Title

Kidiolects: How children perceive child speech

Author

Creel, Sarah C

Publication Date

2026

DOI

10.1016/bs.acdb.2026.08.003

Kidiolects: How children perceive child speech

Sarah C. Creel
Department Cognitive Science
University of California San Diego
screel@ucsd.edu

10230 words including references

Abstract

A longstanding assumption is that children acquire language mainly from adult models, but newer research suggests this cannot be correct. Children experience a great deal of child speech as well, particularly across a broader sample of cultures. They enjoy listening to child speech. They can comprehend and learn from child speech. This chapter discusses acoustic properties unique to child speech; argues for its prevalence and importance as an input source to other children; outlines an account that children recognize speech contingent on the speaker's age and identity; and discusses some outstanding questions about children's comprehension of and learning from child speech. The goal is to convince readers that there is an enormous gap in our knowledge of a major component of children's language input, which shapes both child comprehension and production.

Keywords

Child speech, speech production, speech perception, word recognition, word learning, comprehension, sociolinguistics, dialect variation

Kidiolects: How children perceive child speech

This chapter, and to some extent large parts of my research program, asks how children process speech patterns that vary across speakers, situations, and pronunciation detail. How do you learn and recognize speech patterns when the patterns themselves are so variable? A longstanding simplifying assumption—which is almost certainly wrong, especially outside WEIRD (Western educated industrialized rich developed nations) populations—is that the acoustic-phonetic structure of language knowledge derives from caregiver speech that is directed at the child. I say almost certainly wrong because in everyday life and in the media, children hear a plethora of talkers with varied accents, voice qualities, and skill levels. Since most research on speech sound development considers adult child-directed speech, it is much less well understood how children recognize words in other speech varieties or how and when they differentiate those varieties.

Here I consider a particular situation of variable, diverse input: how children learn and recognize spoken language *from other children*. Because of a longstanding focus on propitious properties of parental input—child-directed speech—most of the literature on child word recognition and speech sound processing focuses on children's understanding of adult child-directed speech. In what follows, I will try to convince readers that **child speech is interestingly different** from adult speech; that it is an **important input source** to other children (at the speech level and probably other levels too); that we **need a better understanding** of how children process child speech; and that there are some **consequential and under-studied questions** out there with respect to child speech.

Like many linguists, I take a descriptive, not prescriptive, stance toward child speech: while it is different from adult speech, and these differences may have consequences for communicative efficiency, child speech characteristics are not wrong or bad. As a result, if I use words such as "errors" or "misarticulations" as shorthand, they may appear in quotation marks or with qualifications. This isn't to minimize potential impacts on comprehension or social interaction, but to emphasize that my goal is to characterize variation, not judge it.

CHILD SPEECH CONTAINS LESS CATEGORY SEPARATION AND MORE ACOUSTIC VARIABILITY THAN ADULT SPEECH

Child speech sounds different from adult speech in multiple respects. Compared to adults, children have smaller vocal folds and shorter vocal tracts. As a result, children's fundamental frequency (f0, or voice pitch, though that's probably an oversimplification) is higher, and their formant frequencies are also higher (i.e., resonances in the vocal tract that change across vowels and consonant place of articulation, often enumerated F1, F2, F3 and so on).[1] Another feature, somewhat underappreciated, is that their rate of speech production is measurably slower (Kent & Forner, 1980; Lee et al., 1999).

[1] Though there are more than three formants, the first three are viewed as most consequential for identifying vowel sounds.

Children also produce more variable speech sounds than adults, at both a large scale and a small scale. Before going into more detail, an important factor to consider about our understanding of pronunciation is how production accuracy/variation is being evaluated. One major source of data on child speech production accuracy is skilled listener evaluation, that is, impressionistic judgments (e.g. Smit et al., 1990). On the plus side, listener judgments can respond to anything that sounds typical or atypical in a speech sound. However, impressionistic judgments have the disadvantage of potential observer bias: for example, if a listener expects the word *rainbow*, or knows they are hearing a child, they may be biased to report a word-initial speech sound as more r-like than it actually is.

Another source of data is acoustic measurement. Acoustic measures are both more precise and less prone to human bias. (Human bias can still enter the picture unless the process is completely automated, which may introduce its own errors; see Ahn et al., 2023, for discussion on error checking in automated analysis pipelines.) A disadvantage of acoustic measures is that they are limited in the acoustic and thus articulatory features that they investigate. What if a child producer doesn't use F2 and F3 to distinguish between r and w, as adults do, but changes something else that's idiosyncratic, say, breathiness? A human listener might or might not detect this, whereas acoustic measurements targeted to known acoustic cues would surely miss it.

In any case, even if we correct for differences between children's and adults' absolute formant frequencies, children show both more large-scale and small-scale speech sound variation than adults do. At a large scale, some children may **merge** some speech sound categories that are distinguished by adult speakers (see, among others, Smit et al., 1990 for normative data). Several speech sound mergers or substitutions are considered typical developmental speech patterns. For instance, many English-acquiring children produce the word-initial r sound much like they produce the w sound (e.g. Hoffman et al., 1983; McAllister Byun & Tiede, 2017), or the sh sound much like the s sound (e.g. Li et al., 2009; Roepke & Brosseau-Lapré, 2019). Thus, *wing* and *ring* sound virtually identical, and *sell* and *shell* sound virtually identical (though see Munson et al., 2010, for listener ability to detect substitutions). These differences decrease with development (Shriberg, 2009; Smit et al., 1990) but in some cases may persist into later childhood or adulthood (Shriberg, 2009). On top of this, some children produce the adult speech sound contrasts while others merge or reduce them, resulting in more variability across child speakers than across adult speakers. There is incomplete understanding of how much these sound differences reflect inaccurate perception of sounds, which then leads to inaccurate production, vs. accurate perception but inability to produce the exact motor commands to the speech articulators. I return to the topic of perception-production interactions later.

On the small-scale side, there is more variation *within* a speech sound category for child speech than for adult speech. This variability shows up at the level of a single speaker and across a set of child speakers, for a number of years (Idemaru & Holt, 2013; Lee et al., 1999; Macken & Barton, 1980; Whiteside et al., 2003; Zlatin & Koenigsknecht, 1976). These differences are measurable at the acoustic level, and possibly by gradient or aggregate human listener judgments (e.g. McAllister Byun et al., 2016; Schellinger et al., 2017). This suggests that, despite

impressionistically sounding adultlike, children tend to be less precise at hitting exact articulatory (speech-motor) targets than adults are.

All of these sources of variability in child speech, plus the fact that children across a sufficient age range have wider variation than adults in body size (a correlate of vocal tract length), suggest that there is more cross-child speech variability than cross-adult speech variability—at least first-language speaking adults with fairly uniform accent characteristics. Cross-child variability in speech production is similar to findings from adult second-language speech, in that individual speakers vary in how closely they approximate first-language adult sound patterns (Wade et al., 2007; Xie & Jaeger, 2020). Greater within-child speech variability may differ from second-language speech (Vaughn et al., 2018; Xie & Jaeger, 2020), where *individual speakers* may produce particular sounds in their second language as consistently as first-language speakers. Children appear not to show this individual level of consistency.

