Impacts of Acoustic-Phonetic Variability on Perceptual Development for Spoken Language: A Review

Carolyn Quam[1],
Department of Speech and Hearing Sciences, Portland State University, USA

Sarah C. Creel
Department of Cognitive Science, University of California, San Diego, USA

Abstract

This article reviews research on when acoustic-phonetic variability facilitates, inhibits, or does not impact perceptual development for spoken language, to illuminate mechanisms by which variability aids learning of language sound patterns. We first summarize structures and sources of variability. We next present proposed mechanisms to account for how and why variability impacts learning. Finally, we review effects of variability in the domains of speech-sound category and pattern learning; word-form recognition and word learning; and accent processing. Variability can be helpful, harmful, or neutral depending on the learner's age and learning objective. Irrelevant variability can facilitate children's learning, particularly for early learning of words and phonotactic rules. For speech-sound change detection and word-form recognition, children seem either unaffected or impaired by irrelevant variability. At the same time, inclusion of variability in training can aid generalization. Variability between accents may slow learning—but with the longer-term benefits of improved comprehension of multiple accents. By highlighting accent as a form of acoustic-phonetic variability, and considering impacts of dialect prestige on children's learning, we hope to contribute to a better understanding of how exposure to multiple accents impacts language development and may have implications for literacy development.

Introduction

Acoustic-phonetic variability is rampant in infants' early language experience, yet infants somehow learn words and categories and patterns of sounds. The question of how we experience constancy despite fluctuation in sensory input stems from at least 500 BCE in remarks by Heraclitus. Heraclitus' ideas have been interpreted in two ways: Plutarch gleaned the idea that "it is not possible to step twice into the same river" (Graham, 2019), suggesting environmental variability is a challenge to learning and processing. More recent interpretations posit that constancy and change are fundamentally linked and that variability is a precondition for constancy (Graham, 2019; Marcovich, 1966). Tension between variability-as-nuisance and variability-as-fundamental pervades research on spoken-language development. For modern scientists examining learning by humans and machines, lack of invariance is a perpetual challenge. Nonetheless, some recent work

[1]Corresponding author. Portland State University Speech and Hearing Sciences, PO Box 751, Portland, OR 97207-0751, USA. 1-503-725-3558. cquam@pdx.edu.

suggests acoustic-phonetic variability lends stability to children's representations of words and sounds.

This article reviews research on when acoustic-phonetic variability facilitates, inhibits, or does not impact perceptual development for spoken language, elucidating mechanisms by which variability helps form robust sound categories. We review different sources and structures of variability; proposed mechanisms for how variability impacts language learning; and effects of variability in the domains of speech-sound category and pattern learning, word-form recognition and word learning, and accent processing. We hope that highlighting accent as a form of acoustic-phonetic variability, and considering impacts of dialect prestige on accent processing, will help illuminate how exposure to multiple accents impacts language development and may impact literacy development.

In the context of recent calls to diversify STEM disciplines (e.g., #ShutDownSTEM; Chen, 2020), we aim to critically evaluate our field, consciously reduce citation bias, and call for increased diversity in this research area, in researchers, topics, and participants. Much of basic research on perceptual development for spoken language in the US has been conducted by and with predominantly White, educated people (Heinrich, Heine, & Norenzayan, 2010) speaking White American English (WAE, also sometimes referred to as Mainstream or Standard American English; Wolfram, 2007; Y. Holt & Rangarathnam, 2018). Research on other dialects often treats WAE as a standard from which other forms are derived, rather than one of multiple variants. This bias limits generalizability and does a disservice to the diverse population who should benefit from research (DeGraffe, 2018; see also DeGraffe, 2005). Yet the limited research on other dialects that does exist has generated crucial insights into basic-science questions about effects of input variability on language learning, which are intimately connected to important applied topics, such as reducing racial disparities in literacy outcomes (Washington & Craig, 2000; Thompson, Craig, & Washington, 2004) and speech-language evaluation (e.g., Moland, 2011). Accordingly, we highlight work by and with those from minoritized groups, including work by scholars identifying as Black, Indigenous, or People of Color (BIPOC), focusing in particular on African American English (AAE). We aim to draw attention to high-quality, relevant research, some of which has been underappreciated and under-cited in proportion to its value.

We also emphasize strongly that future work connecting basic research on perceptual learning for spoken language with topics such as racial disparities in literacy outcomes must center perspectives of children and communities, working in careful partnership with—and ideally led by—communities of color, given their historic exploitation in research (Ndlovu-Gatsheni, 2019; Patel, 2015). The community-based participatory research framework provides a roadmap for such partnerships (Israel, Schulz, Parker, & Becker, 2001; Holkup, Tripp-Reiner, Salois, & Weinert, 2004). Future research should also consciously apply a strengths-based rather than deficits-based lens (Dudley-Marling & Lucas, 2009; Michaels, 2013; Avineri et al., 2015; Baugh, 2017; Cole, 2013; Sperry, Sperry, & Miller, 2019).

Sources and Structures of Variability

Sources of variability.

Acoustic-phonetic variation comes from a variety of sources. One source is contrastive differences between words and sounds (i.e., those the language uses to differentiate meanings), and this type of variability is critical input to distributional learning (Maye, Werker, & Gerken, 2002; Vallabha et al., 2007). For example, sounds like *b* and *p* differ in voice-onset time (VOT) and pitch. Many other sources of variability are non-contrastive. Elman and McClelland (1984, 1986) distinguished two types of non-contrastive variability. The first, which they aimed to account for in the TRACE model of speech perception, was "lawful" sources of variability, from, e.g., phonological rules or segmental context, e.g., impacts of preceding consonants on acoustic-phonetic realizations of speech sounds.

Elman and McClelland (1986) distinguished lawful variability from such variability as "cross-speaker and within-speaker token-to-token variability." The TRACE model did not attempt to account for variability that was not lawful, as it is more difficult to adapt to systematically (as it varies more randomly). Nevertheless, its prevalence means it must still be accommodated by learners. Realizations of sounds differ *between talkers*, in part due to biological features including age (Schötz & Müller, 2007), vocal-tract size, health, and articulatory strategies (Johnson, Ladefoged, & Landau, 1993). Between-talker differences influence talker-voice characteristics (e.g., absolute formant frequencies) and can impact word learning and recognition.

Within-talker variability can also lead to different realizations of a sound from instance to instance. It can be due to myriad factors including changes in health status, fatigue (Greeley et al., 2007), circadian rhythm (Heald & Nusbaum, 2015), cognitive load (Segbroeck et al., 2014; Yap, Epps, Choi, & Ambikairajah, 2010), affective state (Cowie & Cornelius, 2003; Barrett & Paus, 2002), and recent linguistic experience (Evans & Iverson, 2007). Another source of not-lawful variability is essentially random variability that pervades speech, such that the same speaker using the same accent, register, affect, and sentence context will not produce identical realizations of a sound twice.

Finally, individual speakers also differ systematically in socially conditioned features (e.g., accent, sociolect). Accent variability occurs due to regional and social variation in language sound patterns (for example, AAE vs. WAE, or Chicano English vs. California Anglo English; Guerrero, 2014). Speakers can also employ distinct speech registers (such as formal vs. informal). It is not clear whether Elman and McClelland would consider differences across accents, sociolects, and registers as lawful or not.

Structures of variability.

Variability can be structured in a number of ways. Four particular structures of variability appear in Figure 1, which uses VOT and second-formant frequency as example dimensions. Table 1 summarizes experimental and real-world examples of each variability structure. Variability can be *patterned*, i.e., organized by modes or clusters. VOT, which differentiates sounds like English *b* and *p* at syllable onset, patterns bimodally (e.g., Lisker & Abramson,

1964). Alternatively, variability can be (seemingly) *random noise* (unimodally or uniformly distributed).

Variability can either contribute or not contribute to distinguishing words. Some researchers use the terms relevant vs. irrelevant somewhat synonymously with contrastive vs. non-contrastive, but this the use of the terms relevant and irrelevant involves at least two presuppositions: (1) that learners' goal is to distinguish words, not to (say) use coarticulatory cues to speed up word recognition, or to differentiate parents' voices to identify the talker (e.g., the top vs. bottom two clusters in Figure 1, Panel 3); (2) that speech-sound variability is uncorrelated with non-segmental variability. Even accepting the first presupposition, that learners are learning words, the second is sometimes wrong, as experimental or natural situations may result in correlations between speech-sound and (e.g.) talker variability.

