongly in the face recognition literature, but has received less attention in the language acquisition literature (though see Walley, Metsala, & Garlock, 2003). One could certainly hold a perceptual precocity view without assuming hierarchical representations, or a protracted learning view that also assumes some hierarchy, but these seem to be less represented.

Another reason scientists should care about the differing views of the time course of speech sound pattern learning is that one's view on this matter shapes the questions that get asked. An informal survey of the literature by my colleague Carolyn Quam and me suggests that the body of research on typically-developing speech processing in developmental populations is largest for infants, with a precipitous decline around age 2 years, a local maximum around 5 years, and another decline from age 7 years up through the undergraduate years. (Ironically, there appear to be more undergraduates represented in research on developmental speech processing than 3-year-olds or 7-17-year-olds, due to undergraduates' use as a convenience population of adult control participants.) Further, the local maximum at 5 years appears to be accounted for almost entirely by studies in the *Journal of Speech Language and Hearing Research*, which has a greater interest in language disorders. This may be because language disorders tend to become evident, and to become a target for remediation, near the start of formal schooling (around 5 years).

A third reason it's important to nail down the timing of speech sound pattern learning is to understand whether speech is "special:" does it develop on a rapid time course using a distinct module of brain tissue, a hard-wired network to which other learning domains do not have access? Or does it use the same brain regions to different ends? From this perspective, it is useful to understand perceptual learning across development in other auditory domains. These include non-speech regularities in the speech signal itself (talker-specific acoustic characteristics; accent variability; vocal emotional cues), as well as nonlinguistic sound patterns like musical systems in one's culture. Evidence from each of these domains suggests a relatively slow time course (talkers: Creel & Jimenez, 2012; Mann, Diamond, & Carey, 1979; accents: Creel, 2018; Floccia, Butler, Girard, & Goslin, 2009; Girard, Floccia, & Goslin, 2008; Jones, Yan, Wagner, & Clopper, 2017; Wagner, Clopper, & Pate, 2014; vocal affect: Friend & Bryant, 2000; Morton & Trehub, 2001; Nelson & Russell, 2011; Quam & Swingley, 2012; Van Lancker, Cornelius, & Kreiman, 1989; musical pitch: Creel, 2014b, 2016; Creel, Weng, Fu, Heyman, & Lee, 2018; Fancourt, Dick, & Stewart, 2013; Stalinski, Schellenberg, & Trehub, 2008). If all of these domains show relatively slow learning, is speech sound pattern learning different in showing fast learning? Instead, is it possible that we have the story wrong, and speech sound pattern learning is also slow?

Evidence for early learning

What is the evidence that has led to the perceptual precocity story? A lot of the evidence comes from the development of ingenious techniques to assess cognitive processing in nonverbal organisms: babies. Since babies can't talk, much less reason aloud about their thought processes, researchers have had to develop a special toolkit to figure out what they are thinking. Much of this toolkit leverages one of two things: infants' attention to unexpected things; or infants' long-term knowledge. I briefly discuss some common methods before detailing the findings they have yielded.

Methods used to investigate infant sound pattern processing

Inspecting the unexpected: habituation and conditioned head-turn paradigms

One common thread through multiple infant methods is that infants react to things that mismatch the preceding context. To perhaps overinterpret a bit, they respond to things that are surprising. If a sound becomes overly familiar, their attention starts to wane, a process called habituation. Researchers use this over-familiarity to test infants' abilities to tell one stimulus from another: they wait for waning attention to, say, the sound "pa", and then present a different sound, say, "ba." If the infant shows renewed attention when "ba" occurs, they must have detected the difference between "pa" and "ba." Such habituation-dishabituation paradigms can index pacifier sucking rate in very young infants (e.g. Eimas, Siqueland, Jusczyk, & Vigorito, 1971), or visual fixations in older infants (e.g. Polka & Werker, 1994).

A related paradigm for investigating infant speech perception is the conditioned head-turn paradigm (Eilers, Wilson, & Moore, 1977). Infants are conditioned (taught) that every time they hear a change stimulus (B) in a series of standard stimuli (AAA…), they will receive a visual reward. The visual reward, often a creepy animated toy that lights up and moves, appears in a location to the side of the infant's visual fixation point. After a training phase that uses starkly-different B stimuli, the researcher can measure the likelihood of a head-turn response to individual change stimuli. The ability to present multiple change stimuli to the same participant—rather than a single moment of dishabituation—makes this paradigm especially appealing (for various uses of this paradigm with speech and music materials, see Eilers et al., 1977; Kuhl, 1979; Polka, Colantonio, & Sundara, 2001; Sundara, Polka, & Genesee, 2006; Trehub, Bull, & Thorpe, 1984; Werker & Tees, 1983, 1984). An early commentary by Trehub (Trehub, 1979) expresses optimism that this paradigm may even bridge developmental divides. Unfortunately, it seems to have declined in popularity relative to habituation-dishabituation studies. A later commentary, again by Trehub (Trehub, 2012), notes that it suffers high dropout rates due to failure to meet criterion or failure to stay engaged throughout the test session, perhaps explaining its relative lack of popularity (see Gravel & Traquina, 1992, for low response rates in children at 1.5 years and older).

In recent decades, researchers have applied electroencephalography (EEG) and magnetoencephalography (MEG) to the study of spoken language development (Cheour et al., 2002; Cheour, Leppänen, Kraus, Cheour, & Leppa, 2000; Háden et al., 2009; lmada et al., 2006; Kuhl, Ramírez, Bosseler, Lin, & Imada, 2014). Again, things that do not match the surrounding context tend to generate neural processing differences that can be picked up by EEG or MEG, typically combined over multiple measurements. If one presents several standard stimuli followed by a deviant stimulus (AAAAB…), a change detection response may occur within hundreds of milliseconds of events in the auditory input. Still, it is not simple to understand how neural measures relate to overt behaviors. That is, it may be a step too far to claim that infants process speech sounds in an adultlike manner if one's infant evidence is from EEG and one's adult evidence is from an overt categorization task.

Infants look to named objects: Eye tracking

The visual world eye tracking paradigm has illuminated the world of infant language comprehension perhaps more than any other technique in the past two decades. If one is hearing speech about items in the visual field (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), one tends to look at those items, even if one is an infant (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998). Thus, visual fixations to named objects can be interpreted as comprehension. In the infant literature, a closely-related paradigm is looking-while-listening (Fernald et al., 1998), but as both the visual world paradigm and looking-while-listening measure fixation proportions to visual referents in response to hearing speech, I discuss them together here.

Multiple studies have looked at infants' recognition of highly-frequent, early-acquired words by assessing whether they look to a picture of the named word (say, a car) more than they look to another picture (say, an apple or an unfamiliar object). Infants at 14-15 months are sensitive to single-phonological-feature mispronunciations (Swingley & Aslin, 2002), and 19-month-olds show decreasing proportions of looks to named objects as the pronunciation of the name ventures farther and farther from its canonical pronunciation (White & Morgan, 2008). For example, infants who see two pictures (a car and a novel object) look most to a car picture when they hear "car," a bit less (but still more than at the other picture) when they hear "gar," still less for "dar," and very little for a completely dissimilar name (see Bergelson & Swingley, 2017, for evidence of earlier sensitivity to stronger mispronunciations). But does a weaker looking pattern to "gar" mean that infants can tell the /k/ category from the /g/ category? This is less clear. We just know that infants are sensitive to properties of /k/ in the context of the word 'car.'

But wait, not so vast

However, one set of infant studies suggests weaker sensitivity early in the second year of life: studies of word learning conducted using the Switch procedure (Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002). In this paradigm, instead of habituating to a sound, children are habituated to a sound-image pairing, such as the novel word "bih" with a novel object. In some variants, they are habituated to two sound-image pairings, in other cases just one. Once habituated, they are tested with a mismatch: one sound paired with the other image. While they do dishabituate when the words sound distinctly different ("lif" and "neem"), 14-month-olds do not dishabituate to similar-sounding words like "bih" and "dih" under these conditions, even though much-younger infants will dishabituate to a sound change alone (in the sense of exceeding chance performance as a group; Stager & Werker, 1997). It is not until 17 months or so that infants succeed at the task (Werker et al., 2002). This finding has been replicated multiple times (Pater, Stager, & Werker, 2004; Rost & McMurray, 2009; Werker et al., 2002). Since Stager and Werker's initial finding, a handful of studies have suggested that 14-month-olds can succeed (again, in group performance) given facilitative conditions such as presentation of words in full sentences (Fennell & Waxman, 2010), acoustic variability of various sorts (Galle, Apfelbaum, & McMurray, 2014; Rost & McMurray, 2009, 2010), or a more sensitive dependent measure (Yoshida, Fennell, Swingley, & Werker, 2009). It is not clear how to integrate these results into the picture of early development. On the one hand, it may just suggest a slight delay in the perceptual precocity timeline until 17 months or so. On the other, it may suggest that early learning is not the whole story. I return to this shortly.

Early learning of speech sound patterns: perceptual narrowing and possibly perceptual sharpening

Evidence for early learning in spoken language is irrefutable. Numerous studies using habituation or conditioned-head turn indicate that infants are sensitive to subtle but meaningful (to adults) differences in native-language speech sounds (e.g. Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, 1994; Werker & Tees, 1984). Studies of eye-tracked familiar-word recognition suggest that slightly-older infants (around 15 months onward) can detect deviations from familiar words' typical pronunciations with similar sensitivity (Swingley & Aslin, 2000, 2002). Studies of novel word learning suggest that by 17 months, perhaps earlier in some cases, infants can learn words distinguished by minute differences (e.g. Werker et al., 2002).

Two key aspects of this early learning are perceptual narrowing and perceptual sharpening. Perceptual narrowing is argued to occur over the course of the first year: younger infants detect native and non-native speech sound changes, while older infants only detect changes that are relevant in their native language (Kuhl et al., 1992; Polka & Werker, 1994; Werker & Tees, 1984). This perceptual narrowing pattern has been interpreted as infants beginning with perceptual sensitivity to all humanly-possible speech sounds and then "using or losing" sound contrasts depending on their presence or absence in the native language. Additionally, some research suggests that as non-native sounds are being lost, infants additionally gain sensitivity to native-language contrasts (Kuhl et al., 2006), sometimes referred to as attunement or sharpening.

An important aside is that such studies of speech-sound processing tend not to look at children past the age of 12 months. It is not clear why this cutoff age exists, whether it coincides with onset of word production, increased participant mobility (children who can walk away are not as easy to habituate), or perhaps a developmental decrease in researchers' interest in speech sound learning. As mentioned earlier, this represents a gap in the literature.

So the evidence suggests early learning, but what does all this mean? Commonly, this research is interpreted to suggest that infants' speech sound representations are qualitatively adultlike by 12 months of age or so (see Werker & Hensch, 2015, for a recent account). Whether they are quantitatively adultlike is a much harder question to answer, given that dependent measures vary so drastically between infants and adults. Effects in infant studies are typically only observable at the group level, not at the individual level (with the exception of conditioned head turn studies; Kuhl, 1979, 1983; Polka & Werker, 1994; Werker & Tees, 1984). Further, there may be a file-drawer problem, such that failures to find speech sound discrimination are filed away due to pessimism about their publishability. It would be useful in this respect to see more efforts by researchers, along with more editorial support, to publish null findings. In addition to recent (and deserved) attention to study preregistration and registered replications, this might be done in large groups of researchers with a large collection of studies. As one example, Floccia et al. (2016) recently published a set of 13 studies which largely failed to find a facilitative effect of infant-directed speech on learning. Still, the overarching story of infant speech sound acquisition is one of early achievement of pattern structure.