Higher variability may not be a bad thing. Children can do well at recognizing speech in new accents or within accent-like phonetic distortions under supportive conditions. Studies of infants (e.g. Schmale et al., 2010; Schmale & Seidl, 2012), toddlers (van Heugten & Johnson, 2014, 2015; van Heugten et al., 2015; Swingley, 2016; Swingley & Aslin, 2000, 2002), and young children (3-6 years; Creel, 2012) indicate reasonably good word recognition despite mispronunciations or less familiar accents, though sometimes brief exposure to the accent is needed.

Additionally, there is a sizable literature at this point exploring the hypothesis that within-speaker or cross-speaker variation in acoustic characteristics benefits word learning and accent adaptation. For example, if there is random variation in a word's absolute pitch, then absolute pitch is ruled out as a cue to word identity. So maybe hearing a variable child speaker facilitates learning, or maybe hearing speech across a range of children, or in both child and adult voices, facilitates learning. Some research supports this: toddlers appear to learn similar-sounding words better with variability amongst a set of adult talkers (Rost & McMurray, 2009, 2010) or within an adult talker (Galle et al., 2014). Young adults appear to learn accent characteristics better from multi-talker exposure than single-talker exposure (Bradlow & Bent, 2008; Lively et al., 1993; Logan et al., 1991; Sidaras et al., 2009; Xie et al., 2021; though see Brekelmans et al., 2022 for a null result, and Perrachione et al., 2011, for a complex set of findings).

But how far do these variability benefits extend? Work in my own lab suggests that word variation which crosses sound category boundaries is not helpful but may be slightly harmful to learning (Creel 2014; Frye & Creel, 2022; Muench & Creel, 2013). For example, if a child hears the same novel object picture labeled as both *deev* and *div*, they are slightly less accurate at recognizing it later (labeled with either word) than children trained that the object is just a *div* (and tested on div). (Note that all of these studies used adult speakers.) Thus, high variability may only work up to a point. Bringing this back to child speech, it's possible that higher-variability child speech introduces mild confusion into the learning process, either from speech across children with varying articulation skills, or a combination of child and adult speech.

At any rate, children are more variable than adults, which has potential consequences for comprehension and learning. Researchers quantifying child speech acoustics should also be mindful, though, that we may be missing some dimensions of acoustic variability that children unexpectedly use. For example, perhaps children change loudness, spectral slope, or duration of two sounds that are indistinguishable in their formant frequencies.

Covert contrasts. Acoustic analysis of speech productions suggests that some children are covertly contrasting sounds that are, to adult listeners, indistinguishable (Scobbie et al., 1996). It is less clear that it benefits these child perceivers. Multiple findings suggest that at least some children who impressionistically appear to be merging a contrast may in fact be producing statistically detectable differences at the acoustic level (Li et al., 2009; Munson et al., 2010; Scobbie et al., 1996; though see Cleland et al., 2017, for null articulatory findings in velar fronting.) That is, if you collect enough of their s and sh attempts, or s and th, or r and w; compute the acoustic properties; and run a t-test, those properties might differ. Some such differences may be detectable by human listeners, either by aggregating identification judgments or asking human listeners to provide gradient judgments (McAllister Byun et al., 2016; Munson et al., 2010; Schellinger et al., 2017). Even though these differences are detectable, the distributions of the sounds may still have substantial overlap (e.g. Li et al., 2009). Evidence that children who produce these covert contrasts actually perceive the intended sound is more limited, with some studies suggesting that covert-contrasting children cannot tell the difference between their own sounds (Roepke & Brosseau-Lapré, 2019). A related pattern is *atypical* contrast, where children produce a sound differently than adults would, but distinct from its contrasting sound, such as producing a lateral fricative for s (s-like sound where the sound escapes from the side of the tongue instead of the front). In these cases, there is reasonable evidence for children discriminating their own contrasts (Roepke & Brosseau-Lapré, 2019).

CHILDREN'S SPEECH PRODUCTION IS SHAPED BY THEIR PERCEPTUAL REPRESENTATIONS, WHICH ARE…WHAT, EXACTLY?

It's of course necessary to consider where children's productions come from, and a short answer is their percepts. A general assumption, and one supported by evidence, is that perceptual representations of speech precede and are necessary for speech production (Guenther et al., 1998; Tourville & Guenther, 2011). Infants who do not get exposure to spoken language because they have hearing impairment do not babble with the frequency that hearing infants do (Oller & Eilers, 1988). A rich literature on infant speech perception shows that infants can discriminate many speech sounds, only some of which they can produce (see Johnson & White, 2025, for an overview). Children with speech sound production difficulties often have perceptual difficulties (Hearnshaw et al., 2019), consistent with a role of perception in speech sound production. Thus perceptual representations are a necessary, if not sufficient, ingredient for accurate speech production. Less studied is *whose speech* these perceptual representations are based on. I discuss three possible input sources, with the latter two being more controversial than the first.

Parental input. As noted earlier, a major assumption in the literature is that children's speech percepts consist of adults' child-directed speech (CDS). CDS (sometimes referred to by its

sibling term infant-directed speech, or IDS) is purported to be the critical ingredient in children's early language development. Child-directed speech (CDS), typically defined as slower, higher-pitched, more pitch-variable, and grammatically and lexically simpler form of speech directed at child listeners (see Cox et al, 2023, for a meta-analysis of CDS properties), has been a huge focus of research for years. There is good evidence that children listen more to CDS than to adult-directed speech, or ADS (Fernald & Kuhl, 1987; Zettersten et al., 2024). However, there is still not consensus as to why, or even if, it facilitates language development (see Kempe et al., 2024, for a review; see also Floccia et al., 2016). One older idea was that CDS facilitated language acquisition because it contained clearer speech sound category structure due to hyperarticulation (Kuhl et al., 1997). However, more recent work suggests that CDS may be more phonetically variable than ADS (McMurray et al., 2013; Martin et al., 2015). Despite this, one might still argue that CDS benefits language development for other reasons, such maintaining attention or being better suited to children's nascent syntactic and lexical knowledge.

As an important aside, this emphasis on child-directed adult input has often struck me as problematic in a lot of ways. What if adults in one's culture rarely talk directly to children (see Ochs, 1982)? What if one's personal adults are not very good at doing the 'right' kind of child-directed speech? Why can't we agree on the critical properties—attentional, informational, acoustic—that make child-directed speech so important in the first place? What if it doesn't actually help language development that much? And finally, if children only pay attention when speech is directed at them, how do they pick up cartoon character accents and parental swears?

In any case, as a result of this focus on and presumed dominance of CDS, most of the speech stimuli in infant and child language development studies use CDS-like characteristics, and usually with a female speaker on the assumption that maternal CDS is preeminent. For the record, in one young-child word learning study in my lab (Creel & Frye, 2024), it didn't matter if the speaker was male ADS or female CDS—children learned similarly in both cases. Further, relatively few studies have examined paternal speech to children (see Ferjan Ramírez, 2022; Salo et al., 2016). Because of the very strong representation of CDS across developmental studies of word learning and word recognition, we don't know as much about how other forms of speech might contribute to children's perceptual representations or vocabularies (Bulgarelli, 2026).

Figure 1. Adult and child acoustic measurements of log(F3/F2), a distinguishing cue for r and w productions, balanced to equate child and adult frequency of occurrence and r and w frequency of occurrence. Y axis indicates probability density. Upper left: child (dashed) and adult speech (solid) categories; upper right: adult categories; lower left: combined categories; lower right: child categories. Purple: intended r; orange: intended w. A boundary of log(F3/F2) of 0.7 impressionistically classifies most adult instances correctly, but not child speech.