Instead of conflating contrastive and relevant variability, we use 'relevant' and 'irrelevant' more synonymously with 'lawful' (Elman & McClelland, 1986) vs. 'unlawful,' though we note that relevance must be defined relative to the task. Contrastive variability is lawful, and typically relevant. For example, assuming learners are trying to differentiate words, VOT variability is relevant (Figure 1, x-axes). Non-contrastive variability can be either lawful (and typically relevant), as for variability due to coarticulation and allophonic variation, or unlawful (and typically irrelevant), as for token-by-token variability across and within talkers. For example, in settings without talker variability—language-learning classes with one instructor, households with one caregiver, or experimental contexts—unlawful, irrelevant variation still occurs between instances of a speech sound (Figure 1, Panel 1).

Two additional structures of variability were not considered by Elman and McClelland (1984, 1986), but an argument could potentially be made that they are lawful, as they are patterned. First, in settings with, e.g., a small number of talkers (Figure 1, Panel 3), talker variability can be patterned (as contrasted with settings with a larger number of talkers; Figure 1, Panel 2), even if still irrelevant to linguistic tasks like word recognition. Second, it is possible for typically "irrelevant" variability, such as talker variability, to be correlated with "relevant" variability (Figure 1, Panel 4). When talker characteristics are correlated with speech sounds, either in experiments or natural settings, voice characteristics are actually *relevant* to word learning and recognition.

When acoustic-phonetic variability is correlated with speech sounds, it is important to distinguish two learning cases. First, when cue correlations persist past initial learning (especially for strong correlations), there will be relatively fewer instances where learners must generalize to uncorrelated or anticorrelated cases (blank spaces in Figure 1, Panel 4). Thus, correlated cues should largely facilitate learning and recognition by helping separate input distributions (as in Figure 1, Panel 4). This might happen in an experiment where a particular novel word, say *darg*, is always spoken by Talker 1, while the otherwise-similar *darge* is always spoken by Talker 2, and talker-word correlations persist throughout the experiment (Creel, 2014a; see also Creel et al., 2008, Creel & Tumlin, 2011, in adults; but see Quam, Knight, & Gerken, 2017); or in a less-deterministic case where, e.g., men tend to preaspirate less than women (Foulkes, Docherty, & Watt, 1999).

Second, when cue correlations do not reliably persist past initial learning, learners must generalize past their learning data. This might happen, e.g., when children learn certain words only from one caregiver and others from another caregiver, but must eventually generalize to other speakers. In these cases, short-term facilitation from correlated cues may come at the cost of reduced generalization beyond particular contexts or talkers, reflected in reduced recognition speed or accuracy. Even worse, correlated cues that are salient but eventually irrelevant might pull attention away from relevant dimensions, impairing learning (relative to random variability; see Quam, Knight, & Gerken, 2017).

A particularly complex case of variability patterning (which is again ambiguous in terms of lawfulness) arises from accent differences. How does variability pattern between accents? For some sound contrasts, accent variability is inconsequential because two accents have very similar sound patterns (Figure 2, Panel 1; see Table 2 for experimental and real-world examples). Examples include /i/ and /ɪ/ (peak and pick) in German-accented and US-accented English (Flege, Bohn, & Jang, 1997), or many corresponding vowel contrasts in Los Angeles Chicano English and the local WAE (Santa Ana & Bayley, 2008). Other cases, however, create the possibility of confusion. One accent may merge a distinction another distinguishes (Figure 2, Panel 2). For example, Chicano English and Californian WAE merge the vowels in cot /ɑ/ and caught /ɔ/ (Santa Ana & Bayley, 2008), but many other US accents do not. Or, two accents may distinguish sound categories in different ways, either due to shifted speech-sound boundaries (Figure 2, Panel 3), e.g., lax front vowels in the Californian Vowel Shift (Eckert, 2008), or different sets of cues or cue weightings (Figure 2, Panel 4), as for coda voicing in AAE vs. WAE (Y. Holt, Jacewicz, & Fox, 2016; see also Baran & Seymour, 1976; Farrington, 2018). Unless learners can track accent context, the variability depicted in Figure 2 Panels 2-4 may introduce confusion. If learners can track accent context, they can successfully interpret sounds using accent-specific categories (for a cross-linguistic example, see Quam & Creel, 2017).

Putative Mechanisms for Impacts of Variability on Learning

Models of the earliest stages of perceptual learning for spoken language tend to focus on how infants discover the constituent sounds of their language(s). Despite evidence of continued sound-category development well past infancy (Hazan & Barrett, 2000; McMurray, Danelz, Rigler, & Seedorff, 2018; Nittrouer, Lowenstein, & Packer, 2009; Ohde & Haley, 1997; see Creel & Quam, 2015, for discussion), language-specific tuning within the first year is a focus of substantial theoretical and modeling efforts. Exposure to the native language over the first year tunes infants' early, possibly species-general (Kluender, Diehl, & Killeen, 1987; Kuhl, 1981; Kuhl and Padden, 1982, 1983) sensitivities to sound patterns (Eimas et al., 1971; Jusczyk et al., 1980; Lasky et al., 1975; Sundara et al., 2018, Trehub, 1976). For vowels, this reorganization happens by 6 months (Polka & Werker, 1994); for consonants, by 12 months (Werker & Tees, 1984), though there are exceptions (e.g., Kuhl et al., 2006; Mazuka, Hasegawa, & Tsuji, 2014; Narayan, Werker, & Beddor, 2010; Polka, Colantonio, & Sundara, 2001).

We discuss models in order of their claims about the internal structure of representations, moving from models assuming more- to less-abstract representations of input variation.

Historically, models of spoken-word processing traditionally assumed acoustic-phonetic variability was normalized to match word tokens to *abstract* representations, or models simplified their input to mostly lawful variability. Exemplar approaches, on the other hand (Goldinger, 1996, 1998; Johnson, 1997), store all variability in individual, detailed experiences, out of which emerge clusters representing words and speech sounds (and talkers and accent-specific patterns). Variability on irrelevant dimensions during learning helps relevant dimensions emerge as stable (Pierrehumbert, 2003), while also helping learners detect newly relevant information, like talker-word correlations (Goldinger, 1996, 1998). Thus, exemplar-based models address impacts of input variability on learning more directly than other models, though all the models we highlight focus on learning from input distributions.

Several theoretical frameworks attempt to capture reduced discrimination of nonnative speech-sound contrasts across infancy. Kuhl et al.'s (2008) Native Language Magnet Theory Expanded (NLM-e) argues decreased discrimination results from magnetic pull of category prototypes, which abstract away from input variability. Kuhl et al. predict this learning mechanism would apply to monolingual and bilingual learners, but bilinguals may show slower and/or different learning trajectories in response to the demands of bilingual input. One might infer that children hearing multiple accents would also show slower learning trajectories within each accent due to greater input complexity. Best's (1994) Perceptual Assimilation Model predicts relative discriminability of nonnative speech-sound contrasts based on their relationship to native sound categories. If each sound in a contrast falls into a different native category, infants are predicted to continue to discriminate them (akin to between-accent relationships in Figure 2, Panel 1). If the sounds both fall within one native category, discrimination is predicted to worsen (akin to Figure 2, Panel 2).

The PRIMIR framework (Werker & Curtin, 2005), accounts for different sensitivity to speech-sound distinctions across ages and tasks. Speech is accessible through general-perceptual, word-form, or phonemic planes, but children's biases, developmental level, and task requirements constrain access. PRIMIR was developed to address a complex pattern of findings about infants' word learning as assessed in the "Switch" habituation procedure (Werker, Cohen, Lloyd, Casasola, & Stager, 1998). In one version, infants habituate to two word-object associations, and then at test, word-object pairs are maintained or switched. Word learning is inferred from a novelty response: longer looking to switched word-object pairings. Across many studies, 14-month-olds fail to distinguish similar-sounding words like /bɪ/ vs. /dɪ/—even words differentiated by two features (/pɪn/ vs. /dɪn/; Pater, Stager, & Werker, 2004)--despite distinguishing the sounds (e.g., /b/ vs. /d/, Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002). According to PRIMIR, 14-month-olds fail because they are processing word-form information (and representing it with an exemplar-based structure), while 17-month-olds succeed (Werker et al., 2002) by processing phonemic information. This might predict irrelevant variability would facilitate learning most strongly at 14 months (Rost & McMurray, 2009), prior to phonemic processing of words.

Jusczyk's (1993) Word Recognition and Phonetic Structure Acquisition (WRAPSA) model is distinct from the approaches described above in assuming infants store specific instances of syllables (not phonemes), comparing them to traces of prior exemplars (variants), not

prototypes. As experience accumulates, the auditory-analysis system adjusts weights of different features to highlight contrastive dimensions. In WRAPSA, young infants treat non-contrastive and contrastive variability as equally relevant, but non-contrastive variability receives reduced attention with age, suggesting (in line with PRIMIR) that younger infants would be more sensitive to facilitation from irrelevant variability than older learners. Apfelbaum and McMurray's (2011) associative-learning model aims to explain facilitative effects of variability on 14-month-olds' Switch word learning (Rost & McMurray, 2009). The model is somewhat consistent with exemplar models, but incorporates dimensional weighting very similar to WRAPSA (Jusczyk, 1993). Variability facilitates learning in the model by affecting association strengths between acoustic dimensions and visual referents. Thus, it may predict variability would not facilitate learning in tasks that lack visual referents, such as typical laboratory speech-sound discrimination tasks.