Why the story is wrong (even though the evidence is right)

Perceptual sharpening past the first year

There are several holes in the perceptual precocity story. First, it can't account for all of the infant data. Some speech sounds appear inaccessible until after a year of perceptual exposure, such as differences between the sounds "n" and "ng" (as in ri<u>ng</u>) at the beginnings of words in Tagalog (Narayan, Werker, & Beddor, 2010). Other speech sounds, such as the contrast between the English sounds in "<u>d</u>oze" and "<u>t</u>hose," appear to be distinguishable for multiple years in the absence of input (that is, in French speakers who do not experience this contrast; Polka et al., 2001; Sundara et al., 2006). As I already mentioned, there is evidence that infants increase in their speech sound sensitivity over the first year of life (Kuhl et al., 2006), rather than just maintaining sensitivity to within-language speech sounds. This sharpening, rather than plateauing at 12 months, also appears to occur in older children, possibly through the teen years (Hazan & Barrett, 2000; McMurray, Danelz, Rigler, & Seedorff, 2018; Nittrouer, 2001; Ohde & Haley, 1997; Rigler et al., 2015). This is consistent with young children's greater susceptibility to perceptual distortion vs. older children or adults, such as comprehending accented speech (Bent, 2014; Bent & Holt, 2018; Nathan, Wells, & Donlan, 1998) or speech in noise (Elliott, 1979; Fallon, Trehub, & Schneider, 2000, 2002).

Plasticity persists past the first year

Another hole in perceptual precocity is its implication that perceptual narrowing is profoundly consequential. If non-native sound categories are truly lost during the first year of life, it should be difficult, if not impossible, to discriminate speech sounds after one year. However, this is not the case. Learners show substantial residual plasticity in speech perception and production well past 12 months of age. Numerous studies demonstrate that child learners up to age 7 years or so can achieve large gains in speech sound perception and production (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004; Oh et al., 2011), not to mention many even-later learners who achieve excellent results (see Flege, 1987 for discussion). One might postulate that certain exceptional conditions reopen the learning mechanism to explain these findings (see, e.g., Werker & Hensch, 2015), but it would be more parsimonious to assume that the same perceptual learning system that operates during the first year of life continues to operate throughout the lifespan (as opined by Flege in the Speech Learning Model, e.g. Flege, 1992, 1995).

Perceptual precocity not necessary

A further reason to doubt the first-year-of-life time course for speech sound learning is that infants appear capable of learning words before their speech sound acquisition has solidified, at 6-9 months for common objects (Bergelson & Swingley, 2012; Jusczyk, 1993; Tincoff & Jusczyk, 1999) and even earlier for their own names (Bouchon, Floccia, Fux, Adda-Decker, & Nazzi, 2014; Mandel, Jusczyk, & Pisoni, 1995). Thus, speech sound learning appears unnecessary for the beginnings of word learning, damaging the case for strictly hierarchical acquisition of language structure.

Perceptual precocity not sufficient

Just as infant perceptual narrowing and increasing native sensitivity to speech sounds may not be necessary for word learning, it appears not to be sufficient either. Simply put, learning to tell apart similar-sounding words is hard, and stays hard well after infants show qualitatively nativelike sensitivity to speech sounds. This insufficiency is evident on a small scale in the 14-month-olds studied by Stager and Werker (Stager & Werker, 1997), but is on full display in my own research on preschool-aged children.

My own research: perceptual immaturity persists into early childhood

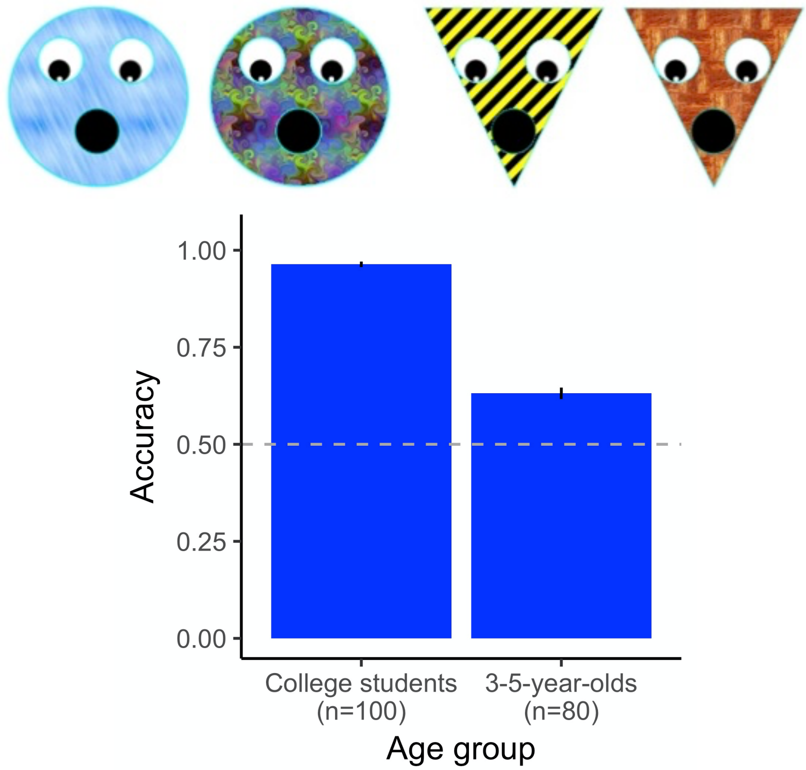
Young children are not great at learning words

The short story of my developmental research is that young children are still very much under construction, perceptually speaking. The longer story is that I've spent a great deal of time and energy considering Keen's (2003) question of why toddlers (and preschoolers) look so dumb relative to infants (see also Cowan, 2016, 2017; Haith, 1990; for consideration of these issues). I offer my keen insights here.

Maybe it's worth noting that my initial realization that young children are still undergoing perceptual learning stemmed from my findings in other auditory domains: talker recognition, accent detection, music perception. I hadn't dared explore this phenomenon in the word learning domain, because I had uncritically accepted the idea that children are brilliant word learners. I even dismissed my own lab's findings of weak word learning in children relative to young adults as operator error: we as experimenters didn't give them just the right words in just the right way.

An early study, and what's wrong with it

In one of the first studies my lab ran at UC San Diego, we tested word learning in 3- to 5-year-old children and a control sample of college-aged adults. Children and young adults learned four to eight novel words (e.g. vaymo, zyghel) for four to eight novel facelike pictures (Figure 1). They were then tested with two alternatives at a time: they saw two faces, and one was named. They were asked to point to the named face. Children were 63% correct (Figure 2). Adults were 96% correct. This difference was highly statistically significant.

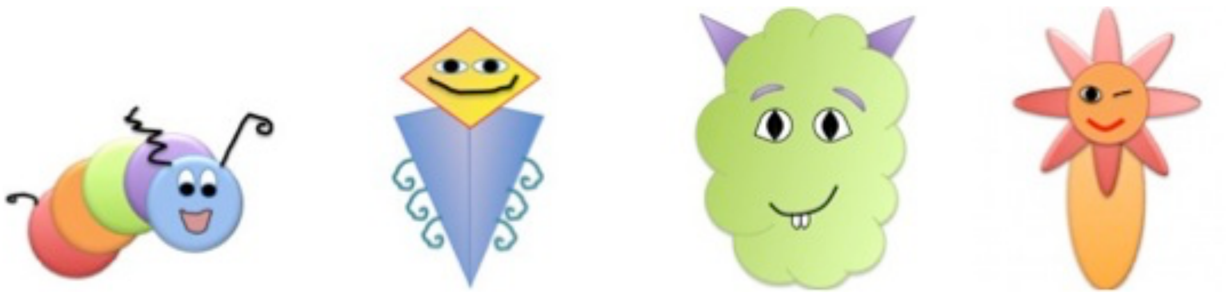


One might take this as evidence that adults are better word learners. However, immediately one could come up with things that are wrong with this study: the visual referents (faces) were unfamiliar to children, and too similar to each other. Some words were low in phonotactic probability, which makes words harder to learn (Graf Estes, Edwards, & Saffran, 2011). Maybe learning 8 words at once is too much (though it wasn't too much for the college students). Perhaps some children didn't understand what we were asking them to do. For any of these reasons, it could look like children are not great at learning words when in fact they are excellent at learning words. Of course, one might think (and I might agree with one) that positing all of these exceptional conditions is to grade children on a "curve" that would not apply to adults. More on that later.

Learning similar-sounding words

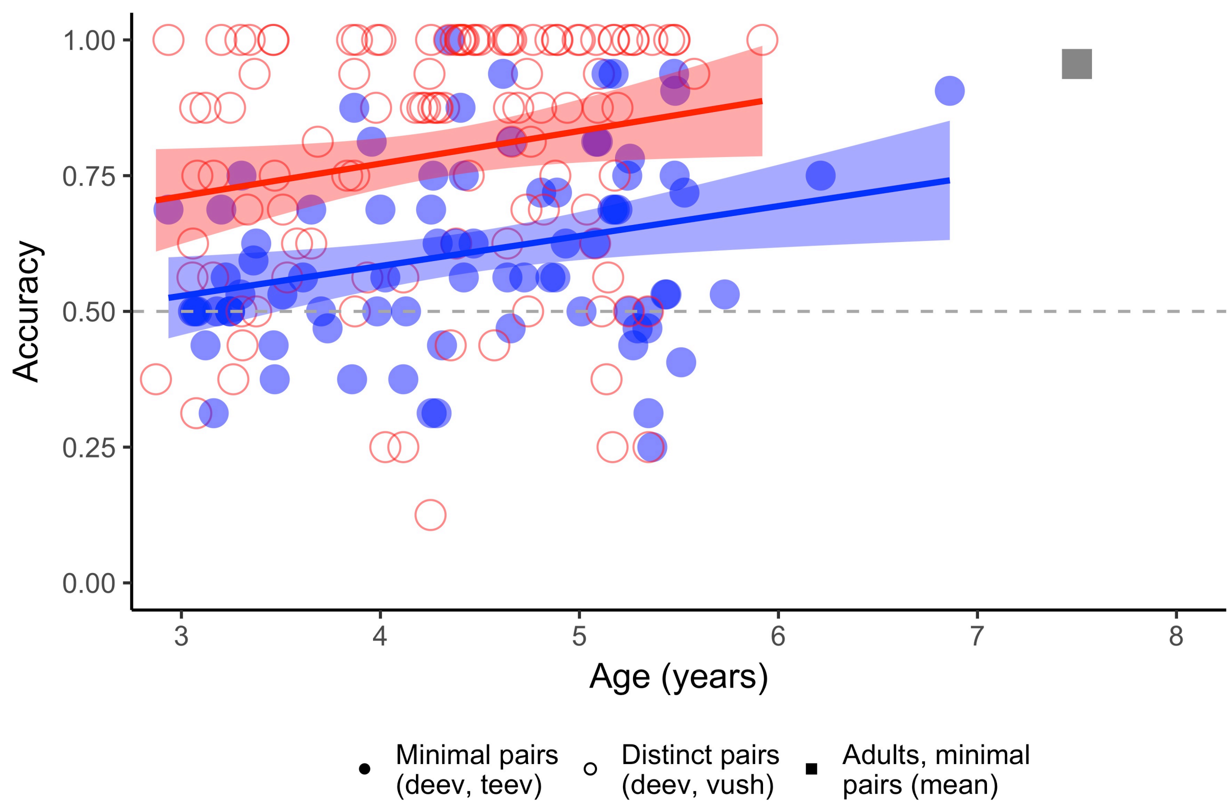
More recently, with graduate student Conor Frye, I have tested young children's ability to learn extremely similar words, also known as minimal pairs, such as deev and teev. Such words differ by speech sounds that are distinguished by a single phonological feature. In some sense this is a direct follow-on to Stager and Werker's studies (Stager & Werker, 1997; Werker et al., 2002), where 17-month-olds but not 14-month-olds showed evidence of learning minimal-pair words (bih and dih). Notice that I am specifically testing whether they have learned the sound forms well enough to tell deev and teev apart; they do much better when we ask them to tell apart deev and vush, or fush and teev.