Overheard adult speech. Another possible source of input to children is **adult-directed speech (ADS)** or overheard speech. It's important to consider what characteristics distinguish ADS from CDS in any given experiment or observational study. Is it characteristics of *form* (Fernald & Kuhl, 1987; Cox et al., 2023)—are ADS acoustics potentially less interesting to children? Is it instead *content*—does it have adult-directed syntax, vocabulary, and topics? Is ADS instead a characterization of speech where the adult is not undergoing joint attention with the child (that is, not talking *to* the child)? The more general literature on learning from overheard speech is not very large, but it is a little ambivalent. On the one hand, Kuhl, Tsao, and Liu (2003) found that 9-month-olds learned speech sounds in an unfamiliar language only from a live person but not from a recording of that person. However, Akhtar (2005; Floor & Akhtar, 2006) reported that children at 1.5-2 years can learn vocabulary by overhearing. I can personally attest that my young children have learned to produce some nouns that I have only directed at other adult (very bad) drivers.

Child speech. A third possible source of input to children's perceptual representations is child speech (see Bulgarelli, 2026, for recent discussion of the importance of other-child speech).[2] Children are frequently exposed to other children's speech from siblings (Hippe & Ferjan Ramírez, 2022), in child care settings (Weisleder & Fernald, 2013), and in many cultural contexts outside the WEIRD bubble (Cristià et al., 2023; Loukatou et al., 2022; Schick & Stoll, 2025). Cristià et al. (2023) estimated that children in their Tsimané sample heard as much or more child speech than adult speech. Even more overlooked, children (like most people) also frequently hear their own voices: one estimate I derived from Bergelson et al. (2023), who assessed a world-spanning sample of 1001 children's language environments, is that by age 4 years, children produce about as many utterances as adults do. This means that a child has substantial exposure to their own speech output. Unless they can somehow stop themselves from encoding all this child speech, then it must be the case that child speech is a component of their perceptual representations. This might be useful in cultures and settings where there is less speech directly from adults (see Bulgarelli, 2026, for similar arguments), and it might also be a particularly apt perceptual model because child speech is more acoustically similar to children's own productions. Schick et al. (2025) remark that child speech acoustically resembles CDS—or perhaps the other way around: maybe children like CDS because it sounds more childlike.

Is it even possible that children somehow tune out or ignore speech from other children? This seems unlikely, given evidence that they enjoy listening to child speech. Masapollo et al. (2016) report that 4-6-month-old infants prefer to listen to childlike synthetic vowels rather than adultlike synthetic vowels. Schick et al. (2025) find that both Swedish and Peruvian-Amazonian children ages 8-20 months show equal preference for adult CDS and non-CDS child speech, and they actually prefer child speech to ADS. While children do appear to *understand* adult speech somewhat better than child speech (as discussed below), they nonetheless understand other-child speech in many cases (Cooper et al., 2018; Creel et al., under review; Żygis et al., 2023), and can learn words from child speakers (Erskine & Seidl, 2026 in 2-year-olds; Levy & Hanulíková, 2023 in 7-11-year-olds).

How children perceive child speech: Three possibilities in order of decreasing absurdity

The above evidence that children *experience* quite a lot of child speech is intertwined with the question of how they perceive it. More specifically, what perceptual representations do they compare incoming child speech to—adult representations, some undifferentiated amalgam of all their speech input, or more specific representations such as child vs. adult or even specific speaker representations? I discuss each of these possibilities below, starting with the standard account: recognition by comparison to adult speech.

1 – By the standard account of child spoken language acquisition, their representations of words and speech sounds should resemble adult speech. This means that despite experiencing child speech, they must encode adult speech more robustly. This might happen because in general, there are more adults than children, which would lead to more adult input (though see

[2] On a personal note, I experience occasional episodes of child-to-child learning. For example, my 5-year-old now says "bruh", which I'm sure she learned from her 7-year-old sibling because neither parent knows how to use the term.

previous section). A second reason adult speech might be encoded more robustly is that the higher "cue noise" in children's speech sounds might lead learners—including other children—to downweight child speech input. Finally, a literature on informant trust suggests adult input might get higher "weight" in word learning because, all else being equal, children credit adults with greater knowledge (Jaswal & Neely, 2006; see also Taylor et al., 1991, and Sumner et al., 2014, for an account of prestige-related input weighting in word representations). Whether greater adult epistemic trust weakens child word learning from children, or affects distributional learning of speech sound detail, is less clear.

Supporting the standard account, children understand adult speech better than child speech. In the small number of studies that have investigated this question, children comprehend adult speech better than child speech. Dodd (1975) asked children ages 3;2-4;7 (years;months) to produce words that were difficult for each child. Dodd (1975) then presented those recorded words back to the child along as they saw a set of four pictures. One picture's name was the intended production (e.g. *snake*), another's name was similar to the target word (e.g. *cake*), and two that were dissimilar (*hat, boat*). Looking at Dodd's (1975) Table 2, children chose the correct word about 47% of the time, the rhyming word about 19%, and each of the other two 10% of the time (the remaining 14% were refusals to respond). They were more accurate on adult speech (~94%). Dodd (1975) suggests that children cannot identify their own productions. Still, one might notice that their *predominant* response was the intended word, suggesting they recognized some phonetic similarity to their own forms. In a more recent study, Cooper et al. (2018) asked 30-month-old children and their caregivers to name familiar pictures and later tested them on their own, another child's, their parent's, and another parent's productions of the words in a picture recognition task. Children spent more time looking at the named picture for both of the adults' speech than for both children's speech, including their own. Żygis et al. (2023) looked at children's (ages 2;11-7;11) perception and production of Polish fricatives, which are challenging for children to produce. They found in their picture recognition task, which contrasted a minimal triplet of fricative-containing words, that even the oldest children recognized adult instances of the fricatives more readily than their own productions, suggesting that perception precedes production for these sounds. Finally, Creel et al. (under review) asked children to name 24 pictures in similar-sounding pairs, like Żygis et al. (2023), these were mostly minimal pairs. Children then heard themselves and a control adult name the pictures as they attempted to select a replayed name (e.g. "wing") from a set of four pictures (e.g. *ring, wing, chip, ship*). Children's accuracy on adult speech (90%) exceeded their accuracy on self-speech (78%), though most of the incorrect responses were selections of the similar-named picture (75% of incorrect responses). Thus, across multiple studies, young children between 2.5 and 8 years show better comprehension of adult speech than their own speech.

Interlude: objections to standard account

However, as I have already implied in a very heavy-handed manner, there are a lot of objections one could raise to the standard account that adult speech dominates children's mental representations. One objection is that children are pretty good at child speech, at least at ages 3 years and up: they can be reasonably accurate on child speech as long as they are not required to distinguish highly phonologically similar words. In a number of the studies described in the previous section, children performed well at getting to the right phonological neighborhood.

While children were only 78% accurate on self-speech in Creel et al. (under review), they chose a picture that sounded *similar* to the intended word 94% of the time, and in a preliminary study with dissimilar-named pictures, children reached 90% self-speech accuracy. In Żygis et al.'s (2023) study of fricatives, children performed at high levels of accuracy on their own speech (95%) on non-minimal pair trials. A related literature on children's sensitivity to adult speech mispronunciations (Swingley, 2016 in toddlers; Creel, 2012, in 3-6-year-olds) concurs that small speech alterations to a single sound such as pronouncing *fish* as "fesh" only cause mild confusion, evident in slightly lower pointing accuracy or decreased looks to named pictures, but children usually arrive at the phonologically similar answer.