Sumner, Kim, King, and McGowan's (2014; Sumner, 2015) exemplar-based framework aims to account for better memory for less-frequent but socially prestigious[2] forms of a word. (Note that, while the framework was based on adult studies, it is potentially relevant to children.) Sumner and colleagues posit that more-prestigious exemplars of a word are weighted more strongly (Sumner et al., 2014; Sumner, 2015). On this account, children learning multiple dialects may successfully learn prestigious forms from limited input, though this is not consistent with data indicating that AAE-speaking children's adjustment to WAE, as the dialect of school, is often challenging (Washington & Craig, 2000). An alternative social-weighting approach, in line with traditional exemplar approaches, is that social context (talker, register, accent) is encoded within each exemplar. This would allow perceivers to up-regulate activation of contextually appropriate forms, yielding more efficient recognition. On this account, evidence for amplification of prestige forms could actually reflect contextual expectations: college students might expect to hear "prestigious" speech in scientific experiments. For multiple-accent learners, this account predicts that hearing fewer instances of words in each accent than single-accent learners will lead to slower accrual of exemplars, resulting in slower cluster formation (see discussion of NLM-e above). Thus, children learning multiple accents might show both context sensitivity and more difficulty with prestige forms compared to children hearing primarily the prestige dialect.

Evidence for Effects of Variability

We summarize evidence from three domains in which variability impacts children's perceptual development for spoken language: learning and processing of speech-sound categories and patterns; word-form recognition and word learning; and accent. For the first two domains, we first describe how children learn linguistic structure from distributional patterns in the input, and then address impacts of variability on dimensions irrelevant to the task (e.g., sound discrimination or word learning). For accent, the distinction between relevant and irrelevant variability is less clearcut.

---

[2]We note the term "prestigious"/"prestige" implies a value judgment we do not intend; however, we are not aware of an existing, less-loaded term.

We focus mainly on child learners. We sometimes include brief descriptions of learning in adults, in some cases because a research question has been more thoroughly investigated in adults, and more generally because adults represent a developmental target or endpoint. For example, we note that a complementary literature in adult word and speech-sound learning focuses on influences of word knowledge on speech-sound retuning (termed lexical feedback; Norris, McQueen, & Cutler, 2003), with some research on bottom-up perceptual learning (e.g., Goudbeek, Swingley, & Smits, 2009; Idemaru & L. Holt, 2011). Like children, adults must balance stability of representations with perceptual flexibility (Baese-Berk, 2018; Earle & Quam, 2022). While we focus on development, we refer interested readers to Baese-Berk (2018) and Barriuso and Hayes-Harb (2018), among others, for more information.

Learning and processing of speech-sound categories and patterns.

In this section, we first review evidence from learning and processing of speech-sound categories, and then infants' learning of patterns of speech sounds. In each domain of learning, we consider impacts of irrelevant variability that is clearly non-phonological, e.g., talker, pitch, and affect, on learning.

Learning and processing of speech-sound categories.—Contrastive variability between speech sounds is essential for infants' phonological learning. Infants appear to infer one sound category if the distribution of a cue is unimodal and two if it is bimodal (Maye, Werker, & Gerken, 2002; see Vallabha, McClelland, Pons, Werker, & Amano, 2007 for a computational model). Infants can also use bimodal visual information to learn two phonemes from a unimodal phonetic distribution (Teinonen et al., 2008). Thus, variability between speech sounds can facilitate learning of sound categories by serving as input to distributional-learning mechanisms.

Variability during phonetic learning can also occur on non-contrastive dimensions. Such variability can be *lawful* (Elman & McClelland, 1986), as for variability across different within-word contexts, or *unlawful,* as for within-talker or between-talker variability across tokens. Here, we consider impacts of unlawful variability on infants' speech-sound processing, but in the next section we consider infants' learning of lawful variability (patterns of speech sounds). Even in adult learners, research on monitoring for phonemes suggests that unlawful talker variation interferes with speech-sound identification (e.g., Mullennix & Pisoni, 1990). Compared with mature listeners, infants could be even more susceptible to interference, given that some adult studies suggest variability can overwhelm learners with weaker perceptual skills (Antoniou & Wong, 2016; Perrachione, Lee, Ha, & Wong, 2011; Sadakata & McQueen, 2014).

To our knowledge, no studies have found facilitative effects of irrelevant, unlawful acoustic-phonetic variability on infants' speech-sound processing, with a few cases of inhibitory effects. Two infant studies have found null effects. Quam et al., 2020 found that talker variability (4 habituation talkers) did not impact 7.5-month-old infants' discrimination of either a native, salient onset contrast /b/ vs. /p/ or a non-native, less-salient onset contrast /n/ vs. /ŋ/. Kuhl (1983) trained 6-month-old infants to discriminate a vowel pair produced by a

single synthesized "talker." Infants generalized discrimination to simulated men's, women's, and children's voices, suggesting synthesized talker variability does not disrupt infants' discrimination. Two other early studies incorporating variability into training stimuli found mixed null and inhibitory effects. In the high-amplitude-sucking procedure, 4- to 16-week-old infants' pitch discrimination was impaired by vowel variation (which we consider to be "irrelevant" for pitch discrimination), but not vice-versa (Kuhl & Miller, 1982; see also Lebedeva & Kuhl, 2010). The asymmetry likely reflected differences in acoustic salience (vowel>pitch) rather than phonological knowledge. When testing immediately followed familiarization, Jusczyk, Pisoni, and Mullennix (1992) found 2-month-olds detected changes to syllables (/bʌg/ vs. /dʌg/) whether they heard a single talker or multiple talkers. When a 2-minute delay was incorporated before the test phase, only infants in the single-talker condition detected changes. Thus, while infants track variability between speech sounds themselves to learn speech-sound categories, unlawful, irrelevant variability appears to either not impact or inhibit infants' detection of changes in speech sounds.

Learning of lawful variability: patterns of speech sounds.—Some variation in speech sounds is non-contrastive, but lawful. One common type is allophonic variation, where a phoneme has different, but related, sound forms, lawfully conditioned by different phonetic contexts. For example, *r* in "braid" is voiced, while *r* in "prayed" is voiceless due to the preceding voiceless consonant. Infants can learn a novel phonological alternation and group allophones (surface realizations) into a single "functional category" by 12 months (White, Peperkamp, Kirk, & Morgan, 2008; see also Chong & Sundara, 2015). Infants' ability to learn this patterning depends on phonetic similarity between sounds, consistent with the tendency for allophones to be phonetically similar (White & Sundara, 2014). Whether sounds are allophonic or contrastive also impacts infants' learning of a novel phonotactic pattern (sound combination) at 11 months. Infants can only learn patterns relying on sound distinctions that are contrastive—not allophonic—in their language (Seidl et al., 2009).

Unlawful variability can sometimes facilitate infants' learning of lawful variability. Seidl, Onishi, and Cristia (2014) reported facilitative effects of talker variability, for infants at 4 and 11 months, on phonotactic-pattern learning: discrimination of nonce words that were phonotactically compatible vs. incompatible with familiarization. Infants use talker differences to track grammatical-like rules (Gonzales, Gerken, & Gómez, 2018), but apparently not to track phonological statistics (Benitez, Bulgarelli, Byers-Heinlein, Saffran, & Weiss, 2020). Richtsmeier, Gerken, & Ohala (2011) found facilitation from talker and word-type frequency on preschoolers' production of phonotactic sequences (see also Richtsmeier, Gerken, Goffman, & Hogan, 2009). In summary, infants and toddlers appear to correctly attribute variability in segments to allophony and other types of phonological alternations, differentiating allophonic variation from constrastive variation. Learning of such lawful variation in speech sounds can sometimes be facilitated by variability on *irrelevant* dimensions.

Word-form recognition and word learning.

In this section, we first review impacts of unlawful variability on early word-form recognition, the component of word learning that involves forming a representation of the sound pattern of the word, as opposed to learning the meaning or referent of the word or the association between sound and meaning. We then review impacts of unlawful variability on word learning—the mapping of word forms to *meanings*. As will become apparent here, and even more apparent in discussing accent variation, the boundaries between relevant and irrelevant variation are somewhat murky. Thus, the last topic we address in this section is the fact that learners must contend with ambiguity about whether segmental distinctions are relevant or irrelevant.