One especially nice feature of the study is that we assessed eight different minimal contrasts across children: four consonant voicing contrasts (voicing is a single phonological feature; b/p, d/t, f/v, s/z), and four vowel contrasts (i/ɪ, eɪ/ɛ, ɑ/ʌ, u/ʊ). Thus, these findings should generalize past a single speech sound contrast. Mindful of all of the reasons that children might have difficulty learning words, we presented only two words at a time for learning and testing, and engaged children in a gamelike setting where learning occurred as a series of short animations in which cartoon creatures (Figure 3) were named. We also tested a group of adult controls. Adults were tested mainly as a check that our stimuli were sufficiently discriminable to allow learning, but also allowed assessment of developmental change at a coarse scale.



Despite our efforts at child-ifying the task, the effects were the same as in earlier work: children were statistically above chance (59%) but adult learners were far more accurate (95%). Another study from Frye's dissertation research suggested that children perform better (78%) when asked to learn dissimilar-sounding words for the same cartoon characters. This suggests that perceptual discriminability is a major contributor to their learning difficulty.

Still, mindful of all the reasons that children might have difficulty learning words, I replicated the study. The study I just described, since it adapted materials from an early study with adults, used adult-directed speech with isolated words ("Deev. Deev."); the replication used child-directed speech embedded in sentences ("Do you see the deev over there? Isn't it a nice deev?"). I also included a within-subjects control condition with dissimilar-sounding words, to make sure that children who performed poorly on minimal-pair word learning were grasping the task itself. The outcome was strikingly similar to Frye's dissertation research: children were less accurate on minimal-pair words (61%) than on dissimilar words (78%, though both were above chance). While children's below-ceiling accuracy on dissimilar words suggests that additional factors constrain children's performance, it is clear that perceptual similarity has an effect. There appears to be little benefit to sentence frames or child-directed speech. However, across all of these studies, both perceptual similarity and age predicted pointing accuracy (Figure 4). We also collected visual fixation data, but as I will describe later, this correlated strongly with pointing accuracy, so I am leaving it out for now.



A further follow-up attempted to match learning conditions more closely to the Switch procedure, in which 17-month-olds show learning of minimal pairs. I tested our youngest age group—3-year-olds—in a paradigm closely modeled on Yoshida et al. (2009), who showed learning in 14-month-olds using an eye tracking measure after Switch-style training. Each word was heard 64 times (vs. only 16 times in our original study). However, this did not increase word learning relative to the first two experiments: novel-word accuracy was 55%, which did not exceed chance. This did not seem to result from inattention or poor discrimination skill, in that a familiar minimal pair post-test (doll/ball and boot/boat) showed high accuracy (87%). It's not clear why 3-year-olds might fail while 17-month-olds succeeded. One possibility is that factors that optimize for 17-month-olds' attention are not optimal for 3-year-olds. Another consideration is that my lab tested 3-year-olds in a preschool environment, which is likely more distracting than the lab settings where infants are typically tested. Still, preschools are, one hopes, the site of much 3-year-old word learning, so it is interesting that word learning does not appear to be very successful here.

Accompanying studies (Creel, in preparation) have examined whether word learning is a particularly difficult way of assessing sensitivity. Instead of teaching words as labels, I present minimal-pair words back to back in a same-different discrimination task. As a group, children are sensitive to minimal pair differences in consonant voicing (deev vs. teev) and vowel quality (deev vs. div), in that their discrimination scores are well above chance (d' exceeds zero). However, they are less sensitive to minimal pair differences than to whole-word differences (deev vs. vush). Further, it's tough to know how to compare discrimination accuracy to word-learning accuracy, and children under age 4 years often have difficulty with same-different discrimination, a point I return to later.

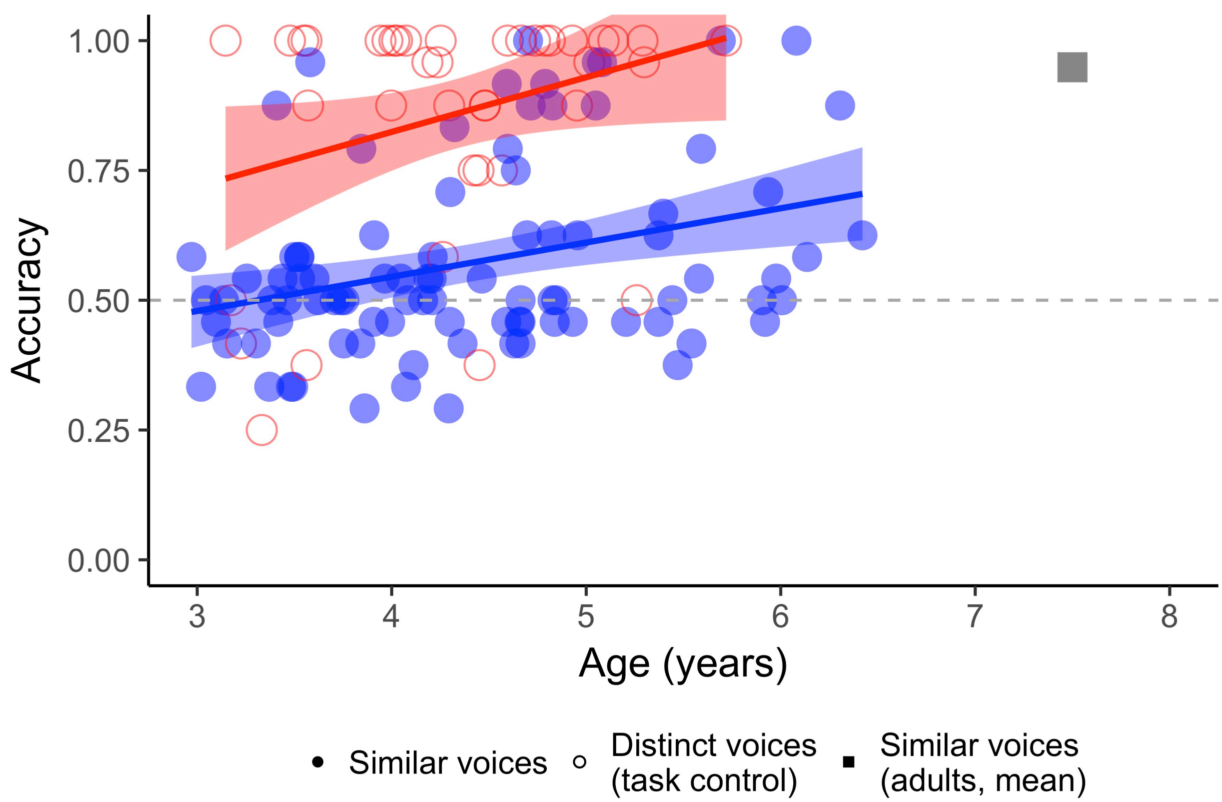
In any case, children's weak learning of minimal pair words seems contradictory to research using the Switch procedure, which suggests minimal-pair word learning by 17 months. Yet, other researchers (Barton, 1976; Garnica, 1971) have previously reported that toddlers have difficulty distinguishing minimal pairs, especially newly-taught ones, in word-recognition contexts. I argue that differences in methodology between my word-learning studies and Switch habituation studies of word learning are sufficient to explain the difference in outcome. In any case, I was relieved to find this data pattern because it fit much better with results I had obtained in other domains: children in this age group are far from adultlike in learning to recognize voices or discern pitch differences in melodies. I review these studies below to illustrate that there is additional evidence for protracted perceptual learning of auditory pattern structure.

Young children also have difficulty learning voices

My student Sofia Jiménez and I were interested in how children learn to recognize voices. Voice recognition is interesting in that it is accomplished using the same vocal signal as speech recognition, but presumably using a different (but probably overlapping) set of auditory attributes. It also helps us discern why we don't attach meaning to voice variation: why are "cat" and "cap" in mom's voice two different words, but "cat" in mom's voice and "cat" in dad's voice are the same word, despite sounding quite different? Is the irrelevance of voice variation hard-wired, or learned? Numerous studies suggest that infants can distinguish voices (e.g. DeCasper & Fifer, 1980; Kisilevsky et al., 2003), especially if the voices are speaking the language in which infants are immersed (Johnson, Westrek, Nazzi, & Cutler, 2011). Further, 7.5-month-olds have difficulty generalizing a newly-learned word from one voice to another dissimilar voice (Houston & Jusczyk, 2000), suggesting that voice information is encoded early in development (see also Creel, 2014a; Creel, Aslin, & Tanenhaus, 2008; Creel & Tumlin, 2011, for evidence of voice-sensitivity in word learning for children and adults).

Limited previous research suggests reasonably good voice identification of classmates and cartoon characters by children around 4 years of age (Bartholomeus, 1973; Spence, Rollins, & Jerger, 2002; van Heugten, Volkova, Trehub, & Schellenberg, 2013). However, these studies used voices familiar to the children, and cartoon voices in particular may have distinctive voice qualities not encountered in everyday life. Other research, particularly a study by Mann, Diamond, and Carey (1979), suggested very poor voice differentiation in children as old as 6 years, with slow improvement through adulthood. Yet Mann et al.'s (1979) study used a multi-alternative listening task that was challenging even to adult participants. Thus, it is not clear whether young children are good or only middling at learning voices like those they typically experience.

In our study (Creel & Jimenez, 2012), we asked participants to learn new voice-individual associations, controlling for amount of exposure to particular voices across participants of varying ages. The individuals were, in most cases, cartoon characters I created (see Figure 3). We recorded a large number of young adult female speakers from the United States and selected two who had distinct formant structure (Experiments 1-3), and two who had distinct prosody (Experiment 5). In the latter case, one speaker sounded quite animated or child-directed, with large pitch contours across the course of each sentence, while the other sounded more monotone, with minor pitch excursions across each sentence. Children ages 3 to 6 years received several learning trials where each character appeared on screen and "spoke." After this, they received several test trials where they saw the two characters side by side, one of them "spoke" (without any movement of either character), and the child was asked to point to the one speaking. We also eye tracked children during the test phase, but those results were highly similar to the pointing responses, so I omit them for simplicity. In all cases where the speakers were two young-adult females, children showed low (though above chance) accuracy, around 60%, at associating voices with characters. Performance improved with age (Figure 5, circles). Adults hearing the same talkers exceeded 90% accuracy.