Another objection to the standard account is that children's weaker comprehension of self-speech under phonologically challenging conditions isn't even the right framing of the question in the first place, given what we know about dialect variation and speech perception. Perhaps a better framing to investigate child speech is that it is akin to *learning a dialect with some mergers*. For example, Californian speakers produce *cot* and *caught* indistinguishably, while speakers in the Northeast and Southeast of the United States still distinguish these vowels. This *cot-caught merger* explains why my California-raised 7-year-old asked "What does stocking mean? Not stocking like on your legs, but like, 'the snow leopard is stocking its prey'?" Perhaps asking children to distinguish their own merged *ring* and *wing* is like asking a Californian to distinguish their own *cot-caught*: it's silly because they are homophones in that dialect, like stock and stalk in my child's dialect. Note that treating these words as homophones means that both alternatives are equally consistent with the input: a Californian won't default to assuming they're hearing (e.g.) cot in all cases. They will likely guess. Similarly, perhaps a child who merges r and w to /w/ will perceive their own attempts at ring and wing not as all being *wing*—as would be the case if their percepts are completely adultlike—but as ambiguous between *ring* and *wing*. I pursue this account of child speech as a type of **dialectal merger awareness**.

A final objection to the standard account is a theoretical one. It is on the hook for explaining how children hear child speech, enjoy listening to it, yet discard this input from their perceptual repertoire. We could certainly come up with seemingly plausible answers. Perhaps children's higher cue noise in speech sound production gives it less weight in perceptual learning. Perhaps there is some sort of cognitive weighting, a la Sumner et al.'s (2014) theory of dialect variation, where more-prestigious dialects get more attention. Perhaps children tend to trust adults more as language informants (e.g. Jaswal & Neely, 2006). A final answer, and my preferred one: perhaps they don't discard child speech at all, but learn it as one of multiple possible variant forms of speech.

2 - Combine across child and adult cue spaces

If we grant that children's representations contain child speech, how does it factor into their mental representations of words and speech sounds? One simple way would be to combine adult and child input (and input from a variety of other speakers) into a unified set of representations, without respect to who is speaking. Of course, one would have to collapse over absolute differences between adult and child f0 and formants and voice quality, and speech categories would have greater variability than for adult-only speech. Taking our running example of r-w

variation into account, there would be, potentially, one set of words (*wing, one, window, wiggle, weather, web*…) that contain a fairly clear /w/ distribution and another set of words (*ring, run, rainbow, ripple, rest, red*…) that contain a broader distribution encompassing some r-like sounds and some w-like sounds (lower left panel of Figure 1). This is not the sort of question that gets addressed in distributional learning studies, which typically consider 1 vs 2 categories or 2 categories that vary in mean difference or variability, but it predicts that, for both child speech and adult speech, listeners might treat clear r tokens as r, but w-like tokens as ambiguous and perhaps w-leaning. It also might predict generally higher levels of uncertainty for child-varying sounds, since the variability of these distributions is larger than for adult speech (Figure 1, upper right).

3 - Speaker-contingent processing

A third possibility that I've already heavily alluded to is dialect switching between child speech (Figure 1, lower right) and adult speech (upper right). On this account, children learn and maintain separate, dialect-like representations of (at minimum) child speech and adult speech. My hypothesis is that this learning occurs over the long term via extensive exposure. This leads them to treat adult sounds as fairly certain, but child speech—particularly sounds that map to two distinct sounds in the adult dialect—as ambiguous. As a specific example, if an adult says "wing," children are pretty certain they are hearing *wing*. However, if they hear a young child say "wing," they may regard it as ambiguous between *wing* and *ring*, because they've experienced many children (potentially including themselves) producing words like *ring* with an initial sound much like /w/. This dialect-sensitive responding would be consistent with a literature on learning of adult dialect variation in real-world settings (K. Johnson et al., 1999; Niedzielski, 1999) and in the lab (Trude & Brown-Schmidt, 2012; Dahan et al., 2008). In those studies, listeners recognize words or sounds *contingent on* the identity or presumed group membership of the speaker. There's even some evidence supportive of second-language or heritage speakers learning their own speech features (Cheung & Babel, 2022; Creel, manuscript in prep; Eger & Reinisch, 2019). My work with child speakers (Creel et al., under review) is consistent with this merger pattern, though the analysis supporting the merger pattern was after the fact and considered only a small subset of the stimuli (50 word pairs that individual children merged). It's not clear from that work how consistent these response patterns are within individual children, and whether these child listeners are responding to child speech generally or self-speech specifically. More on this topic later.

My account is that children learn childlike speech patterns via distributional learning over long time periods. Yet a different route to speaker-contingent processing is that perceivers adapt rapidly to whomever they are listening to, without the need for long-term learning. This would predict, for example, that children who hear a speaker produce a new merger—for example, merging /b/ and /p/ so that both sound like /b/, not a common merger in English—would also show "merger aware" behavior. Rapid adaptation is one possible explanation of the child speech perception pattern in Creel et al. (under review): children hear their own productions, or another child's productions, and then adapt rapidly to that speaker's merger pattern. Adaptation this fast and complete would be an impressive feat, and there is some evidence that young children can adapt somewhat to recently presented accent features (e.g. White & Aslin, 2011). I have my doubts that anything but extensive perceptual learning can lead to children's merger awareness in

kidiolects, as adaptation that rapid and pervasive would seem to allow the possibility of catastrophic forgetting of already learned contrasts. Still, it is likely that *gradient* adaptation can occur from brief exposures, and invoking a familiar dialect set would seem to require some exposure to the speaker to decide to invoke that dialectal set. This leads into broader questions of how many speech variants a listener can differentiate and switch between. Just as an adult might shift flexibly between perceptual sets for speakers from California vs. South Carolina vs. Mexico City vs. Shanghai, can a child oscillate between speakers who merge r/w, split r/w, merge s/sh, split s/sh, and so forth, or do they maintain a more generalized child-speech set of representations? I discuss this further below.

OPEN QUESTIONS TO BETTER UNDERSTAND CHILD PROCESSING OF CHILD SPEECH

Here I discuss some of the limitations of the current literature on perception and production of child speech, focusing on unknown questions. There are far more open questions than I can feasibly address, so I have tried to focus on the most critical ones that have not come up explicitly so far.

Questions about self-speech sensitivity

Whose kidiolect is represented? I've argued that children may form robust perceptual representations of their own speech. It makes sense that one would have pretty fine-grained representations of speakers that one is exposed to the most, such as oneself. Still, maybe children have more general "young child kidiolect" representations that they apply to themselves and other children. Perhaps they represent self and a few other highly familiar children, and use these to generalize to new instances (Kleinschmidt & Jaeger, 2015). Perhaps they represent a small set of common child speech patterns, such as speakers who merge r/w, speakers who split r/w, speakers who merge s/sh, and speakers who split s/sh. These related cases make some slightly differing predictions that are worth testing.