Impacts of unlawful, irrelevant variability on word-form recognition and word learning.—As with speech-sound learning, early word-form recognition and word learning require recognizing acoustic-phonetic forms by tracking variation on relevant dimensions, while ignoring irrelevant variation. A fundamental question is when children discover which variation is contrastive. This question has been probed somewhat differently for word-form recognition vs. word learning. For word-form recognition, researchers have tested how readily children generalize a word form (typically presented in isolation during familiarization) across a change in irrelevant properties to recognize it in fluent speech. By 9-10.5 months, infants recognize familiarized word forms despite unlawful variability in talker gender (Houston & Jusczyk, 2000), affect (Singh, Morgan, & White, 2004), or pitch (Singh, White, & Morgan, 2008) between familiarization and test.

Earlier, at 7.5 months, word recognition is disrupted by unlawful variability, unless familiarization words are embedded in fluent, natural speech (van Heugten and Johnson, 2012). While 7.5-month-olds generalized familiarized word forms to a new *same*-gender talker (Houston & Jusczyk, 2000), recognition was disrupted by a 1-day delay between familiarization and test (though reminding infants of the original talker in the test session helped them to generalize; Houston & Jusczyk, 2003). Relatedly, in some cases, 7.5-month-old infants only recognize words spoken in a familiar voice (their mother's; Barker & Newman, 2004). Thus, unlawful acoustic-phonetic variability challenges early recognition of word forms. While 7.5-month-old infants' word-form recognition is easily disrupted by irrelevant, unlawful variability, strategic introduction of training variability (e.g., multiple talkers or multiple vocal affects) can aid generalization across acoustic differences. While 7.5-month-olds do not generalize familiarized word forms across changes in vocal affect (Singh et al., 2004), incorporating affect variability into familiarization facilitates generalization (Singh, 2008). Thus, irrelevant, unlawful variability is challenging for infants' early word-form recognition. However, it is also critically important for the formation of robust, generalizable word-form representations.

For word learning, as for word-form recognition, unlawful, irrelevant variability during learning may introduce additional complexity, but several studies suggest it facilitates learning. Talker variability facilitates infants' associative word learning in the Switch procedure: when habituation words /buk/ and /puk/ were spoken by 18 male and female talkers, 14-month-olds successfully learned words (Rost & McMurray, 2009, 2010; see

replication in Quam, Knight, & Gerken, 2017). Galle, Apfelbaum, and McMurray (2015) demonstrated that even single-talker variability (noise variability around voicing-category means) facilitates learning if it presents a wide enough scope of acoustic values. Facilitation at 14 months, specifically, is consistent with PRIMIR (which argues 14-month-olds are processing word forms holistically and storing exemplars; Werker & Curtin, 2005); as well as with WRAPSA (which argues irrelevant dimensions are weighted more heavily earlier in infancy; Jusczyk, 1993). Facilitation in word-learning tasks but not sound-discrimination tasks (summarized earlier) is also consistent with Apfelbaum & McMurray's putative mechanism, that variability operates on cue weights between acoustic dimensions and visual referents (the latter of which are not included in common laboratory discrimination tasks).

Yet not all irrelevant variability facilitates infants' word learning. While Höhle et al. (2020) replicated the talker-variability facilitation effect, variability in the visual appearance of objects did *not* facilitate word learning (see Sommers & Barcroft, 2013, for a related finding in adults). Variability could need to be in the speech signal to help infants rule out acoustic-phonetic dimensions as potentially relevant for differentiating words. However, Quam and Swingley (in prep.) found a null effect of acoustic-phonetic variability: 18-month-olds learned words in the Switch habituation procedure irrespective of whether variability was introduced on an irrelevant dimension (pitch contour for vowel-contrasted words, or vowel for pitch-contour-contrasted words).

While in many cases learners use variability on irrelevant dimensions to downweight those dimensions, sometimes they can use non-contrastive variability as a correlated cue to help them track differences in words (or grammatical patterns). Three- to five-year-olds (Creel, 2014a; and adults, Creel, Aslin, & Tanenhaus, 2008; Creel & Tumlin, 2011) use talker differences, such as a male voice always saying "marv" and a female voice always saying "mard," to recognize otherwise-similar-sounding words more quickly in real time. These results are consistent with an exemplar approach in which learners store both speech-sound and talker information. However, in one study, 14-month-olds did not learn words when voice gender covaried with word, despite learning words with random talker variation (Quam, Knight, & Gerken, 2017). This set of findings suggests an age trajectory for word learning in which irrelevant variability, rather than being tuned out over development, may increase in usefulness over development. What is not clear is whether early presence of covarying cues may be detrimental to post-learning generalization, though as discussed in the section on structures of variability, there is less pressure to generalize if cue correlations are stable in learners' environments.

In summary, irrelevant variability impairs early word-form recognition, but inclusion of irrelevant variability during familiarization appears to aid generalization across irrelevant changes in word forms. While some studies have shown facilitation from noise-like variability for word learning, others have found null effects. The structure and source of variability may interact with age to determine whether variability will facilitate learning: random acoustic (but not visual) variability facilitates word learning at 14 but not 18 months. Correlated talker-gender cues seem to facilitate certain types of learning, perhaps only at certain ages. They do not appear to facilitate 14-month-olds' word learning, though they facilitate word recognition in preschool and beyond.

Coping with ambiguity in whether variation is contrastive.—For early word learning, differentiating contrastive vs. non-contrastive changes in words is complicated by the fact that what might appear to be contrastive changes in segments (not attributable to allophony, accent, etc.) could actually be inadvertent misspeakings. Thus, there is a tension for listeners: how does one know if a variant is inadvertent or deliberate? If inadvertent, one should disregard the variation and activate the closest-matching real word. If intentional, one should treat the change as phonologically relevant (Swingley & Aslin, 2000, 2002) and infer a new word with a distinct referent (Quam & Swingley, 2010; though see Swingley, 2016). One factor listeners consider is the degree or severity of the mispronunciation, e.g., how many phonetic features are changed. We illustrate with a series of word-learning studies by Quam and Swingley (2007, 2010, 2022), which focused primarily on toddlers, but included adults for comparison. Each 2.5-year-old or adult was repeatedly exposed to a novel object labeled "deebo" (/dibo͡ʊ/). A distracter object was visually familiarized, but not named. In test, they saw the two objects and heard either "deebo" or a mispronunciation. Visual fixations to objects assessed learners' interpretations of words.

Results (Figure 3) differed both by age and degree of mispronunciation. Toddlers were sensitive to the more substantial two-feature vowel mispronunciation "dahbo" /dabo͡ʊ/, exhibiting more looks to the target for "deebo" than for the mispronunciation (Figure 3). But only 46% of children fixated the distracter more than the target (Quam & Swingley, 2010), suggesting most children did not fully reject the target as the referent of the mispronounced word. For the milder mispronunciation, toddlers did not show a decrement in target looking, implying acceptance or even obliviousness to the mispronunciation (Quam & Swingley, 2022; see White & Morgan, 2008, for similar findings with familiar words in 19-month-olds). Adult listeners were sensitive to both mispronunciations, but even adults seemed to interpret mispronunciations differently depending on their magnitude. For the two-feature vowel mispronunciation "dahbo" /dabo͡ʊ/, 75% fixated the distracter more than the target, suggesting they interpreted "dahbo" as a novel word. By contrast, for the single-feature consonant-voicing mispronunciation "teebo" /tibo͡ʊ/, only 33% fixated the distracter more (Quam & Swingley, 2007).

One might ask how well looking measures match children's final interpretations. Does a high proportion of looks to a novel object indicate the child thinks it was the object named, or could it instead indicate uncertainty or interest in the novel object? Studies using both manual responses and looks suggest strong agreement between the two measures. Swingley (2016) presented 27-month-olds with familiar-word mispronunciations (e.g., "guck" for duck) in the presence of one familiar (duck) and one novel object. Children's manual (touchscreen) responses and looking patterns both suggested children chose the duck the majority of the time for "guck," while they mostly looked at and touched the novel object when hearing an unrelated form (e.g., "meb"). In older children, aged 3-5 years, Creel (2012) found something similar: children looked and pointed less at a picture (e.g., fish) when its name was mispronounced (e.g. "fesh", and even less for more-distant "fosh"), but the familiar object (fish) was still the predominant response.

These studies taken together suggest children do not automatically assume a mildly mispronounced word is an unfamiliar one. Thus, interpretation of potentially relevant

variation is gradient: in many cases—particularly for toddlers—variant forms are treated as close enough, with familiar(ized)-word activation decreasing as mispronunciation severity increases. Still, this gradient sensitivity must be reconciled with children's demonstrated ability to distinguish similar-sounding familiar-word pairs. For example, children in Swingley (2016) treated "puhk" (a non-word rhyming with book) like "book," but same-age children in Barton (1976) readily distinguished "bear" and "pear." Why are *b* and *p* distinguishable in familiar words but not in novel words? This pattern of results is consistent with an exemplar account. If children learn entire word forms rather than individual sounds, knowing "bear" and "pear" does not imply differentiation of "book" and "puhk." Differential sensitivity across acoustic contexts might also allow learners to use variability in one accent (e.g., where a contrast is meaningful) but not another (e.g., where it is merged).