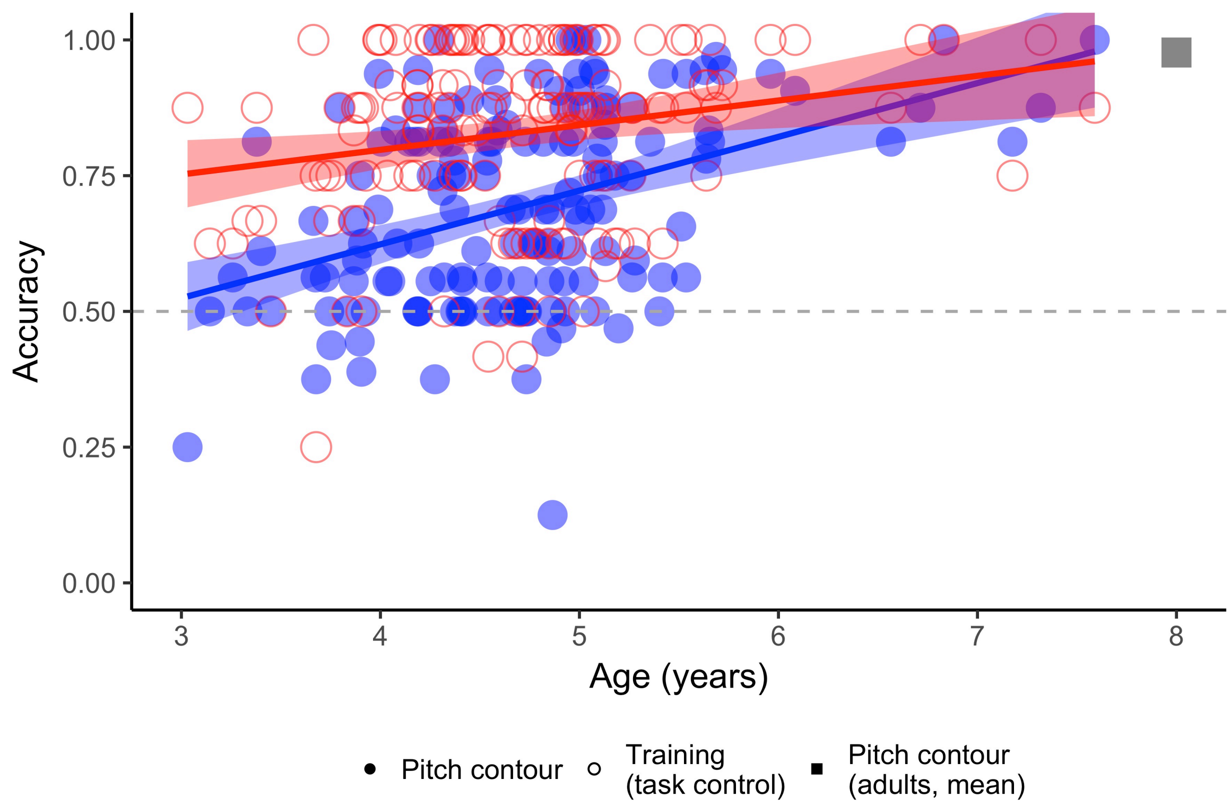


To confirm that children were not simply failing to grasp the task, we also tested male-female speaker pairs (Experiment 2) and adult-child speaker pairs (Experiment 6). Children were far more accurate in these cases (80%-90%, Figure 5, hollow circles). This suggests they understood the nature of the task, but had difficulty perceptually distinguishing the voices. It's also possible that they did not socially distinguish the voices despite excellent perception. However, I later conducted a study (unpublished) where we presented pairs of voices in a same-different discrimination task. Each voice pair in a trial spoke the same brief sentence. Children's ability to perceptually distinguish female-female voice pairs was well above false-alarm baseline (49% "different" responses, with 10% false alarms on same trials). Still, this was much lower than their ability to perceptually distinguish male-female pairs (84% "different" responses), consistent with a perceptual basis for their difficulty in learning voice-character associations. It also suggests that discrimination ability is not a perfect predictor of learnability, much as we saw with minimally-different words.

Pitch processing tunes up slowly

Much as with research on infant word learning and infant voice sensitivity, infants are reputed to be expert at processing pitch contours (Trehub et al., 1984; Trehub, Thorpe, & Morrongiello, 1985). I was thus initially surprised to find that children have difficulty processing pitch patterns (Creel, 2014b, 2016, 2019; Creel et al., 2018). Adult participants find this relatively easy: while your average undergraduate might not notice that a note is mistuned by a quarter of a semitone, they have little difficulty identifying changes in pitch direction (see, e.g., Halpern & Bartlett, 2010). However, young children that I tested had great difficulty learning that one particular individual (a cartoon character) liked a rising melody, while the other liked a falling melody (Creel, 2014). And when I say great difficulty, I mean that they were at 50% accuracy in a two-alternative test. This is especially striking in that these are the two most-distinguishable pitch contours possible—one consisting of multiple rising intervals and the other of multiple falling intervals. I tried different instructions, and much wider pitch changes, and learning accuracy didn't budge (see Creel, 2016). Again, it did not seem to be difficulty with the task itself, because children were above chance when the melodies differed in gross pitch height (one was an octave higher than the other—a pitch doubling; Creel, 2014b), and were above chance when melodies differed in timbre (e.g. a rising melody played on a vibraphone and a rising melody played on a trumpet; Creel, 2014b). It really seemed to be about pitch itself.

I followed up these results with discrimination studies where I played pairs of melodies back to back and asked children if they were the same sound or different sounds (Creel, 2014, 2016). Children were well above floor in distinguishing melodies differing in pitch contour, but they were better at distinguishing melodies differing in timbre. This pattern replicated in a study where I asked each child both to discriminate pitch contours, and then to learn them as the favorite songs of cartoon characters (Creel, 2016). The same children who distinguished rising and falling melodies with 82% accuracy were not able to associate those melodies with different characters (48% accuracy, where 50% is chance), and good discrimination did not predict good learning (r = .04). Across a set of six melody-discrimination experiments from my lab (as yet unpublished), children's pitch contour performance improved with age (Figure 6), but accuracy only began to approach adult performance around age 7.



An exception to this poor pitch processing seems to be familiar melodies. In a recent study (Creel, 2019), I presented children with pairs of familiar melodies or pairs of matched unfamiliar melodies. I found that children are better at discriminating two familiar melodies than two unfamiliar melodies. They are also better at learning associations with familiar melodies than with unfamiliar melodies, if the associations are semantically supported. These effects, while not huge, are analogous to findings from the word recognition literature that infants readily recognize and differentiate familiar minimal pair words like doll and ball (Fennell & Werker, 2003), even though same-age infants have difficulty learning new words (Stager & Werker, 1997). The fact that I find the same pattern in much-older children suggests that familiarity with specific sound patterns shapes association learning, rather than or in addition to chronological age. (See also Pająk, Creel, & Levy, 2016, for a related pattern in adult second-language learners.)

As with the word learning findings, these results appear to contradict early learning results from the infant literature. As many Reviewers 2 have written to me, we all know that even infants can detect melodic differences. But when one digs deeper, one finds widespread evidence of less-than-amazing learning or perception results in the preschool age range. Other researchers have also reported weak pitch contour processing in young children (Fancourt et al., 2013; Stalinski et al., 2008).

Children's generally weak processing of pitch information also fits with findings from pitch processing in language. In particular, vocal affect (some of which is expressed by pitch height and pitch variability) and question intonation (differences in pitch direction) are relatively late to emerge developmentally. Specifically, children have difficulty identifying vocal cues to a speaker's affect (Friend & Bryant, 2000; Morton & Trehub, 2001; Nelson & Russell, 2011), including a study by Quam and Swingley (2012) where pitch cues were the focus of study. Some researchers suggest that this is due to interference from verbal content (Morton & Trehub, 2001), such as hearing "I got ice cream for being good" in a sad voice and reporting that it sounds happy. In other papers with less biasing verbal content, children have difficulty differentiating happy from sad utterances based on prosody alone (Quam & Swingley, 2012; though see Berman, Chambers, & Graham, 2010 for a different perspective). Further work on question intonation ("That's a dog." vs. "That's a dog?") suggests that children as old as age 8 experience difficulty using pitch to distinguish questions from statements (Saindon, Trehub, Schellenberg, & Van Lieshout, 2016, 2017). Finally, Ito, Bibyk, Wagner, and Speer (2014) find that children slowly increase in their ability to use contrastive pitch accent cues, which may be related to their pitch processing. All of these findings together suggest that children's pitch representations improve slowly with age, which may place an upper limit on children's abilities to apprehend some pragmatic aspects of language.

Auditory pattern learning: summary

Across multiple studies in multiple auditory domains—speech, voice, and musical pitch—I have found evidence consistent with nonadultlike perceptual sensitivity well past infancy. Other researchers have found similar patterns in all of these domains. I have argued (Creel, 2018; Creel & Quam, 2015) that these findings are consistent with long-term perceptual learning. However, there is a raft of infant studies that seem to contradict this by revealing qualitatively adultlike perceptual sensitivity. That is, we need an explanation of why infants look so smart and preschoolers look so dumb (Keen, 2003). My explanation is that it's due to vast methodological differences: I assert that the totality of the data are consistent with protracted perceptual learning for speech and non-speech sound patterns, obfuscated by massive methodological discontinuities (see related arguments in Creel & Quam, 2015). These methodological discontinuities are partly driven by affordances of the subject population. So it isn't infants that are so smart—it's infant researchers that are so smart as to have optimized paradigms to coax out evidence of nascent knowledge. There is not an easy way around these methodological divides, but I offer some possibilities later on.

Alternative explanations for age-related changes in performance

Before suggesting bridges across methodological divides, I want to discuss some non-perceptual explanations for the results I've observed. A lot of these boil down to competence/performance distinctions—the notion that children look like they're bad at sound pattern learning for extraneous reasons, when actually they're quite good at it. One just has to find the perfect task, remove an extraneous roadblock, ask the question in the right way.

Lack of metalinguistic awareness: an accent waiting to happen

One counterexplanation—a reason why children might be great at word learning or perception but we have failed to uncover it—is that they simply don't understand what we are asking them to respond to. There is a lot of truth to this. In my lab, my researchers once returned from a preschool where they asked children if they knew what an accent was. Of the few children who said that yes, they did know, one of them described an accent as "when your car goes off the road."

So it is definitely the case that children do not have the metalinguistic equipment that adults do. Still, there are a number of ways to make up for this. One is using child-friendly language ("sounds funny" instead of "has an accent," though even then you have to make sure that they don't think you mean "tells good jokes" or "is being silly"). Another is to include a training phase with easy correct answers—and then filtering out children who did not pass the training phase as those who did not understand the task. Similarly, the researcher can include easy stimuli throughout a test phase, or as one of multiple conditions, to check for continued task compliance and assess performance where perceptual difficulty is lessened, again filtering out children who do not register the "check" stimuli.

Attentional differences

Another reasonable counterexplanation that doesn't depend on perceptual learning is that my findings are ascribable to attentional differences. Maybe children "space out" every few trials, causing them to answer incorrectly (in a discrimination study) or to fail to encode enough learning instances (in a learning study). On this account, it is attention that improves with age, not perceptual representations. Thus, while perceptual representations become language-specific and perhaps sharpen in infancy, poor performance past this point reflects high competence but weak attention.

To an extent, this is an insoluble problem: by most measures, children's attention improves from infancy to childhood through young adulthood. Thus, any improvement with age is confounded with attentional improvements. (It's also confounded with massive changes in size, mobility, vocalizations, motor skill, and teenage sarcasm.) However, it is a confound not only in the performance of infants vs. young children, but also in the performance of younger infants vs. older infants. Thus, invoking attention and not perception as the explanation for children's weak performance relative to adults invites a reinterpretation of presumed perceptual learning effects in infants. For example, perhaps perceptual representations, or perceptual acuity, do not change at all between 4 months and 8 months, but attention improves. Further, this leaves unresolved the question of whether attention itself is partly an epiphenomenon of the level of resolution of one's perceptual representations: perhaps attention can only be diverted to a novel stimulus once perceptual resolution is sufficiently keen to detect it.

Nonetheless, there are a number of ways one might assess the attention hypothesis. Some of the performance-filtering measures described in the previous section might be useful. For learning studies specifically, one can measure on-task-ness by using eye tracking during the learning phases to gauge degree of visual attention to the display. A peevish reviewer might then argue that children are attending more to visual aspects of the display at the expense of cooccurring speech. Of course, this argument would question the general usefulness of looks as a measure of infants' language processing, potentially unraveling visual habituation and conditioned-head turn studies.