Questions of whether one's own voice is represented mentally are intertwined with a related question: do you have to recognize that it's you speaking in order to hear in kidiolect? If you just have to recognize that it's child speech, that's not difficult for children (Creel & Jiménez, 2012; see also unpublished 3-alternative man/woman/child data from my lab, shown in Table 1). If you have to recognize that you're hearing yourself specifically, that might seem more challenging, but a small amount of research suggests that children can do so. Strömbergsson (2013) reports that children as young as 4 years of age are good at recognizing self-speech vs. others' speech, performing at 76% accuracy in selecting self out of four different voices even at a 1–2-week delay (see also Cooper et al., 2018, p. 18, which states that 30-month-olds were "above chance" at recognizing which voice was their own). In any case, it may not matter whether you recognize yourself so much as hearing phonetic properties that are similar to your own: Eger and Reinisch (2019) report that female German speakers of English recognized words in their own speech better even though recordings had been morphed to sound like male speakers. My operative

hypothesis is that knowing it's you is not as relevant as the perceptual self-likeness of the speech characteristics, though of course future work could disconfirm this hypothesis.

Table 1. Unpublished data on child and adult listeners' rates of identifying child, adult female, and adult male speakers in a three-picture selection task. Proportion correct is on diagonals (bolded); confusion errors are the off-diagonal elements (not bolded).

	actual speaker	**reported speaker**		
		child	**female**	**male**
adult listener (n=24)	child	**1.00**	0.00	0.00
	female	0.10	**0.90**	0.00
	male	0.02	0.04	**0.94**
child listener (n=24)	child	**0.81**	0.11	0.08
	female	0.19	**0.73**	0.09
	male	0.07	0.10	**0.82**

Is perception impaired when production is going on at the same time? One possible reason that children might have adultlike perception but childlike production is that, while speaking, they conflate their *intended production* with their *actual production.* If so, they might *think* they are producing more accurately in the moment than they actually are. This has been observed in pitch production in professional singers (Vurma & Ross, 2006), who are better at judging their own pitch accuracy in after the fact recordings than immediately after the moment of production. Even more relevant, 4-6-year-old children with speech sound substitutions like /t/ for /k/ (Strömbergsson et al., 2014) are less sensitive to their own atypical productions in real time than hearing their own productions after a delay, or hearing another child's productions. Still, it is not the case that one completely blots out one's own auditory production. Anecdotally, you can hear yourself accidentally misspeak or produce an unexpected voice quality (e.g. due to laryngitis). Further, studies presenting altered auditory feedback show that listeners update their vocal output to match auditory intent (e.g. Houde & Jordan, 1998; see Coughler et al., 2022, for a developmental review), suggesting listeners constantly monitor whether their productions match expectations. More careful work in child populations with and without speech sound production difficulties is needed.

How do children sociolinguistically or prescriptively process child speech? I've been framing the problem as one of recognizing words, but spoken language is also a source of information about the speaker. This means children may perceive aspects of listener identity or make value judgments based on others' speech or their own speech. We know that young children are sensitive to accent variation and can make simple sociolinguistic judgments (Jones et al., 2017; Kinzler et al., 2007; see Creel, 2018, for a slower developmental trajectory for accent sensitivity). Other studies, mostly in the speech sciences literature, have addressed what we might call prescriptive norms or normativity judgments. For example, Dugan et al. (2019) asked older children (8-16 years) with and without residual speech sound errors (atypicalities) on r whether instances of adultlike and misarticulated r were "correct" or not. They found that both child groups showed milder but non-zero categorization slopes than adult listeners, that is, they

were less certain about correctness. Dugan et al. (2019) did not test their child listeners on adult speech to assess whether the slope mildness was specific to child speech, nor did they test child listeners with residual speech sound errors with their own speech. Still, the finding that older children can make these prescriptive judgments to some degree is interesting. More research needs to be done in younger children, though one wonders if such judgments are too metalinguistic for them to grasp.

Questions about the role of child speech in recognition and learning

How much do studies of minimal pair word recognition in child speech tell us about recognition and processing more generally? Several studies have examined the somewhat unnatural situation of testing children on identification of minimal pairs such as *ring* and *wing* or *seat* and *sheet*. These are a special case, an acid test of phonetic discrimination. Someone might reasonably object that in everyday comprehension, you don't constantly have to distinguish words by single phonological features. Most of the time, the set of competitor words that are discourse-relevant, syntactically licensed, and semantically similar aren't going to sound similar to the intended word. Despite this, principles from minimal-pair studies likely apply in more general cases.

First, minimal pair recognition may be especially sensitive due to its difficulty, but there doesn't need to be strong phonological competition present for unusual productions of a word to slow down recognition. Several studies report that mispronunciations which cross phonetic boundaries are recognized less rapidly and (in cases where children point or touch a screen) somewhat less accurately by children ages 15 months through 6+ years. Children are slower to look at pictures whose names are mispronounced by an adult speaker (e.g. "fesh" for *fish* or "vaby" for *baby*; ages 15-24 months: Swingley & Aslin, 2000, 2002, White & Morgan 2008; 24-30 months: Swingley, 2016; 3-6-year-olds: Creel, 2012). When presented with (7-year-old) child mispronunciations vs. correct pronunciations, Bernier and White (2019a, 2019b) report that 22-month-olds looked less to named pictures, and Krueger and colleagues (Krueger et al., 2018; Krueger & Storkel, 2020) report that 4-6-year-old children chose mispronounced pictures less often than correctly pronounced ones, though they chose frequent child(like) pronunciations more often than less frequent ones.

Second, while needing to distinguish between minimal pairs in a real-world context is not very common, *somewhat*-similar words can and do occur in real-world contexts. That is, it's somewhat more frequent to experience *partial* overlap between words, which may still present competition for recognition, such as *red* and *web,* which only overlap in one sound (the vowel) but contain related consonants. These more common situations of partial overlap may show increased in phonological competition when child speech properties move one spoken word closer to one or more other words. For example, suppose a child attempts to say "red" and it emerges as "wed." How much will a listener *temporarily* consider web, slowing down recognition of the intended word *red*? If an adult produces "we-," there are good odds that it's really *web* and not *red*. If a young child produces "we-," the odds are much more ambiguous, with *red* ("wed") a distinct possibility. This sort of temporary ambiguity might lead to slower

comprehension of some child speakers even in the absence of strong phonological competition like that in minimal pairs.

How does *large-scale* child speech variability affect learning of new words? Children not only need to recognize words in child speech, they also need to learn them. Does variability between child speakers, or across a mixture of child and adult speakers, affect word learning? As mentioned earlier, Frye and Creel (2022) as well as Creel (2014), asked children to learn phonemically variable object labels (e.g. Object 1 can be called both "deev" and "div") and found only mild evidence of increased learning difficulty, similar to Muench and Creel (2013) in adult learners. This suggests that learning from phonemically variable input, as might occur when learning from both adults and children, may not be all that difficult, and it appears to be different than the process of learning phonologically unrelated synonyms, which is difficult for both young children (Frye & Creel, 2022) and young adults (Muench & Creel, 2013). A more uncertain question is whether learners can keep track of who produces a particular word variant. Suppose you teach a child that an adult calls Object 1 *wutch* and Object 2 *ruvv*. Can they then infer that a child who describes a stuffed toy as a "gween wabbit" will call Object 2 *wuvv*, or will they temporarily think the child is in the process of saying *wutch*? If they readily recognize *wuvv*, this would imply that children understand phonological relationships across kidiolects and adultlects, which I might call a type of phonological translation (Oller et al., 1998). Oller et al. (1998) coined this term in the sense of children converting word forms between Spanish and English phonological systems, but the underlying principle is the same.