In summary, toddlers' detection of variation in segments in newly learned words is somewhat less robust than adults'. It is also gradient, with stronger detection for multiple-feature than single-feature mispronunciations. Sensitivity to partial similarity provides a mechanism for recognizing word forms not only when speakers make errors, but also when speakers systematically produce less-familiar forms, as in cases of unfamiliar accents. Thus studies of mispronunciation sensitivity provide an interesting counterpoint to the literature on accented-speech comprehension and adaptation (discussed below).

Accent variability.

We now address accent variability, which is highly complex, for several reasons. First, it is not clear whether or not cross-accent differences should be considered lawful—they were not directly addressed by Elman and McClelland (1984, 1986), who described how segments pattern systematically *within* an accent (though they did describe how bilinguals interpret sounds with respect to the language context). Second, while differences across accents do not signal lexical contrast, as described above, there can be ambiguity between accent differences vs. contrastive changes. Third, a difference between accents typically encompasses multiple sound distinctions. In many experimental studies, young learners' detection of mispronunciations of words has been interpreted as a response to "accent" variability, but naturalistic accent differences are rarely as clear-cut as experimental mispronunciations of words. Rather than always shifting from one speech-sound category familiar to a child (e.g., /d/) to another familiar category (/t/), an accent might (partially) merge a category or change cues, as illustrated in Figure 2.

Fourth, naturalistic accent covaries with non-phonological features of dialect. In the sections below, we mainly use the term 'accent,' but we will also sometimes use the term 'dialect,' which has a somewhat distinct meaning. Accent refers specifically to pronunciation: either to differences between first- and second-language (L2) speakers' pronunciations, which may be impacted by the speech sounds and prosody of their first language (L1); or to variation in pronunciation across different dialect groups. Dialect refers more broadly, not just to pronunciation, but also to vocabulary and grammar. As one aspect of dialect, accent often activates stereotypes, that is, overgeneralized language-external notions about speakers based on their perceived identity. This means that, in some cases, it is unclear whether recognition difficulty results directly from accent properties in speech or from stereotype

activation. Listeners may experience subjective difficulty understanding speech because of negative associations with speakers' cultural or linguistic backgrounds (Kang & Rubin, 2009; Kang, Rubin, & Lindemann, 2015) and then blame speakers for their difficulty understanding (Lippi-Green, 2012). In our review, we assume that recognition difficulty results from unfamiliarity of sound forms, but alternative social-bias explanations should be kept in mind.

Finally, given the high perceived social relevance of accent (e.g., as a cue to group membership), accent variability blurs distinctions between relevant (or lawful) and irrelevant (or unlawful) variability. Thus, instead of organizing the subsections below by relevant vs. irrelevant variability, we follow a different structure. We first discuss recognition and learning of words in unfamiliar accents. Then, we discuss learning in multiple accents, addressing impacts of dialect prestige.

Recognition and learning of words in unfamiliar accents.—Infants', toddlers', and preschoolers' word learning in the face of accent variability is a research topic of increasing interest. As is typically the case in the experimental literature, there is a longer record of research on accent comprehension and adaptation in adults, which provides an endpoint measure of sorts for child accent processing. Briefly, adults are less adept at comprehending unfamiliar-accented than familiar-accented speech, but with exposure to an accent they rapidly improve (e.g., Bradlow & Bent, 2008; Clarke & Garrett, 2004; McQueen, Norris, & Cutler, 2003). Thus there is a cost to accent variation, but experience can reduce it.

In infants, a possible precursor to comprehension or adaptation to accented speech may be the ability to detect that two accents differ. Some research suggests accent-difference detection by 12 months (Butler, Floccia, Goslin, & Panneton, 2011; Cristia et al., 2014; Kinzler, Dupoux, & Spelke, 2007; Mulak, Bonn, Chládková, Aslin, & Escudero, 2017; Nazzi, Jusczyk, & Johnson, 2000). However, other studies indicate children as old as 5 years or more have difficulty differentiating their accent from another (Creel, 2018; Floccia, Butler, Girard, & Goslin, 2009; Girard, Floccia, & Goslin, 2008; Jones, Yan, Wagner, & Clopper, 2017), with slow developmental improvement. The apparent discrepancy between these research areas likely reflects methodological differences (Creel & Quam, 2015), so it is not clear how findings at earlier vs. later ages relate to each other.

There is also a lengthy timeline for recognition of words in unfamiliar accents, but it is not clear whether this skill is strongly linked to accent-difference detection. Children must develop "phonological constancy," recognizing an accented form as a familiar word. With natural exposure to an accent, infants can demonstrate phonological constancy as early as 6 months (Kitamura, Panneton, & Best, 2013). For unfamiliar accents, infants can recognize familiarized words despite changes to the talker and the accent around 13 months (Schmale & Seidl, 2009; see also Schmale, Cristià, Seidl, & Johnson, 2010). However, Best et al. (2009) reported a later time-course: 15-month-olds only recognized a word as familiar in their native accent (see also Mulak et al., 2013), while 19-month-olds did so in a different accent. Infants under 2 can only recognize familiar words in an unfamiliar accent if they are first familiarized with the accent in a highly familiar story, but by age 2 they do not need accent familiarization, suggesting good generalization to unfamiliar accents (van Heugten

& Johnson, 2014; van Heugten, Krieger, & Johnson, 2015). For novel-word learning, by 24 months, toddlers can learn a word in an unfamiliar accent after under 2 minutes of exposure to the accent (Schmale, Cristià, & Seidl, 2012). However, when taught a word in their native accent, only 30-month-olds—not 24-month-olds—recognize it in another accent (Schmale, Hollich, & Seidl, 2011).

Studies in older children suggest that difficulty understanding accented words does not end in toddlerhood, but extends at least into mid-childhood. Studies using sentence-repetition measures have found that children have greater difficulty comprehending words (Bent, 2014) and sentences (Bent & Atagi, 2015, 2017) in unfamiliar-accented speech (including unfamiliar native accents: Bent & R. Holt, 2018; see also Nathan, Wells, & Donlan, 1998). While performance improves with age and phonological skill (Bent & Atagi, 2015) and when semantic cues are provided (R. Holt & Bent, 2017), children as old as 8 still have more difficulty than adults (typically evaluated via transcription measures) in comprehending unfamiliar accents (Bent & Atagi, 2015).

Why do studies of older children show extended difficulty when studies with toddlers suggest good accent comprehension by 24-30 months? One reason may be the difficulty of the task. Toddlers are often tested with picture-fixation tasks that provide a small set of response alternatives (hearing "feesh" while seeing fish and banana pictures), while older children are tested with a repetition task, which provides less context (no pictures) and has a free response set. Creel, Rojo and Paullada (2016) found that this task difference is substantial: monolingual English-learning 3- to 5-year-olds showed good recognition of Spanish-accented words and sentences in a picture-recognition test—nearly comparable to their performance on WAE-accented words—but a much larger gap on the unfamiliar (Spanish) accent when children had to repeat heard words. In addition to emphasizing the importance of methodological factors (Creel & Quam, 2015), this collection of studies, taken together, suggests children may need more contextual support to recognize unfamiliar-accented variants at younger ages, with less support needed as development progresses.

Learning in multiple accents.—The question we address here is whether dual-accent input slows learning in each accent. Multiple studies make clear that children *can* learn to recognize and produce at least some types of cross-accent variability. Infants and toddlers can learn to interpret word forms differently depending on the accent of the speaker, both via lab training (Weatherhead & White, 2016; see also White & Aslin, 2011) and real-world exposure (van der Feest & Johnson, 2016; and see Smith et al., 2007, 2013 for children as young as 3 producing multiple phonological variants depending on the formality of a situation). However, several lab and observational studies suggest accent variation or multiple phonological variants may slow learning, relative to learning primarily in a single accent. To our knowledge, only one set of studies has simulated multi-accent input in the lab. Creel and colleagues have found that 3- to 5-year-old children (Creel, 2014b; Frye & Creel, under review) have more difficulty learning novel words when they are pronounced with segment variability (e.g., an object is called both "beesh" and "peesh") vs. segment consistency (an object is always called "beesh"). This effect is gradient, with greater difficulty (lower pointing accuracy) when two labels for an object are phonologically

dissimilar (an object is called "dedge" and "vush"; see Muench & Creel, 2013, for similar findings with adults).