Another possibility is to manipulate interestingness or learning dosage and assess whether this modulates performance. More interesting things presumably increase children's attention, which should lead to improved learning. Recall the minimal-pair word learning study I described earlier where the first version was run with a male speaker producing adult-directed isolated words, and the replication was run with a (presumably more interesting) female speaker producing child-directed words in sentences. Recall also that there was no difference between the outcomes—children performed no better at learning minimal pair words when child-directed speech was used. This is not very consistent with an account that increasing attention should increase performance. A related way to address the "attentional lapses" hypothesis is to present more learning opportunities. If children perform worse than expected because they are randomly inattentive on some subset of trials, then presenting more learning trials should increase the number of processed learning trials and thus boost performance. Preliminary work in my lab suggests this is not the case: doubling learning trials does not increase performance.

A third possibility is to limit analyses to only those children who score at ceiling on some control measure. If a child responds reliably at endpoint stimuli, or in a relatively perceptually easy condition, but is less sensitive to more subtle variation, this is consistent with perceptual difficulty rather than attentional fluctuation. Hazan and Barrett (2000) did this with speech sound identification data, filtering out children who did not respond deterministically on endpoint stimuli. They found that even those children showed softer categorization functions than adults. Still, one is entitled to wonder whether such an approach invites regression to the mean. That is, perhaps the children who scored high on endpoint stimuli did so partly by chance, and thus selecting them on the basis of endpoint stimuli virtually guarantees that other stimuli will show less deterministic performance. For this reason, one's measure of good attention should be checked for stability.

A final possibility is to provide continual reminders to children about what they are being asked to do. This might include restating the decision rule on each trial ("Which speaker do you want to have as your friend?") or providing corrective feedback ("No, those two words were different."). My lab and I have attempted to improve same-different discrimination performance in new work. Constant corrective feedback leads to slightly higher d-prime scores overall ($p<.01$) but children still make many more errors on similar-sounding trials (deev, teev) than on different-sounding trials (deev, vush).

Possible approaches to studying protracted perceptual learning

The mystery of the missing studies

Here I want to discuss ways that we might track changes in perceptual learning across development. A major challenge is that there are very few tasks that can be used across such a wide age range, though some come closer than others. Some tasks appear to have promise for spanning age ranges, yet this promise appears largely unrealized, with almost no literature on the subject. I suspect that the absence of literature may represent a Bermuda Triangle of studies that were attempted but never published. That is, I would think that numerous researchers must have tried (say) dishabituation in 2-year-olds or 4-year-olds or adults, but to my knowledge no such studies exist.

I would also think that a paradigm like conditioned head-turn might work well in children who are past infancy, but there are almost no papers using anything like conditioned head-turn in young children. The few exceptions to these gaps are illuminating. For example, Sundara et al. (2006) succeeded in getting 4-year-olds to do something like a conditioned head-turn task, but noted that this only worked when they asked children to monitor for a specific word (press the button every time you hear "boat"). Similarly, Trainor and Trehub (1994) tested 5-year-olds in a melody change detection task, after training them on "same" and "different" responding. In both cases, these investigators had previously tested infants (< 12 months), thus skipping over 1-year-old, 2-year-old, and 3-year-old children. This makes one wonder if they attempted testing with younger children and it was not successful. Further, one issue with interpreting this as homologous to conditioned head-turn is that word-monitoring requires attention, whereas dishabituation or a measure such as the mismatch negativity event-related potential may occur more passively as a result of detecting a change in the auditory environment (see Cowan, 2016 for discussion of working memory in infant tasks vs. those used with children and adults).

The measures that come closer to spanning wide age ranges include visual-world eye tracking and event-related electroencephalography or magnetoencephalography. A final possibility involves connecting related behavioral measures across ages.

Visual world paradigm

The visual world eye tracking paradigm has illuminated the world of infant language comprehension perhaps more than any other technique in the past two decades. As discussed earlier, eye tracking allows access into comprehension even in individuals who cannot verbally respond or even manually respond (e.g. pointing). This research suggests that children have formed some associations between word forms and word meanings as early as 6 months of age (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999, 2012), and are sensitive to deviations from typical pronunciation by 14 months or so (Swingley & Aslin, 2002). Some studies have attempted to link visual world responsiveness to language development later in childhood (Fernald & Marchman, 2012; Fernald, Perfors, & Marchman, 2006; Marchman & Fernald, 2008), though none of these tracked visual world measures past about two years of age.

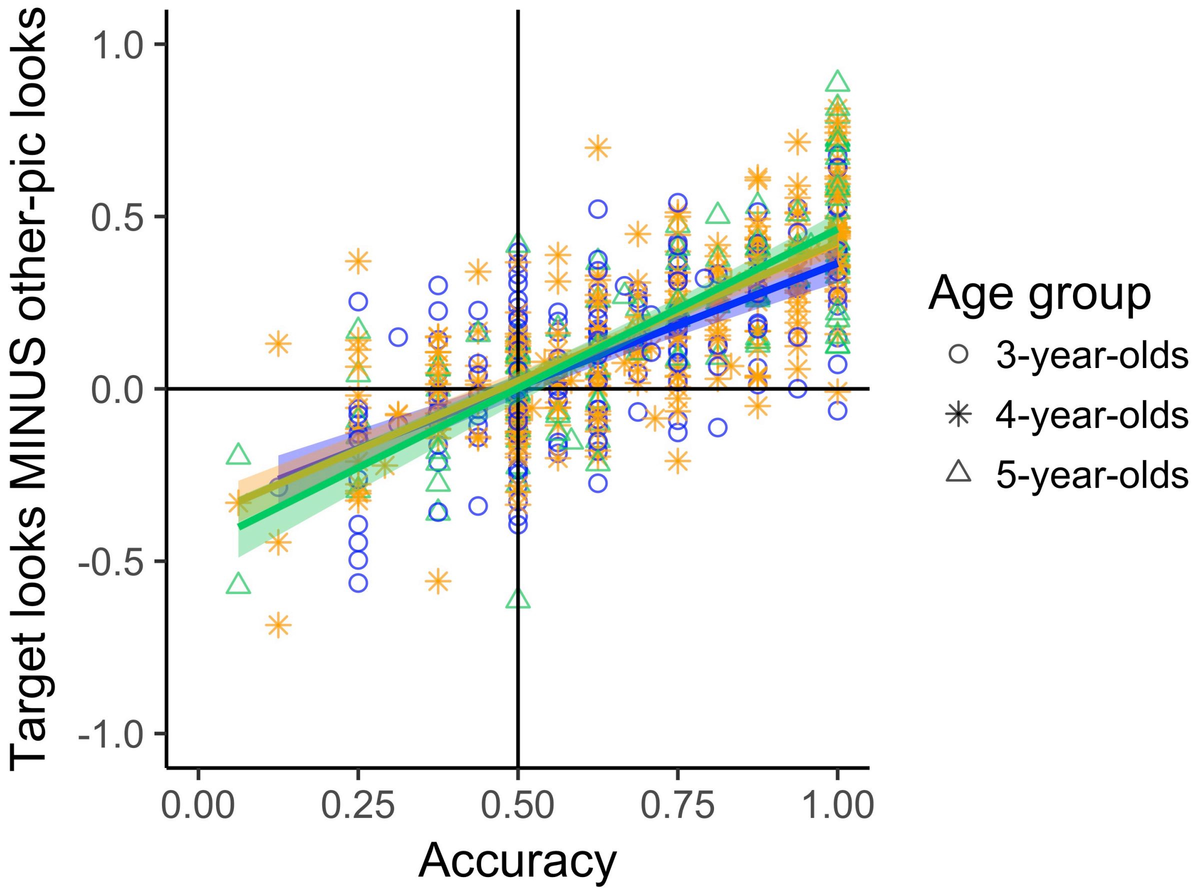
Nonetheless, challenges in using the visual world paradigm abound. First, it's not clear how much children's slower looks represent slower language processing vs. slower eye movement dynamics, slower object recognition, some more general cognitive feature, or even a greater tendency in children to explore the display rather than doing the task the experimenters intend. One hint that it's not (just) about exploring the display is that, if it were, there should be similarly-inflated looks across all non-target pictures, regardless of phonological or semantic similarity to the target, and unrelated to the speech signal. That is, if children are simply doing more visual exploration overall rather than processing language more slowly, they should look more at a cap when hearing "cat", but should also look more at a shoe and a bug. However, the observed pattern is that cap looks are higher than shoe and bug looks for a lengthy period of time (Sekerina & Brooks, 2007), which is more consistent with extended phonological competitor activation than with off-task looking around. Another hint that it may be about speed of language processing specifically, rather than developmental factors, is that young adults learning novel vocabulary also show slower looks to named pictures (e.g. Creel, Aslin, & Tanenhaus, 2008; Magnuson, Tanenhaus, Aslin, & Dahan, 2003). Thus, weaker knowledge of to-be-recognized words can lead to the patterns we see in children. Still, it would be reassuring to see studies of children's saccades in response to auditory signals when word recognition is out of the picture. Studies of children's developing eye movements suggest early maturation but tend not to examine saccades (eye movements) based on an external signal.

Another challenge in interpreting visual world paradigm results is that, if familiar words are used, those words are likely to be much more familiar the older the participant is. For example, a 2-year-old has probably heard the word "cat" twice as often as a 1-year-old, but half as often as a 4-year-old. Thus, it's possible that slower visual fixations to named pictures in younger children are driven by weaker knowledge of specific words. One might in theory control for exposure to particular words by teaching novel words to individuals at a range of ages. In cases where my lab has attempted this, the result has been much lower accuracy in children than in adults, and eye movements tend to follow suit. It would be especially interesting to study learning over a longer term to find out whether younger learners need larger "doses" (exposures per word), and whether, given equivalent doses, younger learners are any more sensitive than older learners to nonnative sound patterns.

What's the point of looking measures?

Discussion of visual-world language processing brings up a related question: how does looking proportion in infancy map onto overt responses in children and adults? Numerous researchers have found increases in named-picture looks in older vs. younger infants (Bergelson & Swingley, 2012; Fernald et al., 1998) as well as young children vs. older children vs. adults (Borovsky & Creel, 2014; McMurray et al., 2018; Rigler et al., 2015; Sekerina & Brooks, 2007). The reader may have noticed that I have collected visual-world data on word learning, voice learning, and melody learning, but have only reported the pointing accuracy data in this chapter. One reason for that is that looking times and points tend to lead to similar conclusions in all of these studies, and it's simpler to explain pointing accuracy.

There is a sometimes-implicit assumption that eye movements to pictures reflect more implicit knowledge than does an overt measure like pointing accuracy. But is this assumption correct? One way to look at this is to compare an overt measure to looking time in the same individuals. I did this over a large number of children who have taken part in word-learning studies in my lab over the years. This includes 286 children and 567 word pairs learned (some children learned multiple word pairs). There is a strong relationship ($r = .73$) between pointing accuracy and the proportion of looks toward the correct (target) picture vs. the non-target picture (in the time window 200-2000 milliseconds after word onset), suggesting that pointing accuracy and visual fixations may be measuring the same construct. Further, it does not appear that visual fixations are more sensitive to knowledge than pointing. If visual fixations reflected implicit knowledge, then there should be above-chance looks to targets even when pointing accuracy is at chance. This is very much not the case: as evident in Figure 7, when accuracy is at chance (50%), looking proportions are also at chance (target minus other looks = 0). It remains to be seen whether this holds for even younger children and for other situations. It is likely to break down in cases where accuracy is at ceiling (or floor), and some of my own most interesting findings from eye tracking are in cases where accuracy is high and indistinguishable. Still, it would be interesting to know how well this relationship between implicit and explicit measures holds up in even younger children. While a 6-month-old definitely can't point, a 2-year-old can. This would perhaps give us some hints about what looking measures mean in children so young that they can't point.