Most of this discussion revolves around variation such that a child seems to produce a different speech sound than an adult. But children also show small-scale pronunciation variability. Is this enough to slow recognition or stymy learning? For example, are very r-like child productions recognized faster or more accurately than moderately r-like or ambiguous productions? We might also ask how such small-scale variability affects novel word learning. While one line of thought is that variability will facilitate learning (Galle et al., 2014; Rost & McMurray, 2009, 2010), another is that it might cloud recognition slightly (Frye & Creel, 2022).

A final question concerns production. How does a child use a combination of child and adult input to decide how to say things? That is, how do child learners decide what speaker(s) they should sound like? This is not a trivial question, though again the default assumption is that young children mostly learn speech patterns from caregiver input (see, e.g., Roberts, 2004). This may be based on presumptions that children's main language exposure is in the home up until primary school age. This may be more true of children who have a parent with a flexible enough schedule to visit an infant lab, but less true of the many children who spend time in day care or preschool settings, or children in more diverse cultures with more child speakers (Cristià et al., 2023; Loukatou et al., 2022; Schick & Stoll, 2025). One possibility is that relative proportions of child vs. adult input, and the skill levels of child speakers, contribute to the child's productions.

Concluding thoughts

I'm hopeful that in the coming years, researchers will begin broadening the scope of their research on spoken input to children. There's a lot of speech out there, and it would be a shame

not to take advantage of it. I'm also hopeful that child language researchers and scholars of dialect variation will interact more, as there are multiple fruitful areas of overlap in understanding age-related speech variation and social group-related variation. Finally, input from child speakers contains nonadultlike properties at multiple levels of language structure, so there are additional interesting questions about semantic and syntactic effects of child input that equally deserve exploration.

References

Ahn, E. P., Levow, G.-A., Wright, R. A., & Chodroff, E. (2023). An Outlier Analysis of Vowel Formants from a Corpus Phonetics Pipeline. *Proceedings of Interspeech.*

Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science*, *8*(2), 199–209. https://doi.org/10.1111/j.1467-7687.2005.00406.x

Bergelson, E., Soderstrom, M., Schwarz, I. C., Rowland, C. F., Ramírez-Esparza, N., Hamrick, L. R., Marklund, E., Kalashnikova, M., Guez, A., Casillas, M., Benetti, L., van Alphen, P., & Cristia, A. (2023). Everyday language input and production in 1,001 children from six continents. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(52), 1–12. https://doi.org/10.1073/pnas.2300671120

Bernier, D. E., & White, K. S. (2019). Toddlers process common and infrequent childhood mispronunciations differently for child and adult speakers. *Journal of Speech, Language, and Hearing Research*, *62*(11), 4137–4149. https://doi.org/10.1044/2019_JSLHR-H-18-0465

Bernier, D. E., & White, K. S. (2019). Toddlers' sensitivity to phonetic detail in child speech. *Journal of Experimental Child Psychology*, *185*, 128–147. https://doi.org/10.1016/j.jecp.2019.04.021

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

Brekelmans, G., Lavan, N., Saito, H., Clayards, M. A., & Wonnacott, E. (2022). Does high variability training improve the learning of non-native phoneme contrasts over low variability training? A replication. *Journal of Memory and Language*, *126*(January 2021), 104352. https://doi.org/10.1016/j.jml.2022.104352

Bulgarelli, F. (2026). Beyond adult speech: why research on language development should consider speech from children. *Child Development Perspectives*. https://doi.org/10.1093/cdpers/aadaf004

Cheung, S., & Babel, M. (2022). The own-voice benefit for word recognition in early bilinguals. *Frontiers in Psychology*, *13*(September), 1–14. https://doi.org/10.3389/fpsyg.2022.901326

Cleland, J., Scobbie, J. M., Heyde, C., Roxburgh, Z., & Alan, A. (2017). Covert contrast and covert errors in persistent velar fronting. *Clinical Linguistics & Phonetics*, *31*(1), 35–55. https://doi.org/10.1080/02699206.2016.1209788

Cooper, A., Fecher, N., & Johnson, E. K. (2018). Toddlers' comprehension of adult and child talkers: Adult targets versus vocal tract similarity. *Cognition*, *173*, 16–20. https://doi.org/10.1016/j.cognition.2017.12.013

Coughler, C., Quinn de Launay, K. L., Purcell, D. W., Oram Cardy, J., & Beal, D. S. (2022). Pediatric Responses to Fundamental and Formant Frequency Altered Auditory Feedback: A Scoping Review. *Frontiers in Human Neuroscience*, *16*. https://doi.org/10.3389/fnhum.2022.858863

Cox, C., Bergmann, C., Fowler, E., Keren-Portnoy, T., Roepstorff, A., Bryant, G., & Fusaroli, R. (2023). A systematic review and Bayesian meta-analysis of the acoustic features of infant-directed speech. *Nature Human Behaviour*, *7*(1), 114–133. https://doi.org/10.1038/s41562-022-01452-1

Creel, S. C. (2012). Phonological similarity and mutual exclusivity: on-line recognition of atypical pronunciations in 3-5-year-olds. *Developmental Science*, *15*(5), 697–713. https://doi.org/10.1111/j.1467-7687.2012.01173.x

Creel, S. C. (2014). Impossible to _gnore: Word-form inconsistency slows preschool children's word-learning. *Language Learning and Development*, *10*(1), 68–95. https://doi.org/10.1080/15475441.2013.803871

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

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., & Frye, C. I. (2024). Minimal gains for minimal pairs: Difficulty in learning similar-sounding words continues into preschool. *Journal of Experimental Child Psychology*, *240*, 105831. https://doi.org/10.1016/j.jecp.2023.105831

Creel, S. C., Vu, A., & McCrary Kambourakis, K. (manuscript under review). Secret kid language: Preschoolers' perception precedes production, but they memorize mergers.

Cristia, A., Gautheron, L., & Colleran, H. (2023). Vocal input and output among infants in a multilingual context: Evidence from long-form recordings in Vanuatu. *Developmental Science*. https://doi.org/10.1111/desc.13375

Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*, *108*, 710–718. https://doi.org/10.1016/j.cognition.2008.06.003

Dodd, B. (1975). Children's understanding of their own phonological forms. *Quarterly Journal of Experimental Psychology*, *27*(2), 165–172. https://doi.org/10.1080/14640747508400477

Dugan, S. H., Silbert, N., McAllister, T., Preston, J. L., Sotto, C., & Boyce, S. E. (2019). Modelling category goodness judgments in children with residual sound errors. *Clinical Linguistics and Phonetics*, *33*(4), 295–315. https://doi.org/10.1080/02699206.2018.1477834

Eger, N. A., & Reinisch, E. (2019). The impact of one's own voice and production skills on word recognition in a second language. *Journal of Experimental Psychology: Learning Memory and Cognition*, *45*(3), 552–571. https://doi.org/10.1037/xlm0000599

Erskine, M. E., & Seidl, A. (2026). Children may learn words best from other children, not adults. *Language Learning and Development*. https://doi.org/10.1080/15475441.2026.2614444

Ferjan Ramírez, N. (2022). Fathers' infant-directed speech and its effects on child language development. In *Language and Linguistics Compass* (Vol. 16, Issue 1). John Wiley and Sons Inc. https://doi.org/10.1111/lnc3.12448