Related evidence from infant studies suggests natural exposure to multiple accents may lead to slower comprehension in a given accent (Durrant, Delle Luche, Cattani, & Floccia, 2014; Floccia, Delle Luche, Durrant, Butler, & Goslin, 2012; Buckler, Oczak-Arsic, Siddiqui, & Johnson, 2017). Children exposed to multiple accents sometimes show more difficulty recognizing words in one of the accents than do children exposed to a single accent. However, this slower learning may occur because there is simply more for bi-accentual children to learn. A tradeoff often underemphasized in deficits-based accounts of language development in bilingual and bi-accentual children is that, while a monolingual or mono-accentual child may learn one accent faster, a bilingual or bi-accentual child can communicate effectively with more people and maintain stronger family and community connections (Yu, 2013).

To summarize, several laboratory studies show children can cope with accent variability, but it increases learning complexity and thus introduces challenges for learners, both in and out of the laboratory. At the same time, natural dialect exposure can be viewed as cultural and linguistic enrichment. The complexity-of-learning account predicts that learning difficulty will be proportional to the degree and nature of dissimilarity between a child's two accents. This makes it critical to accurately characterize cross-accent variability, as we illustrate in the next section.

Impacts of dialect prestige on learning in multiple accents.—Another important consideration in learning and processing of multiple accents is impacts of dialect prestige (as imposed by society). Because of the ubiquity of prestige-dialect (WAE) speakers in media and society (their currently greater US population share, overrepresentation in political office and media), children who speak non-prestige dialects become familiar more quickly with prestige dialects (Perrachione, Chiao, & Wong, 2009) and have better comprehension of them (see Baran & Seymour, 1976; Nober & Seymour, 1974, 1979) than prestige-dialect learners have for non-prestige dialects. And because a prestige or majority dialect is likely to be the dominant dialect of teachers and educational materials (Craig & Washington, 2004), children who speak a non-prestige dialect at home are much more likely to learn to dialect switch (i.e., code-switch between dialects) than speakers of mainstream dialects (Thompson et al., 2004). Dialect switching requires adapting not only to phonological features (e.g., Thomas, 2007) but also to syntactic, morphological, and lexical patterns (Washington & Craig, 1994). Given that natural dialects vary on many dimensions simultaneously, learning a dialect other than WAE outside of school and then learning to dialect switch within an academic setting (Craig & Washington, 2004) might benefit children cognitively, akin to cognitive benefits attributed to bilingualism (Bialystok, 2007).

However, other work suggests AAE speakers experience substantial difficulty in WAE-dominated learning contexts (Washington & Craig, 2000), due to added complexity of adjusting to another dialect while simultaneously processing information needed for academic success (Craig & Washington, 2004; Thompson et al., 2004). This suggests that even if there is a boost to prestigious forms (Sumner, 2015; Sumner et al., 2014),

it is insufficient for many learners. Recent research has noted that these challenges are heightened by the fact that content in standardized assessments (Moland, 2011) and curricular materials is almost always in WAE (Thompson et al., 2004). Children with more prior exposure to WAE in more integrated communities and/or stronger reading and vocabulary scores are better able to dialect switch (Craig & Washington, 2004), indicating they are mastering the complexity of bi-dialectal input to successfully acquire two dialect systems.

We now illustrate a case from AAE where mainstream characterizations of so-called "non-mainstream" dialects and neglect of relevant literature has impeded research progress. Among other features, AAE is described as having a merger of /ɪ/ and /ɛ/ before nasal consonants (as in pin and pen) and final-consonant devoicing ("had" sounds like "hat"; for an overview see, e.g., Thomas, 2007). Thus, for these sounds, the AAE/WAE relationship is characterized as shown in Figure 2, Panel 2. However, it is possible that in some cases, accents like AAE with apparent mergers actually preserve contrasts, but along different dimensions (that WAE-speaking researchers do not hear), as illustrated in Figure 2, Panel 4.

Y. Holt et al. (2016) acoustically document a pattern in AAE speakers suggesting possible maintenance of coda-voicing differences. Multiple cues have been linked to coda-voicing differences, including preceding vowel length (the vowel in "had" is longer than "hat"), consonant-closure duration, and closure voicing (see, e.g., Hayes-Harb, Smith, Bent, & Bradlow, 2008). Y. Holt and colleagues measured vowel durations in 28 word pairs ending in *d* vs. *t*, such as *had* and *hat*, in AAE and WAE speakers from the same region in western North Carolina. AAE speakers produced a *larger* vowel-duration difference in these pairs than WAE speakers did (Y. Holt et al., 2016). This suggests AAE speakers in this region maintain coda-voicing contrasts but in a different way than WAE speakers—via stronger weighting on preceding vowel duration. This is consequential in that some research programs start with the assumption that AAE-speaking children have difficulty learning to read because merged distinctions lead to decreased spelling-sound correspondence. If instead AAE speakers *do* distinguish these sounds but rely on vowel-length distinctions over other cues, these could be highlighted by teachers and mapped onto spelling differences. Future research should focus both on developing best-practices for dialect-specific literacy instruction and on other sources of cross-dialect learning difficulty, such as reduced comprehension of instructions from differently accented instructors.

A study by Baran and Seymour (1976) suggests the acoustic distinctions AAE speakers produce are more intelligible to AAE than WAE listeners. AAE-speaking children surpassed WAE-speaking children in identifying AAE speech contrasts (predominantly coda-voicing pairs). In addition, AAE-speaking children were better at understanding WAE speech than vice-versa. Further, Nober and Seymour (1974, 1979) showed a similar pattern in pre-service teachers who spoke either AAE or WAE and tried to identify AAE and WAE child speech. AAE listeners were equally good at comprehending AAE and WAE, comparable to WAE speakers on WAE speech, and better than WAE speakers on AAE speech (Nober & Seymour, 1974, 1979). Neglect of Seymour's prior research (e.g., Baran and Seymour, 1976, has been cited only 24 times) has obscured evidence for the presence of potentially contrastive patterning of coda-voicing distinctions in some AAE speakers, counter to the

common narrative that AAE obscures such distinctions. Happily, some newer work (Y. Holt et al., 2015, 2016; Y. Holt, 2018) is beginning to fill this gap. Further research is needed, along with greater acknowledgment from reviewers, editors, and funding agencies that dialect variation is an important topic of inquiry and that non-"mainstream" dialects are not special, rare cases, but widespread in learners' experiences. Better understanding of the phonological characteristics of AAE could inform more accurately targeted, phonologically based literacy instruction.

In summary, the question of how children learn in multiple accents is inextricably linked with sociological issues of dialect prestige. Dialect prestige introduces many asymmetries between the experiences of different groups of children as they cope with accent variability. Children dominant in a dialect of relatively lower prestige (as imposed by society) show some advantages over children dominant in a prestige dialect, such as greater familiarity with their non-dominant dialect and greater ability to dialect switch. At the same time, the tendency for academic settings to teach in the prestige dialect can introduce barriers for learning. Efforts to reduce so-called 'achievement gaps' for BIPOC children should carefully consider impacts of pressures to dialect (and language) switch on children's academic access. More attention and resources must be paid to important research suggesting that speakers of AAE may maintain contrasts on different acoustic dimensions than speakers of WAE.

Conclusion

In summary, our review indicates that children learn language sound patterns in the face of rampant variability in the acoustic-phonetic realizations of sounds and words. Variability can be helpful, harmful, or neutral depending on the learner's age and learning objective. Irrelevant, unlawful variability can facilitate children's learning, particularly for early learning of words and phonotactic rules, but facilitation may depend on children's age and the nature of the variability. In other domains, such as speech-sound change detection and word-form recognition, children seem either unaffected or impaired by unlawful variability, with more evidence of inhibitory effects of variability at or before 7.5 months of age. At the same time, inclusion of variability in training can aid generalization for young learners. While in some ways a source of "irrelevant" variation for sound or word learning, accent differences often interact with contrastive dimensions, so they can slow word learning temporarily and introduce academic barriers. However, exposure to multiple accents has benefits, including dialect-switching flexibility and likely better maintenance of family and community ties. More research is needed to better characterize how pressure to dialect switch contributes to so-called 'achievement gaps' for BIPOC children. Better characterizing the phonological properties of dialects like AAE could also contribute to more targeted and effective literacy instruction.

Acknowledgments

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

#ShutDownAcademia #ShutDownSTEM. (2020, June 10). www.shutdownstem.com

Antoniou M, & Wong PCM (2016). Varying irrelevant phonetic features hinders learning of the feature being trained. The Journal of the Acoustical Society of America, 139, 271–278. [PubMed: 26827023]

Apfelbaum KS, & McMurray B (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. Cognitive Science, 35, 1105–1138. doi: 10.1111/j.1551-6709.2011.01181.x [PubMed: 21609356]

Avineri N, Johnson E, Brice-Heath S, McCarty T, Ochs E, Kremer-Sadlik T, Blum S, Zentella AC, Rosa J, Flores N, Alim HS, & Paris D (2015). Invited forum: Bridging the "language gap." Journal of Linguistic Anthropology, 25, 66–86.