Event-related potentials
Event-related potentials, ERPs, as well as their magnetic analogs, may hold promise for measuring continuous change across age. Some researchers have leveraged ERPs to examine change in auditory sensory memory with age. A particular ERP of interest is the mismatch negativity (MMN), which is elicited across a wide range of ages when a deviant occurs after a series of standards (see Näätänen, Paavilainen, Rinne, & Alho, 2007 for a review). What is especially appealing about the MMN is that attention is not required to elicit it. This means that one can test children as young as infancy without having them do any sort of task. Gomes et al. (1999; see also Keller & Cowan, 1994) explored growth of auditory sensory memory with age by mapping out the maximum delay between standard and deviant that still generated an MMN. Children aged 11-12 years and adults showed MMNs at 8-second delays (see also Sams, Hari, Rif, & Knuutila, 1993 for MEG evidence), but only at 1-second delays for children from 6-10 years of age. Related work by Cheour et al. (2002) explored the duration of auditory sensory memory in newborns and found the duration of auditory sensory memory to be around one second or less. These studies taken together suggest an increase in the duration of auditory sensory memory from infancy through the early teenage years.

One challenge to ERP approaches is uncertainty about whether mismatch waveforms at different ages reflect the same underlying brain activity (Cheour et al., 2000). Another challenge is matching the “entertainment” across ages: a movie that will keep a 6-month-old amused and relatively still while listening to a lengthy series of unattended sounds may not have the same effect on 4-year-olds or adults.

Connecting the tots: Tracking developmental change using overlapping related tasks

Another part of the answer may be chaining together a series of tasks that overlap age ranges, and creating new tasks that are modified to be more like those use at an adjoining age. Then one can test different tasks at the same age to determine rough equivalences across tasks. For example, if a child is at ceiling accuracy in a same-different task, how well do they do in a word-learning task that requires storage in long-term memory?

One persistent feature of tasks that work well with infants is repetition of a standard, followed by a deviation from that standard (habituation; conditioned head turn; mismatch negativity). This repetition might make them bored. It also might build up a stronger representation of the standard in short-term memory, allowing for finer-grained detection of a deviant. The representation-strengthening role of repetition is evident in multiple theoretical accounts of adults' perceptual processing (Cowan, Winkler, Teder, & Näätänen, 1993; Davelaar, Tian, Weidemann, & Huber, 2011; Viemeister & Wakefield, 1991).

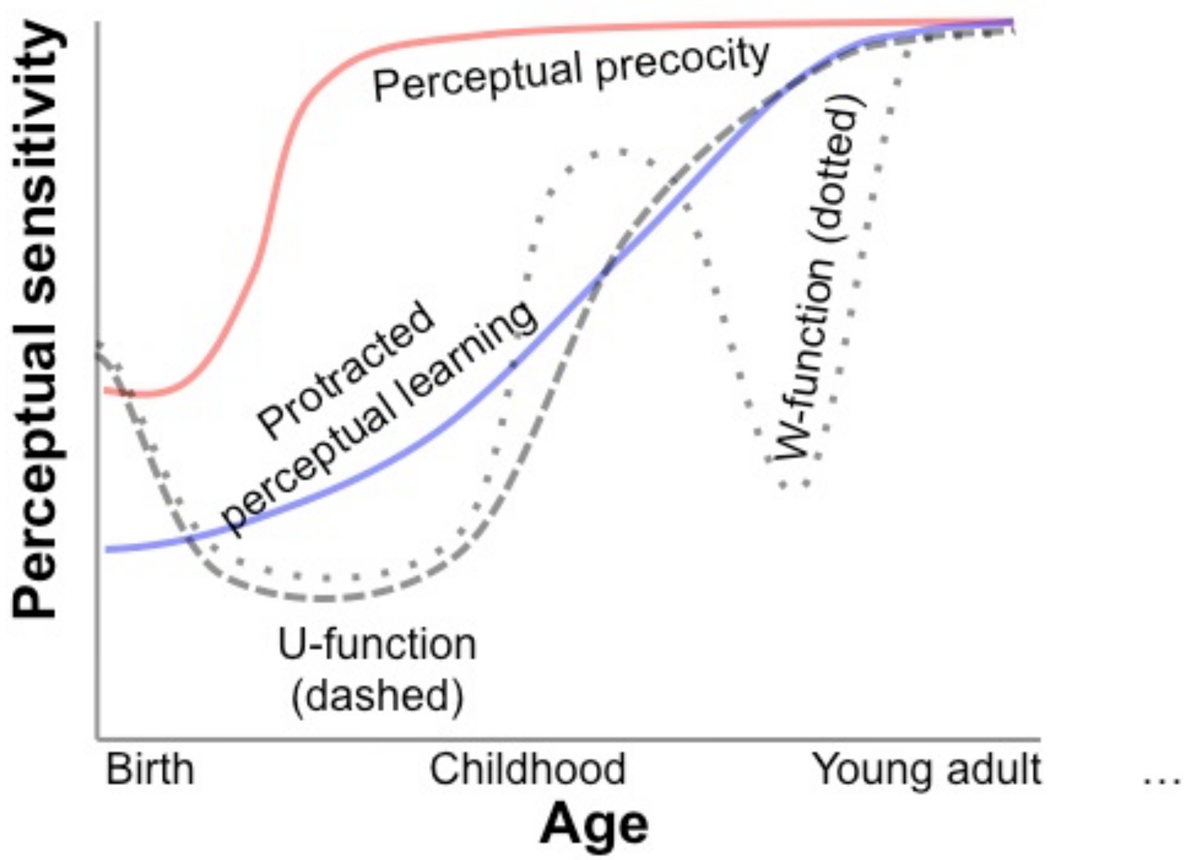
It is tempting to think that a habituation task or a deviant-detection task might connect well to same-different discrimination tasks. Both involve comparing one sound pattern to a second sound pattern in auditory working memory, either implicitly or explicitly. However, the explicit nature of same-different seems to make the task especially challenging for young children. In one unpublished set of same-different music perception studies from our lab, only about 40% of children between 3.0 and 4.0 years of age succeed on the highly-discriminable training trials, even when they initially receive visual same-different training trials and accuracy feedback. I don't know why this is difficult for younger children, but I think at least part of it is the fact that they need to understand what the terms same and different mean, and this may be distinct from detecting a discontinuity in a stream of sounds.

Further, auditory stimuli are fleeting: unlike visual stimuli, you can't look back and forth between two sounds. You must hold one in memory while listening to the second one. Thinking back to the findings from habituation, conditioned head turn, and MMN studies, one wonders if children have difficulty in this task partly because they have trouble holding on to an auditory representation. In this sense, they might have an easier time of it if the auditory representations were stronger. Thus, one possible overlapping task might be a blend between habituation-style tasks of infants and same-different tasks of young children and adults. This can be done by repeating the to-be-judged auditory stimuli, within trials ("deev deev … teev teev") or across trials ("deev deev… deev teev") or both. Holt and Lalonde (2012) manage to coax older 2-year-olds into above-floor performance in speech sound discrimination using a task with repeated standards (see also Holt & Carney, 2007), though the procedure reportedly took 2-3 hours per child, which might be impractical for some researchers. Still, those researchers' methods may suggest ways to bridge, or at least narrow, the developmental data gap.

Open questions

Gaps in the data

As I outlined earlier, there are major age gaps in the developmental record of perceptual learning in speech processing. In particular, there are very few studies between infancy and early childhood. This means that we don't know what the full developmental picture is for word learning, speech sound discrimination, voice encoding, and same-different discrimination of melodies. Setting aside arguments about where such developmental trajectories come from—are they specific to speech perception or language, or reflective of more general cognitive changes—we don't even know what the trajectory looks like. It could be a rapid asymptote (perceptual precocity), a slow asymptote (protracted perceptual development), even a U-function as implied by a superficial interpretation of extant infant and child research (Figure 8). Thus, uncovering the relevant developmental functions is a necessary first step toward some of the open questions discussed below.



Structure of the input

There are several areas of interest in considering children's perceptual learning and word learning over time. One unresolved question, particularly in infants and young children, concerns how long an instance of word learning lasts. Most word learning studies measure learning after very brief delays. However, this may yield an overly rosy picture of how long these word-meaning mappings persist. The small number of studies that have examined word learning after longer delays (Gordon et al., 2016; Horst & Samuelson, 2008; Storkel, 2001, 2003) suggest that there are sometimes substantial dropoffs in performance after even a few minutes. This may mean that lengthier learning (a larger number of exposures, presumably) is required for word-meaning mappings to take hold. Further, findings from word production, rather than word recognition, suggest an even more lengthy process: children who are at roughly 50% accuracy in 3-alternative word recognition tests can produce the correct word form roughly 20% of the time (Storkel, 2001, 2003).

Another longstanding question is how contextual factors help (or hurt) children in pulling apart similar sounds. Do children (and learners generally) benefit from minimal contrasts—pressure to distinguish otherwise-identical word forms? Or do they benefit from contextual differences that pull apart sound patterns (as suggested by Feldman, Griffiths, & Morgan, 2009; Feldman, Myers, White, Griffiths, & Morgan, 2013; Thiessen, 2007), which is rather the opposite of the minimal pair hypothesis? Contextual differences could include different referents, different phonological contexts, perhaps even different semantic domains. Contextual differences might supervise the perceptual learning process, akin to the principle of acquired distinctiveness.

This is related to broader questions of massed vs. spaced practice. While adults benefit from interleaved, distributed learning instances (see, e.g., Cepeda et al., 2009), the evidence for vocabulary learning in young children is more mixed, with some studies suggesting benefits for spacing (Vlach, Ankowski, & Sandhofer, 2012; Vlach, Sandhofer, & Kornell, 2008) and others from massing (Schwab & Lew-Williams, 2016; Vlach & Johnson, 2013). Further, even adults may benefit from close succession of similar items to highlight critical differences (Carpenter & Mueller, 2013; Kang & Pashler, 2012). Given that words in child-directed speech tend to occur in "bursts" (massed learning instances; see, e.g., Onnis, Waterfall, & Edelman, 2008), it seems important to know whether this is the ideal structure for learning.

Relation to brain development

A critical area of research is to tie evidence of learning improvements or perceptual improvements to brain development. The brain changes massively during childhood (and beyond; see, e.g., Brown & Jernigan, 2012). Minimally, do these brain changes correlate with changes in perceptual performance? More subtly, do brain developments permit better acuity, or does more perceptual input yield changes in brain structure and function? Or are changes bidirectional?

Summary and conclusions

In this chapter, I have argued for a view of perceptual development in spoken language processing as a lengthy process that spans years, rather than a rapid process that is complete by the end of the first year of life. I want to be clear that this a reinterpretation, not a rejection, of the sizable infant speech perception and word recognition literature. That literature convincingly demonstrates that infants' perception is changing in response to language input, and that children can form associative memories briefly in the lab. What it does not show, I contend, is that infants are genius language learners. I encourage us as a field to reconstrue perceptual learning in spoken language as a protracted process, and at a societal level, to treat learners accordingly: young child learners need more input than one might think, while adult learners have many resources to leverage in acquiring a new language.

References

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, *32*(2), 233–250. http://doi.org/10.1016/S0095-4470(03)00036-6

Bartholomeus, B. (1973). Voice identification by nursery school children. *Canadian Journal of Psychology*, *27*(4), 464–72. http://doi.org/10.1037/h0082498

Barton, D. (1976). Phonemic discrimination and the knowledge of words in children under 3 years. *Papers and Reports on Child Language Development*, *11*, 1–9.