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*(3), 279–293. https://doi.org/10.1016/0163-6383(87)90017-8

Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Delle Luche, C., Durrant, S., White, L., Goslin, J., & Vihman, M. M. (2016). British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition*, *148*, 1–9. https://doi.org/10.1016/j.cognition.2015.12.004

Floor, P., & Akhtar, N. (2006). Can 18-Month-Old Infants Learn Words by Listening In on Conversations? *Infancy*, *9*(3), 327–339. https://doi.org/10.1207/s15327078in0903

Frye, C. I., & Creel, S. C. (2022). Perceptual flexibility in word learning: Preschoolers learn words with speech sound variability. *Brain and Language*, *226*(January), 105078. https://doi.org/10.1016/j.bandl.2022.105078

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. https://doi.org/10.1080/15475441.2014.895249

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A Theoretical Investigation of Reference Frames for the Planning of Speech Movements. *Psychological Review*, *105*(4), 611–633. https://doi.org/10.1037/0033-295X.105.4.611-633

Hearnshaw, S., Baker, E., & Munro, N. (2019). Speech perception skills of children with speech sound disorders: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research*, *62*(10), 3771–3789. https://doi.org/10.1044/2019_JSLHR-S-18-0519

Hippe, L., & Ferjan Ramírez, N. (2022). Sibs and Bibs: Older Siblings and Infant Vocabulary Development. *Proceedings of the 46th Boston University Conference on Language Development*.

Hoffman, P. R., Stager, S., & Daniloff, R. G. (1983). Perception and production of misarticulated /r/. *Journal of Speech and Hearing Disorders*, *48*, 210–215.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, *279*, 1213–1216.

Idemaru, K., & Holt, L. L. (2013). The developmental trajectory of children's perception and production of English /r/-/l/. *Journal of the Acoustical Society of America*, *4232*. https://doi.org/10.1121/1.4802905

Jaswal, V. K., & Neely, L. a. (2006). Adults don't always know best. *Psychological Science*, *17*(9), 757–758. https://doi.org/10.1111/j.1467-9280.2006.01778.x

Johnson, E. K., & White, K. S. (2025). Advancements in phonetics in the 21st century: Infant speech development. *Journal of Phonetics*, *111*. https://doi.org/10.1016/j.wocn.2025.101425

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics*, *27*(4), 359–384. https://doi.org/10.1006/jpho.1999.0100

Jones, Z., Yan, Q., Wagner, L., & Clopper, C. G. (2017). The development of dialect classification across the lifespan. *Journal of Phonetics*, *60*, 20–37. https://doi.org/10.1016/j.wocn.2016.11.001

Kempe, V., Ota, M., & Schaeffler, S. (2024). Does child-directed speech facilitate language development in all domains? A study space analysis of the existing evidence. In *Developmental Review* (Vol. 72). Elsevier Inc. https://doi.org/10.1016/j.dr.2024.101121

Kent, R. D., & Forner, L. L. (1980). Speech segment durations in sentence recitations by children and adults. In *Journal of Phonetics* (Vol. 8).

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, *104*(30), 12577–12580. https://doi.org/10.1073/pnas.0705345104

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust Speech Perception: Recognize the Familiar, Generalize to the Similar, and Adapt to the Novel. *Psychological Review*, *122*(2), 148–203.

Krueger, B. I., & Storkel, H. L. (2020). Children's Response Bias and Identification of Misarticulated Words. *Journal of Speech Language and Hearing Research*, *63*(January), 259–273.

Krueger, B. I., Storkel, H. L., & Minai, U. (2018). The Influence of Misarticulations on Children's Word Identification and Processing. *Journal of Speech Language and Hearing Research*, *61*(April), 820–836. https://doi.org/10.1044/2017_JSLHR-S-16-0379

Kuhl, P. K., Andruski, J. E., Chistovich, I. a, Chistovich, L. a, Kozhevnikova, E. v, Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science (New York, N.Y.)*, *277*(5326), 684–686. http://www.ncbi.nlm.nih.gov/pubmed/9235890

Kuhl, P. K., Tsao, F., & Liu, H. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096–9101.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, *105*, 1455–1468.

Levy, H., & Hanulíková, A. (2023). Spot It and Learn It! Word Learning in Virtual Peer-Group Interactions Using a Novel Paradigm for School-Aged Children. *Language Learning*, *73*(1), 197–230. https://doi.org/10.1111/lang.12520

Li, F., Edwards, J., & Beckman, M. E. (2009). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics*, *37*, 111–124. https://doi.org/10.1016/j.wocn.2008.10.001

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94*(3 Pt 1), 1242–1255. http://www.ncbi.nlm.nih.gov/pubmed/8408964

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /1/: A first report for publication. *Journal of the Acoustical Society of America*, *89*(2), 874–886.

Loukatou, G., Scaff, C., Demuth, K., Cristia, A., & Havron, N. (2022). Child-directed and overheard input from different speakers in two distinct cultures. *Journal of Child Language*, *49*(6), 1173–1192. https://doi.org/10.1017/S0305000921000623

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: a study of voice onset time in word-initial stop consonants. *Journal of Child Language*, *7*, 41–74.

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristià, A. (2015). Mothers speak less clearly to infants than to adults: A comprehensive test of the hyperarticulation hypothesis. *Psychological Science*, *26*(3), 341–347. https://doi.org/10.1177/0956797614562453

Masapollo, M., Polka, L., & Lucie, M. (2016). When infants talk, infants listen: pre-babbling infants prefer listening to speech with infant vocal properties. *Developmental Science*, *19*(2), 318–328. https://doi.org/10.1111/desc.12298

McAllister Byun, T., & Tiede, M. (2017). Perception-production relations in later development of American English rhotics. *PLoS ONE*, 1–22. https://doi.org/10.1371/journal.pone.0172022

McAllister Byun, T., Harel, D., Halpin, P. F., & Szeredi, D. (2016). Deriving gradient measures of child speech from crowdsourced ratings. *Journal of Communication Disorders*, *64*, 91–102. https://doi.org/10.1016/j.jcomdis.2016.07.001

McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence? *Cognition*, *129*(2), 362–378. https://doi.org/10.1016/j.cognition.2013.07.015

Muench, K. L., & Creel, S. C. (2013). Gradient phonological inconsistency affects vocabulary learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1585–1600. https://doi.org/10.1037/a0032862

Munson, B., Edwards, J., Schellinger, S. K., Beckman, M. E., & Meyer, M. K. (2010). Deconstructing phonetic transcription: Covert contrast, perceptual bias, and an extraterrestrial view of Vox Humana. *Clinical Linguistics & Phonetics*, *24*(4–5), 245–260. https://doi.org/10.3109/02699200903532524

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, *18*(1), 62–85. https://doi.org/10.1177/0261927X99018001005

Ochs, E. (1982). Talking to children in Western Samoa. *Language and Society*, *11*, 77–104. https://doi.org/https://doi.org/10.1017/S0047404500009040

Oller, D. K., & Eilers, R. E. (1988). The Role of Audition in Infant Babbling. *Child Development*, *59*(2), 441–449.