Baese-Berk M (2018). Perceptual learning for native and non-native speech. Psychology of Learning and Motivation, 68, 1–29.

Baran J, & Seymour HN (1976). The influence of three phonological rules of Black English on the discrimination of minimal word pairs. Journal of Speech and Hearing Research, 19, 467–474. [PubMed: 979209]

Barcroft J, and Sommers MS (2005). Effects of acoustic variability on second language vocabulary learning, Studies in Second Language Acquisition, 27, 387–414.

Barker BA, & Newman RS (2004). Listen to your mother! The role of talker familiarity in infant streaming. Cognition, 94(2), 45–53. 10.1016/j.cognition.2004.06.001

Barrett J, & Paus T (2002). Affect-induced changes in speech production. Experimental Brain Research, 146(4), 531–537. [PubMed: 12355282]

Barriuso TA, & Hayes-Harb R (2018). High Variability Phonetic Training as a Bridge From Research to Practice. The CATEsOl Journal, 30, 177.

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.

Baugh J (2017). Meaning-less differences: Exposing fallacies and flaws in "the word gap" hypothesis that conceal a dangerous "language trap" for low-income American families and their children. International Multilingual Research Journal, 11, 39–51.

Benitez VL, Bulgarelli F, Byers-Heinlein K, Saffran JR, & Weiss DJ (2020). Statistical learning of multiple speech streams: A challenge for monolingual infants. Developmental Science, 23(2), 1–12. 10.1111/desc.12896

Bent T (2014). Children's perception of foreign-accented words. Journal of Child Language, 41(6), 1334–1355. 10.1017/S0305000913000457 [PubMed: 24423250]

Bent T, & Atagi E (2015). Children's perception of nonnative-accented sentences in noise and quiet. The Journal of the Acoustical Society of America, 138(6), 3985–3993. 10.1121/L4938228 [PubMed: 26723352]

Bent T, & Atagi E (2017). Perception of nonnative-accented sentences by 5-to 8-year-olds and adults: The role of phonological processing skills. Language and Speech, 60(1), 110–122. 10.1177/0023830916645374 [PubMed: 28326989]

Bent T, & Holt RF (2018). Shhh… I Need Quiet! Children's Understanding of American, British, and Japanese-accented English Speakers. Language and Speech, 61(4), 657–673. 10.1177/0023830918754598 [PubMed: 29402164]

Best CT (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In Goodman JC and Nusbaum HC (Eds.), The Development of Speech Perception, 167–224, Cambridge, MA: MIT Press.

Best CT, Tyler MD, Gooding TN, Orlando CB, & Quann CA (2009). Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words. Psychological Science, 20(5), 539–42. 10.1111/j.1467-9280.2009.02327.x [PubMed: 19368700]

Bialystok E (2007). Cognitive effects of bilingualism: How linguistic experience leads to cognitive change. International Journal of Bilingual Education and Bilingualism, 10, 210–223.

Bradlow AR, & Bent T (2008). Perceptual adaptation to non-native speech. Cognition, 106(2), 707–729. doi: 10.1016/j.cognition.2007.04.005 [PubMed: 17532315]

Buckler H, Oczak-Arsic S, Siddiqui N, & Johnson EK (2017). Input matters: Speed of word recognition in 2-year-olds exposed to multiple accents. Journal of Experimental Child Psychology, 164, 87–100. 10.1016/j.jecp.2017.06.017 [PubMed: 28783524]

Butler J, Floccia C, Goslin J, & Panneton R (2011). Infants' discrimination of familiar and unfamiliar accents in speech. Infancy, 16(4), 392–417. 10.1111/j.1532-7078.2010.00050.x [PubMed: 32693512]

Chen S (2020, June 9). Researchers around the world prepare to #ShutDownSTEM and 'Strike For Black Lives.' Science. www.sciencemag.org/news/2020/06/researchers-around-world-prepare-shutdownstem-and-strike-black-lives#

Chong AJ, & Sundara M (2015). 18-month-olds compensate for a phonological alternation. In Grillo E, & Jepson K (Eds.). Proceedings of the 39th Annual Boston University Conference on Language Development, 113–126.

Clarke CM, & Garrett MF (2004). Rapid adaptation to foreign-accented English. Journal of the Acoustical Society of America, 116(6), 3647–3658. 10.1121/1.1815131 [PubMed: 15658715]

Cole M (2013). Differences and deficits in psychological research in historical perspective: A commentary on the special section. Developmental Psychology, 49, 84–91. [PubMed: 23316775]

Cowie R, & Cornelius R (2003). Describing the emotional states that are expressed in speech. Speech Communication, 40(1-2), 5–32.

Craig HK, & Washington JA (2004). Grade-related changes in the production of African American English. Journal of Speech, Language, and Hearing Research, 47, 450–463.

Creel SC (2012). Phonological similarity and mutual exclusivity: on-line recognition of atypical pronunciations in 3-5-year-olds. Developmental Science, 15(5), 697–713. 10.1111/j.1467-7687.2012.01173.x [PubMed: 22925517]

Creel SC (2014a). Preschoolers' flexible use of talker information during word learning. Journal of Memory and Language, 73, 81–98.

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

Creel SC (2018). Accent detection and social cognition: Evidence of protracted learning. Developmental Science, 21(2), e12524. 10.1111/desc.12524

Creel SC, & Quam C (2015). Apples and oranges: Developmental discontinuities in spoken-language processing? Trends in Cognitive Sciences, 19, 713–716. [PubMed: 26456261]

Creel SC, & Tumlin MA (2011). On-line acoustic and semantic interpretation of talker information. Journal of Memory and Language, 65(3), 264–285. 10.1016/j.jml.2011.06.005

Creel SC, Aslin RN, & Tanenhaus MK (2008). Heeding the voice of experience: The role of talker variation in lexical access. Cognition, 106, 633–664. [PubMed: 17507006]

Cristià A, Minagawa-Kawai Y, Egorova N, Gervain J, Filippin L, Cabrol D, & Dupoux E (2014). Neural correlates of infant accent discrimination: An fNIRS study. Developmental Science, 17(4), 628–635. 10.1111/desc.12160 [PubMed: 24628942]

DeGraff M (2005). Linguists' most dangerous myth: The fallacy of Creole Exceptionalism. Language in Society, 34, 533–591.

DeGraff M (2018). Linguistics' role in the right to education. Science, 360(6388), 502. [PubMed: 29724948]

Dudley-Marling C, & Lucas K (2009). Pathologizing the language and culture of poor children. Language Arts, 86, 362–370.

Durrant S, Delle Luche C, Cattani A, & Floccia C (2014). Monodialectal and multidialectal infants' representation of familiar words. Journal of Child Language, 42, 447–465. 10.1017/S0305000914000063 [PubMed: 24655564]

Earle FS, & Quam C (2022). Perceptual flexibility for speech: What are the pros and cons? Brain and Language, Article 105109. [PubMed: 35395491]

Eckert P (2008). Where do ethnolects stop? International Journal of Bilingualism, 12(1–2), 25–42. 10.1177/13670069080120010301

Eimas PD, Siqueland ER, Jusczyk P, and Vigorito J (1971). Speech perception in infants. Science, 171, 303–306. [PubMed: 5538846]

Elman JL, & McClelland JL (1986). Exploiting lawful variability in the speech wave. In Perkell JS & Klatt DH (Eds.), Invariance and variability in speech processes (pp. 360–385). Hillsdale, NJ: Erlbaum.

Farrington C (2018). Incomplete Neutralization in African American English: The case of final consonant voicing. Language Variation and Change, 30, 361–383.

Flege JE, Bohn O-S, & Jang S (1997). Effects of experience on non-native speakers' production and perception of English vowels. Journal of Phonetics, 25(4), 437–470. 10.1006/jpho.1997.0052

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. 10.1177/0165025409103871

Floccia C, Delle Luche C, Durrant S, Butler J, & Goslin J (2012). Parent or community: Where do 20-month-olds exposed to two accents acquire their representation of words? Cognition, 124(1), 95–100. 10.1016/j.cognition.2012.03.011 [PubMed: 22542698]

Foulkes P, Docherty GJ, & Watt D (1999). Tracking the emergence of sociophonetic variation. Proceedings of the XIVth International Congress of Phonetic Sciences, 1625–1628.

Frye CI, & Creel SC (under review). Preschoolers learn words with speech sound variability, but do not show a consonant bias.

Galle ME, Apfelbaum KS, & McMurray B (2015). The role of single talker acoustic variation in early word learning. Language Learning and Development, 11(1), 66–79. 10.1080/15475441.2014.895249 [PubMed: 27594811]

García M (1984). Parameters of the East Los Angeles speech community. In Ornstein-Galicia J (Ed.), Form and Function in Chicano/a English, 85–98, Rowley, MA: Newbury House.