Bent, T. (2014). Children's perception of foreign-accented words. *Journal of Child Language*, *41*(6), 1334–1355. http://doi.org/10.1017/S0305000913000457

Bent, T., & Holt, R. F. (2018). Shhh… I Need Quiet! Children's Understanding of American, British, and Japanese-accented English Speakers. *Language and Speech*. http://doi.org/10.1177/0023830918754598

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–8. http://doi.org/10.1073/pnas.1113380109

Bergelson, E., & Swingley, D. (2017). Young Infants' Word Comprehension Given An Unfamiliar Talker or Altered Pronunciations. *Child Development*, *0*(0), 1–10. http://doi.org/10.1111/cdev.12888

Berman, J. M. J., Chambers, C. G., & Graham, S. A. (2010). Preschoolers' appreciation of speaker vocal affect as a cue to referential intent. *Journal of Experimental Child Psychology*, *107*(2), 87–99. http://doi.org/10.1016/j.jecp.2010.04.012

Borovsky, A., & Creel, S. C. (2014). Children and adults integrate talker and verb information in online processing. *Developmental Psychology*, *50*(5), 1600–13. http://doi.org/10.1037/a0035591

Bouchon, C., Floccia, C., Fux, T., Adda-Decker, M., & Nazzi, T. (2014). Call me Alix, not Elix: vowels are more important than consonants in own-name recognition at 5 months. *Developmental Science*, 1–12. http://doi.org/10.1111/desc.12242

Brown, T. T., & Jernigan, T. L. (2012). Brain Development During the Preschool Years. *Neuropsychological Review*, *22*, 313–333. http://doi.org/10.1007/s11065-012-9214-1

Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & Cognition*, *41*, 671–682. http://doi.org/10.3758/s13421-012-0291-4

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: theoretical analysis and practical implications. *Experimental Psychology*, *56*(4), 236–246. http://doi.org/10.1027/1618-3169.56.4.236

Cheour, M., Ceponiene, R. C., Leppänen, P., Alho, K., Kujala, T., Renlund, M., … Näätänen, R. (2002). The auditory sensory memory trace decays rapidly in newborns. *Scandinavian Journal of Psychology*, *43*, 33–39.

Cheour, M., Leppänen, P. H., Kraus, N., Cheour, M., & Leppa, P. H. T. (2000). Mismatch negativity (MMN) as a tool for investigating auditory discrimination and sensory memory in infants and children. *Clinical Neurophysiology*, *111*(1), 4–16. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10656505

Cowan, N. (2016). Working Memory Maturation: Can We Get at the Essence of Cognitive Growth? *Perspectives on Psychological Science*, *11*(2), 239–264. http://doi.org/10.1177/1745691615621279

Cowan, N. (2017). Mental Objects in Working Memory: Development of Basic Capacity or of Cognitive Completion? In *Advances in Child Development and Behavior* (1st ed., Vol. 52, pp. 81–104). Elsevier Inc. http://doi.org/10.1016/bs.acdb.2016.12.001

Cowan, N., Winkler, I., Teder, W., & Näätänen, R. (1993). Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 909–921.

Creel, S. C. (2014a). Preschoolers' flexible use of talker information during word learning. *Journal of Memory and Language*.

Creel, S. C. (2014b). Tipping the scales: Auditory cue weighting changes over development. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 1146–1160. http://doi.org/10.1037/a0036057

Creel, S. C. (2016). Ups and downs in auditory development: Preschoolers' sensitivity to pitch contour and timbre. *Cognitive Science*, *40*, 373–403. http://doi.org/10.1111/cogs.12237

Creel, S. C. (2018). Accent detection and social cognition: Evidence of protracted learning. *Developmental Science*, *21*(2), e12524. http://doi.org/10.1111/desc.12524

Creel, S. C. (2019). The familiar-melody advantage in auditory perceptual development: Parallels between spoken language acquisition and general auditory perception. *Attention, Perception, & Psychophysics*.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, *106*, 633–664. http://doi.org/10.1016/j.cognition.2007.03.013

Creel, S. C., & Jimenez, S. R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology*, *113*, 487–509.

Creel, S. C., & Quam, C. (2015). Apples and oranges: Developmental discontinuities in processing? *Trends in Cognitive Sciences*, *19*(12), 713–716. http://doi.org/10.1016/j.tics.2015.09.006

Creel, S. C., & Tumlin, M. A. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language*, *65*(3), 264–285. http://doi.org/10.1016/j.jml.2011.06.005

Creel, S. C., Weng, M., Fu, G., Heyman, G. D., & Lee, K. (2018). Speaking a tone language enhances musical pitch perception in 3–5-year-olds. *Developmental Science*, *21*(1), e12503. http://doi.org/10.1111/desc.12503

Davelaar, E. J., Tian, X., Weidemann, C. T., & Huber, D. E. (2011). A habituation account of change detection in same/different judgments. *Cognitive Affective Behavioral Neuroscience*, *11*, 608–626. http://doi.org/10.3758/s13415-011-0056-8

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176.

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Developmental Changes in Speech Discrimination in Infants. *Journal of Speech and Hearing Research*, *20*(4), 766–780. http://doi.org/https://doi.org/10.1044/jshr.2004.766

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–306.

Elliott, L. L. (1979). Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *Journal of the Acoustical Society of America*, *66*(3), 651–653.

Fallon, M., Trehub, S. E., & Schneider, B. a. (2000). Children's perception of speech in

multitalker babble. *Journal of the Acoustical Society of America*, *108*(6), 3023–3029. http://doi.org/10.1121/1.1323233
Fallon, M., Trehub, S. E., & Schneider, B. a. (2002). Children's use of semantic cues in degraded listening environments. *Journal of the Acoustical Society of America*, *111*(5), 2242–2249. http://doi.org/10.1121/1.1466873
Fancourt, A., Dick, F. K., & Stewart, L. (2013). Pitch-Change Detection and Pitch-Direction Discrimination in Children. *Psychomusicology: Music, Mind. and Brain*, *23*(2), 73–81. http://doi.org/10.1037/a0033301
Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, *116*(4), 752–82. http://doi.org/10.1037/a0017196
Feldman, N. H., Myers, E. B., White, K. S., Griffiths, T. L., & Morgan, J. L. (2013). Word-level information influences phonetic learning in adults and infants. *Cognition*, *127*(3), 427–38. http://doi.org/10.1016/j.cognition.2013.02.007
Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development*, *81*(5), 1376–83. http://doi.org/10.1111/j.1467-8624.2010.01479.x
Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*, *46*(Pt 2-3), 245–64. http://doi.org/10.1177/00238309030460020901
Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development*, *83*(1), 203–222. http://doi.org/10.1111/j.1467-8624.2011.01692.x
Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*(1), 98–116.
Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*(3), 228–231. http://doi.org/10.1111/1467-9280.00044
Flege, J. E. (1987). A critical period for learning to pronounce foreign languages? *Applied Linguistics*, *8*, 162–177.
Flege, J. E. (1992). Speech learning in a second language. In C. A. Ferguson, L. Menn, & C. Stoel-gammon (Eds.), *Phonological development: models, research, implications* (pp. 565–604). Timonium, MD: York Press.
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* (pp. 233–277). Timonium, MD: York Press.
Floccia, C., Butler, J., Girard, F., & Goslin, J. (2009). Categorization of regional and foreign accent in 5- to 7-year-old British children. *International Journal of Behavioral Development*, *33*(4), 366–375. http://doi.org/10.1177/0165025409103871
Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Delle Luche, C., Durrant, S., … Vihman, M. M. (2016). British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition*, *148*, 1–9. http://doi.org/10.1016/j.cognition.2015.12.004
Friend, M., & Bryant, J. B. (2000). A developmental lexical bias in the interpretation of discrepant messages. *Merrill-Palmer Quarterly*, *46*(2), 342–369.
Galle, M. E., Apfelbaum, K. S., & McMurray, B. (2014). The Role of Single Talker Acoustic

Variation in Early Word Learning. *Language Learning and Development*, *11*(1), 66–79. http://doi.org/10.1080/15475441.2014.895249

Garnica, O. K. (1971). The Development of the Perception of Phonemic Differences in Initial Consonants by English-Speaking Children: A Pilot Study. *Papers and Reports on Child Language Development*, *3*, 1–31.

Girard, F., Floccia, C., & Goslin, J. (2008). Perception and awareness of accents in young children. *British Journal of Developmental Psychology*, *26*(3), 409–433. http://doi.org/10.1348/026151007X251712

Gomes, H., Sussman, E., Ritter, W., Kurtzberg, D., Cowan, N., & Vaughan, H. G. (1999). Electrophysiological evidence of developmental changes in the duration of auditory sensory memory. *Developmental Psychology*, *35*(1), 294–302. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9923483

Gordon, K. R., Mcgregor, K. K., Waldier, B., Curran, M. K., Gomez, R. L., Samuelson, L. K., & Gordon, K. R. (2016). Preschool Children's Memory for Word Forms Remains Stable Over Several Days, but Gradually Decreases after 6 Months. *Frontiers in Psychology*, *7*(September), 1–15. http://doi.org/10.3389/fpsyg.2016.01439

Graf Estes, K., Edwards, J., & Saffran, J. R. (2011). Phonotactic constraints on infant word learning. *Infancy*, *16*(2), 180–197. http://doi.org/10.1111/j.1532-7078.2010.00046.x

Gravel, S., & Traquina, D. N. (1992). Experience with the audiologic assessment of infants and toddlers. *International Journal of Pediatric Otorhinolaryngology*, *23*, 59–71.

Háden, G. P., Stefanics, G., Vestergaard, M. D., Denham, S. L., Sziller, I., & Winkler, I. (2009). Timbre-independent extraction of pitch in newborn infants. *Psychophysiology*, *46*(1), 69–74. http://doi.org/10.1111/j.1469-8986.2008.00749.x

Haith, M. M. (1990). Progress in the Understanding of Sensory and Perceptual Processes in Early Infancy. *Merrill-Palmer Quarterly*, *36*(1), 1–26.

Halpern, A. R., & Bartlett, J. C. (2010). Memory for melodies. In M. R. Jones, R. R. Fay, & A. N. Popper (Eds.), *Springer Handbook of Auditory Research: Music Perception* (pp. 233–258). New York: Springer.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, *28*(4), 377–396. http://doi.org/10.1006/jpho.2000.0121

Holt, R. F., & Carney, A. E. (2007). Developmental Effects of Multiple Looks in Speech Sound Discrimination. *Journal of Speech Language and Hearing Research*, *50*(6), 1404. http://doi.org/10.1044/1092-4388(2007/098)

Holt, R. F., & Lalonde, K. (2012). Assessing toddlers' speech-sound discrimination. *International Journal of Pediatric Otorhinolaryngology*, *76*(5), 680–692. http://doi.org/10.1016/j.ijporl.2012.02.020

Horst, J. S., & Samuelson, L. K. (2008). Fast Mapping but Poor Retention by 24-Month-Old Infants. *Infancy*, *13*(2), 128–157. http://doi.org/10.1080/1525O(xH)701795598

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1570–1582. http://doi.org/10.10371/0096-1523.26.5

Ito, K., Bibyk, S. a, Wagner, L., & Speer, S. R. (2014). Interpretation of contrastive pitch accent in six- to eleven-year-old English-speaking children (and adults). *Journal of Child Language*, *41*(1), 1–27. http://doi.org/10.1017/S0305000912000554

Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, *14*(5), 1002–1011.