Oller, D. K., Cobo-Lewis, A. B., & Eilers, R. E. (1998). Phonological translation in bilingual and monolingual children. *Applied Psycholinguistics*, *19*, 259–278. https://doi.org/10.1017/S0142716400010067

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472. https://doi.org/10.1121/1.3593366

Roberts, J. (2004). Child language variation. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 333–348). Blackwell. https://doi.org/10.1002/9780470756591

Roepke, E., & Brosseau-Lapré, F. (2019). Perception of sibilants by preschool children with overt and covert sound contrasts. *Journal of Speech, Language, and Hearing Research*, *62*(10), 3763–3770. https://doi.org/10.1044/2019_JSLHR-S-19-0127

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339–349. https://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. https://doi.org/10.1111/j.1532-7078.2010.00033.x

Salo, V. C., Rowe, M. L., Leech, K. A., & Cabrera, N. J. (2016). Low-income fathers' speech to toddlers during book reading versus toy play. In *Journal of Child Language* (Vol. 43, Issue 6, pp. 1385–1399). Cambridge University Press. https://doi.org/10.1017/S0305000915000550

Schellinger, S. K., Munson, B., & Edwards, J. (2017). Gradient perception of children's productions of /s/ and /θ/: A comparative study of rating methods. *Clinical Linguistics and Phonetics*, *31*(1), 80–103. https://doi.org/10.1080/02699206.2016.1205665

Schick, J., & Stoll, S. (2025). Children Learn Best From Their Peers: The Crucial Role of Input From Other Children in Language Development. *Open Mind*, *9*, 665–676. https://doi.org/10.1162/opmi_a_00198

Schick, J., Daum, M. M., & Stoll, S. (2025). Input to the Language Learning Infant: The Impact of Other Children. *Developmental Science*, *28*(5). https://doi.org/10.1111/desc.70045

Schmale, R., & Seidl, A. (2009). Accommodating variability in voice and foreign accent: flexibility of early word representations. *Developmental Science*, *12*(4), 583–601. https://doi.org/10.1111/j.1467-7687.2009.00809.x

Schmale, R., Cristià, A., Seidl, A., & Johnson, E. K. (2010). Developmental Changes in Infants' Ability to Cope with Dialect Variation in Word Recognition. *Infancy*, *15*(6), 650–662. https://doi.org/10.1111/j.1532-7078.2010.00032.x

Scobbie, J. M., Gibbon, F., Hardcastle, W. J., & Fletcher, P. (1996). Covert contrast as a stage in the acquisition of phonetics and phonology. In J. M. Scobbie (Ed.), *Queen Margaret College Working Papers in Speech and Language Sciences* (Vol. 1).

Shriberg, L. D. (2009). Childhood speech sound disorders: From postbehaviorism to the postgenomic era. In R. Paul & P. Flipsen (Eds.), *Speech sound disorders in children* (pp. 1–34). Plural Publishing.

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *Journal of the Acoustical Society of America*, *125*(5), 3306–3316. https://doi.org/10.1121/1.3101452

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa Articulation Norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, *55*(November 1990), 779–798.

Strömbergsson, S. (2013). Children's recognition of their own recorded voice: Influence of age and phonological impairment. *Clinical Linguistics & Phonetics*, *27*(1), 33–45. https://doi.org/10.3109/02699206.2012.735744

Strömbergsson, S., Wengelin, Å., & House, D. (2014). Children's perception of their synthetically corrected speech production. *Clinical Linguistics & Phonetics*, *28*(6), 373–395. https://doi.org/10.3109/02699206.2013.868928

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology*, *4*, 1–13. https://doi.org/10.3389/fpsyg.2013.01015

Swingley, D. (2016). Two-Year-Olds Interpret Novel Phonological Neighbors as Familiar Words. *Developmental Psychology*, *52*(7), 1011–1023.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*, 147–166.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, *13*(5), 480–484. https://doi.org/10.1111/1467-9280.00485

Taylor, M., Cartwright, B. S., & Bowden, T. (1991). Perspective Taking and Theory of Mind: Do Children Predict Interpretive Diversity as a Function of Differences in Observers' Knowledge? *Child Development*, *62*(6), 1334–1351.

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952–981. https://doi.org/10.1080/01690960903498424

Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes*, *27*(7–8), 979–1001. https://doi.org/10.1080/01690965.2011.597153

van Heugten, M., & Johnson, E. K. (2014). Learning to contend with accents in infancy: Benefits of brief speaker exposure. *Journal of Experimental Psychology: General*, *143*(1), 340–350. https://doi.org/10.1037/a0032192

van Heugten, M., & Johnson, E. K. (2015). Toddlers' Word Recognition in an Unfamiliar Regional Accent: The Role of Local Sentence Context and Prior Accent Exposure. *Language and Speech*, 1–11. https://doi.org/10.1177/0023830915600471

van Heugten, M., Krieger, D. R., & Johnson, E. K. (2015). The Developmental Trajectory of Toddlers' Comprehension of Unfamiliar Regional Accents. *Language Learning and Development*, *11*, 41–65. https://doi.org/10.1080/15475441.2013.879636

Vaughn, C., Baese-Berk, M. M., & Idemaru, K. (2018). Re-Examining Phonetic Variability in Native and Non-Native Speech. *Phonetica*, *76*, 327–258. https://doi.org/https://doi.org/10.1159/000487269

Vurma, A., & Ross, J. (2006). Production and perception of musical intervals. *Music Perception*, *23*(4), 331–344.

Wade, T., Jongman, A., & Sereno, J. (2007). Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica*, *64*(2–3), 122–144. https://doi.org/10.1159/000107913

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152. https://doi.org/10.1177/0956797613488145

White, K. S., & Aslin, R. N. (2011). Adaptation to novel accents by toddlers. *Developmental Science*, *14*(2), 372–384. https://doi.org/10.1111/j.1467-7687.2010.00986.x

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

Whiteside, S. P., Dobbin, R., & Henry, L. (2003). Patterns of variability in voice onset time: a developmental study of motor speech skills in humans. *Neuroscience Letters*, *347*, 29–32. https://doi.org/10.1016/S0304-3940(03)00598-6

Xie, X., & Jaeger, T. F. (2020). Comparing non-native and native speech: Are L2 productions more variable? *The Journal of the Acoustical Society of America*, *147*(5), 3322–3347. https://doi.org/10.1121/10.0001141

Xie, X., Liu, L., & Jaeger, T. F. (2021). Cross-Talker Generalization in the Perception of Nonnative Speech: A Large-Scale Replication. *Journal of Experimental Psychology: General*, *150*(11), 22–56. https://doi.org/10.1037/xge0001039

Zettersten, M., Cox, C., Bergmann, C., Tsui, A. S. M., Soderstrom, M., Mayor, J., Lundwall, R. A., Lewis, M., Kosie, J. E., Kartushina, N., Fusaroli, R., Frank, M. C., Byers-Heinlein, K., Black, A. K., & Mathur, M. B. (2024). Evidence for Infant-directed Speech Preference Is Consistent Across Large-scale, Multi-site Replication and Meta-analysis. *Open Mind*, *8*, 439–461. https://doi.org/10.1162/opmi_a_00134

Zlatin, M. A., & Koenigsknecht, R. A. (1976). Development of the voicing contrast: A comparison of voice onset time in stop perception and production. *Journal of Speech and Hearing Research*, *19*, 93–111.

Żygis, M., Pape, D., Jaskuła, M., & Koenig, L. L. (2023). Do children better understand adults or themselves? An acoustic and perceptual study of the complex sibilant system of Polish. *Journal of Phonetics*, *100*. https://doi.org/10.1016/j.wocn.2023.101227