http://doi.org/10.1111/j.1467-7687.2011.01052.x
Jones, Z., Yan, Q., Wagner, L., & Clopper, C. G. (2017). The development of dialect classification across the lifespan. *Journal of Phonetics*, *60*, 20–37. http://doi.org/10.1016/j.wocn.2016.11.001
Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*, *21*(1–2), 3–28. Retrieved from http://ezproxy.library.arizona.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1993-44875-001&site=ehost-live
Kang, S. H. K., & Pashler, H. (2012). Learning Painting Styles: Spacing is Advantageous when it Promotes Discriminative Contrast. *Applied Cognitive Psychology*, *26*(1), 97–103. http://doi.org/10.1002/acp.1801
Keen, R. (2003). Representation of objects and events: Why do infants look so smart and toddlers look so dumb? *Current Directions in Psychological Science*, *12*(3), 79–83. http://doi.org/10.1111/1467-8721.01234
Keller, T. A., & Cowan, N. (1994). Developmental increase in the duration of memory for tone pitch. *Developmental Psychology*, *30*(6), 855–863. http://doi.org/10.1037//0012-1649.30.6.855
Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H. H., … Wang, Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, *14*(3), 220–224.
Kuhl, P. K. (1979). Speech perception in early infancy: perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, *66*(6), 1668–79. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/521551
Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, *6*, 263–285. http://doi.org/10.1016/S0163-6383(83)80036-8
Kuhl, P. K., Ramírez, R. R., Bosseler, A., Lin, J.-F. L., & Imada, T. (2014). Infants' brain responses to speech suggest analysis by synthesis. *PNAS Proceedings of the National Academy of Sciences*, *111*(31), 11238–11245. http://doi.org/10.1073/pnas.1410963111
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*(2), F13–F21. http://doi.org/10.1111/j.1467-7687.2006.00468.x
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*(5044), 606–608.
lmada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. K. (2006). Infant speech perception activates Broca's area: A developmental magnetoencephalography study. *NeuroReport: For Rapid Communication of Neuroscience Research*, *17*(10), 957–962. http://doi.org/10.1097/01.wnr.0000223387.51704.89
Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*(2), 202–227. http://doi.org/10.1037/0096-3445.132.2.202
Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science*, *6*(5), 314–317.
Mann, V. A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology*, *27*(1), 153–165. http://doi.org/10.1016/0022-0965(79)90067-5

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*(3), F9–F16. http://doi.org/10.1111/j.1467-7687.2008.00671.x

McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Developmental Psychology*, *54*(8), 1472–1491.

Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development*, *72*(3), 834–843.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. http://doi.org/10.1016/j.clinph.2007.04.026

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination. *Developmental Science*, *13*(3), 407–20. http://doi.org/10.1111/j.1467-7687.2009.00898.x

Nathan, L., Wells, B., & Donlan, C. (1998). Children's comprehension of unfamiliar regional accents: a preliminary investigation. *Journal of Child Language*, *25*(2), 343–365. http://doi.org/10.1017/S0305000998003444

Nelson, N. L., & Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology*, *110*(1), 52–61. http://doi.org/10.1016/j.jecp.2011.03.014

Nittrouer, S. (2001). Challenging the notion of innate phonetic boundaries. *Journal of the Acoustical Society of America*, *110*(3), 1598–1605. http://doi.org/10.1121/1.1379078

Oh, G. E., Guion-Anderson, S., Aoyama, K., Flege, J. E., Akahane-Yamada, R., & Yamada, T. (2011). A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting. *Journal of Phonetics*, *39*(2), 156–157. http://doi.org/10.1016/j.wocn.2011.01.002

Ohde, R. N., & Haley, K. L. (1997). Stop-consonant and vowel perception in 3- and 4-year-old children. *Journal of the Acoustical Society of America*, *102*(6), 3711–22. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9407663

Onnis, L., Waterfall, H. R., & Edelman, S. (2008). Learn locally, act globally: learning language from variation set cues. *Cognition*, *109*(3), 423–30. http://doi.org/10.1016/j.cognition.2008.10.004

Pająk, B., Creel, S. C., & Levy, R. (2016). Difficulty in learning similar-sounding words: a developmental stage or a general property of learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(9), 1377–1399. http://doi.org/10.1037/xlm0000247

Pater, J., Stager, C. L., & Werker, J. F. (2004). The Perceptual Acquisition of Phonological Contrasts. *Language*, *80*(3), 384–402. http://doi.org/10.1353/lan.2004.0141

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d/–/ð/ perception: Evidence for a new developmental pattern. *The Journal of the Acoustical Society of America*, *109*(5), 2190–2201. http://doi.org/10.1121/1.1362689

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421–35. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8189202

Quam, C., & Swingley, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Development*, *83*(1), 236–50. http://doi.org/10.1111/j.1467-

8624.2011.01700.x
Rigler, H., Farris-trimble, A., Greiner, L., Walker, J., Tomblin, J. B., & Mcmurray, B. (2015). The Slow Developmental Time Course of Real-Time Spoken Word Recognition. *Developmental Psychology*, *51*(12), 1690–1703.
Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339–349. http://doi.org/10.1111/j.1467-7687.2008.00786.x
Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, *15*(6), 608–635. http://doi.org/10.1111/j.1532-7078.2010.00033.x
Saindon, M. R., Trehub, S. E., Schellenberg, E. G., & Van Lieshout, P. (2016). Children's identification of questions from rising terminal pitch. *Journal of Child Language*, *43*, 1174–1191. http://doi.org/10.1017/S0305000915000458
Saindon, M. R., Trehub, S. E., Schellenberg, E. G., & Van Lieshout, P. (2017). When is a Question a Question for Children and Adults? *Language Learning and Development*, *0*(0), 1–12. http://doi.org/10.1080/15475441.2016.1252681
Sams, M., Hari, R., Rif, J., & Knuutila, J. (1993). The Human Auditory Sensory Memory Trace Persists about 10 sec: Neuromagnetic Evidence. *Journal of Cognitive Neuroscience*, *5*(3), 363–370. http://doi.org/10.1162/jocn.1993.5.3.363
Schwab, J. F., & Lew-Williams, C. (2016). Repetition Across Successive Sentences Facilitates Young Children's Word Learning. *Developmental Psychology*, *52*(6), 879–886.
Sekerina, I. a, & Brooks, P. J. (2007). Eye movements during spoken word recognition in Russian children. *Journal of Experimental Child Psychology*, *98*(1), 20–45. http://doi.org/10.1016/j.jecp.2007.04.005
Spence, M. J., Rollins, P. R., & Jerger, S. (2002). Children's Recognition of Cartoon Voices. *Journal of Speech, Language, and Hearing Research*, *45*(1), 214–222. http://doi.org/10.1044/1092-4388(2002/016)
Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*, 381–382.
Stalinski, S. M., Schellenberg, E. G., & Trehub, S. E. (2008). Developmental changes in the perception of pitch contour: distinguishing up from down. *Journal of the Acoustical Society of America*, *124*(3), 1759–63. http://doi.org/10.1121/1.2956470
Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, *44*(6), 1321–1337. http://doi.org/10.1044/1092-4388(2001/103)
Storkel, H. L. (2003). Learning new words II: Phonotactic probability in verb learning. *Journal of Speech, Language, and Hearing Research*, *46*(6), 1312–1323. http://doi.org/10.1044/1092-4388(2003/102)
Sundara, M., Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of /d-th/ in monolingual and bilingual acquisition of English. *Cognition*, *100*(2), 369–88. http://doi.org/10.1016/j.cognition.2005.04.007
Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*, 147–66.
Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, *13*(5), 480–484. http://doi.org/10.1111/1467-9280.00485

Tanenhaus, M. K., Spivey-knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Thiessen, E. D. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language*, *56*(1), 16–34. http://doi.org/10.1016/j.jml.2006.07.002

Tincoff, R., & Jusczyk, P. W. (1999). Some Beginnings of Word Comprehension in 6-Month-Olds. *Psychological Science*, *10*(2), 172–175.

Tincoff, R., & Jusczyk, P. W. (2012). Six-Month-Olds Comprehend Words That Refer to Parts of the Body. *Infancy*, *17*(4), 432–444. http://doi.org/10.1111/j.1532-7078.2011.00084.x

Trainor, L. J., & Trehub, S. E. (1994). Key membership and implied harmony in Western tonal music: developmental perspectives. *Perception & Psychophysics*, *56*(2), 125–32. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7971113

Trehub, S. E. (1979). Reflections on the development of speech perception. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *33*(4), 368–381. http://doi.org/10.1037/h0081733

Trehub, S. E. (2012). Behavioral methods in infancy: pitfalls of single measures. *Annals of the New York Academy of Sciences*, *1252*, 37–42. http://doi.org/10.1111/j.1749-6632.2012.06448.x

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development*, *55*(3), 821–830.

Trehub, S. E., Thorpe, L. A., & Morrongiello, B. A. (1985). Infants' perception of melodies: Changes in a single tone. *Infant Behavior and Development*, *8*(2), 213–223. http://doi.org/10.1016/S0163-6383(85)80007-2

van Heugten, M., Volkova, A., Trehub, S. E., & Schellenberg, E. G. (2013). Children's Recognition of Spectrally Degraded Cartoon Voices. *Ear and Hearing*, *35*(1), 118–125.

Van Lancker, D., Cornelius, C., & Kreiman, J. (1989). Recognition of emotional‐prosodic meanings in speech by autistic, schizophrenic, and normal children. *Developmental Neuropsychology*, *5*(2–3), 207–226. http://doi.org/10.1080/87565648909540433

Viemeister, N. F., & Wakefield, H. (1991). Temporal integration and multiple looks. *Journal of the Acoustical Society of America*, *90*(1), 858–865.

Vlach, H. A., Ankowski, A. A., & Sandhofer, C. M. (2012). At the same time or apart in time? the role of presentation timing and retrieval dynamics in generalization. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(1), 246–254. http://doi.org/10.1037/a0025260

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375–382. http://doi.org/10.1016/j.cognition.2013.02.015

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109*(1), 163–7. http://doi.org/10.1016/j.cognition.2008.07.013

Wagner, L., Clopper, C. G., & Pate, J. K. (2014). Children's perception of dialect variation. *Journal of Child Language*, *41*, 1062–1084. http://doi.org/10.1017/S0305000913000330

Walley, A. C., Metsala, J. L., & Garlock, V. M. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing*, *16*(1), 5–20. http://doi.org/10.1023/A:1021789804977

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*(1), 1–30. http://doi.org/10.1207/15250000252828226

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology*, *66*, 173–196. http://doi.org/10.1146/annurev-psych-010814-015104

Werker, J. F., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. http://doi.org/10.1016/S0163-6383(84)80022-3

Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology*, *37*(2), 278–286.

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, *59*(1), 114–132. http://doi.org/10.1016/j.jml.2008.03.001

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science*, *12*(3), 412–8. http://doi.org/10.1111/j.1467-7687.2008.00789.x

Figure captions

Figure 1. Cartoon face stimuli used in an early word-learning experiment in my lab.

Figure 2. An early study from my lab, in which college-aged adults clobbered preschool-aged children in learning words. Error bars are standard errors; dashed line represents chance performance.

Figure 3. Better child-directed cartoon characters used in later studies in my lab.

Figure 4. Children are better at learning distinct-sounding word pairs as labels for pictures than they are at learning minimal pairs that differ by a single sound contrast. Performance improves with age, but even at 6 years children are not yet adultlike in accuracy. Chance (dashed line) is .50. Throughout the chapter, shaded areas around linear fits represent 95% confidence intervals.

Figure 5. Children are better at associating distinct-sounding voices with cartoon characters than less-distinct voices. Performance improves with age, but even at 6 years children are not yet adultlike in accuracy. Chance is .50.

Figure 6. Children are better at discriminating easy (trained) melody stimuli than they are at distinguishing pitch contours. Performance improves with age, but even at 6 years children are not yet adultlike in accuracy. Chance is .50 in that each data point is an average of correct rejections (same trials) and hits (different trials).

Figure 7. Fixation proportions are linked to pointing accuracy across multiple word-learning studies from my lab (286 children, 567 word pairs).

Figure 8. Possible developmental trajectories of perceptual sensitivity. A W-function seems unlikely, but it's logically possible since we don't really have the data to discount it.