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

Goldinger SD (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. Journal of Experimental Psychology: Learning, Memory, & Cognition, 22(5), 1166–1183. [PubMed: 8926483]

Goldinger SD (1998). Echoes of echoes? An episodic theory of lexical access. Psychological Review, 105, 251–279. [PubMed: 9577239]

Gonzales K, Gerken LA, & Gómez R (2018). How who is talking matters as much as what they say to infant language learners. Cognitive Psychology, 106, 1–20. [PubMed: 30121306]

Goudbeek M, Swingley D, & Smits R (2009). Supervised and Unsupervised Learning of Multidimensional Acoustic Categories. Journal of Experimental Psychology: Human Perception and Performance, 35(6), 1913–1933. 10.1037/a0015781 [PubMed: 19968443]

Graham Daniel W. (2019). Heraclitus. In Zalta Edward N. (Ed.), The Stanford Encyclopedia of Philosophy (Fall 2019 Edition), URL = <https://plato.stanford.edu/archives/fall2019/entries/heraclitus/>.

Guerrero A Jr. (2014). 'You Speak Good English for Being Mexican': East Los Angeles Chicano/a English: Language & Identity. Voices, 2(1), 53–62. Retrieved from https://escholarship.org/uc/item/94v4c08k

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

Heald SLM, & Nusbaum HC (2015). Variability in vowel production within and between days. PLoS ONE, 10(9): e0136791. [PubMed: 26331478]

Heinrich J, Heine SJ, & Norenzayan A (2010). The weirdest people in the world? Behavioral and Brain Sciences, 33, 1–75. [PubMed: 20377929]

Höhle B, Fritzsche T, Meß K, Phillip M, & Gafos A (2020). Only the right noise? Effects of phonetic and visual input variability on 14-month-olds' minimal pair word learning. Developmental Science, e12950. 10.1111/desc.12950 [PubMed: 32052548]

Holkup PA, Tripp-Reimer T, Salois EM, & Weinert C (2004). Community-based participatory research: An approach to intervention research with a Native American community. ANS Adv. Nurs. Sci, 27, 162–175. [PubMed: 15455579]

Holt RF, & Bent T (2017). Children's Use of Semantic Context in Perception of Foreign-Accented Speech. Journal of Speech Language and Hearing Research, 60(January), 223–231. DOI: 10.1044/2016_jslhr-h-16-0014

Holt YF (2018). Mechanisms of Vowel Variation in African American English. Journal of Speech, Language, and Hearing Research, 61, 197–209.

Holt YF, & Rangarathnam B (2018). F_0 declination and reset in read speech of African American and White American women. Speech Communication, 97, 43–50.

Holt YF, Jacewicz E, & Fox RA (2016). Temporal variation in African American English: The distinctive use of vowel duration. Journal of Phonetics & Audiology, 2(2), 1000121. 10.4172/2471-9455.1000121

Holt YF, Jacewicz E, and Fox RA (2015). Variation in vowel duration among Southern African American English speakers. American Journal of Speech-Language Pathology, 24, 460–469. [PubMed: 25951511]

Houston DM, & Jusczyk PW (2000). The role of talker-specific information in word segmentation by infants. Journal of Experimental Psychology: Human Perception and Performance, 26, 1570–1582. [PubMed: 11039485]

Houston DM, & Jusczyk PW (2003). Infants' long-term memory for the sound patterns of words and voices. Journal of Experimental Psychology: Human Perception and Performance, 29, 1143–1154. [PubMed: 14640835]

Idemaru K, & Holt LL (2011). Word recognition reflects dimension-based statistical learning. Journal of Experimental Psychology: Human Learning and Memory, 37(6), 1939–1956. 10.1037/a0025641

Israel BA, Schulz AJ, Parker EA, & Becker AB (2001). Community-based participatory research: Policy recommendations for promoting a partnership approach in health research. Education for Health, 14, 182–197. [PubMed: 14742017]

Johnson K, Ladefoged P, & Lindau M (1993). Individual differences in vowel production. The Journal of the Acoustical Society of America, 94(2), 701–714. [PubMed: 8370875]

Johnson K (1997). Speech perception without speaker normalization: An exemplar model. In Johnson K & Mullennix JW (Eds.), Talker Variability in Speech Processing (pp. 145–165). Academic Press.

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

Jusczyk PW (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. Journal of Phonetics, 21, 3–28.

Jusczyk PW, Pisoni DB, Walley AC, and Murray J (1980). Discrimination of relative onset time of two-component tones by infants. Journal of the Acoustical Society of America, 67, 262–270. [PubMed: 7354194]

Jusczyk PW, Pisoni DB, & Mullennix J (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. Cognition, 43(3), 253–291. 10.1016/0010-0277(92)90014-9 [PubMed: 1643815]

Kang O, & Rubin DL (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. Journal of Language and Social Psychology, 28(4), 441–456. 10.1177/0261927X09341950

Kang O, Rubin DL, & Lindemann S (2015). Mitigating U.S. undergraduates' attitudes toward international teaching assistants. TESOL Quarterly, 49(4), 681–706. 10.1002/tesq.192

Kinzler KD, Dupoux E, & Spelke ES (2007). The native language of social cognition. Proceedings of the National Academy of Sciences, 104(30), 12577–12580. 10.1073/pnas.0705345104

Kitamura C, Panneton R, & Best CT (2013). The development of language constancy: Attention to native versus nonnative accents. Child Development, 84(5), 1686–1700. [PubMed: 23432759]

Kluender KR, Diehl RL, & Killeen PR (1987). Japanese quail can learn phonetic categories. Science, 237(4819), 1195–1197. [PubMed: 3629235]

Kuhl PK (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. Journal of the Acoustical Society of America, 70, 340–349.

Kuhl PK, & Padden DM (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. Perception and Psychophysics, 32, 542–550. [PubMed: 7167352]

Kuhl PK, & Padden DM (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. Journal of the Acoustical Society of America, 73, 1003–1010. [PubMed: 6221040]

Kuhl PK, Conboy BT, Coffey-Corina S, Padden D, Rivera-Gaxiola M, & Nelson T (2008). Phonetic learning as a pathway to language: New data and Native Language Magnet Theory, expanded (NLM-e). Philosophical Transactions of the Royal Society B, 363, 979–1000.

Kuhl PK (1983). Perception of auditory equivalence classes for speech in early infancy. Infant Behavior and Development, 6(2-3), 263–285. 10.1016/S0163-6383(83)80036-8

Kuhl PK, & Miller JD (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. Perception & Psychophysics, 31(3), 279–292. [PubMed: 7088673]

Kuhl PK, 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. [PubMed: 16472309]

Lasky RE, Syrdal-Lasky A, & Klein RE (1975). VOT discrimination by four-to-six-and-a-half-month-old infants from Spanish environments. Journal of Experimental Psychology, 10, 215–225.

Lebedeva GC, & Kuhl PK (2010). Sing that tune: Infants' perception of melody and lyrics and the facilitation of phonetic recognition in songs. Infant Behavior and Development, (33)4, 419–430. 10.1016/j.infbeh.2010.04.006 [PubMed: 20472295]

Lippi-Green R (2012). English with an Accent: Language, Ideology, and Discrimination in the United States: 2nd Edition. Routledge: London and New York.

Lisker L, & Abramson AS (1964). A cross-language study of voicing in initial stops: acoustical measurements. Word, 20(3), 384–422.

Marcovich M (1966). On Heraclitus. Phronesis, 11(1), 19–30.

Maye J, Werker JF, and Gerken L (2002). Infant sensitivity to distributional information can affect phonetic discrimination. Cognition, 82, B101–B111. [PubMed: 11747867]

McMurray B, Danelz A, Rigler H, & Seedorff M (2018). Speech categorization develops slowly through adolescence. Developmental Psychology, 54(8), 1472–1491. [PubMed: 29952600]

Michaels S (2013). Déjà vu all over again: What's wrong with Hart & Risley and a "linguistic deficit" framework in early childhood education? LEARNing Landscapes, 7, 23–41.

Moland CW (2011). Comparison of the Diagnostic Evaluation of Language Variation-Screening Test (DELV-ST) to two other screeners for low-income, African American children. [Ph.D. Dissertation]. Louisiana State University, Baton Rouge, LA.

Mulak KE, Best CT, Tyler MD, Kitamura C, & Irwin JR (2013). Development of phonological constancy: 19-month-olds, but not 15-month-olds, identify words in a non-native regional accent. Child Development, 84, 2064–2078. 10.1111/cdev.12087 [PubMed: 23521607]

Mulak KE, Bonn CD, Chládková K, Aslin RN, Escudero P (2017). Indexical and linguistic processing by 12-month-olds: Discrimination of speaker, accent and vowel differences. PLoS ONE, 12(5), e0176762. 10.1371/journal.pone.0176762 [PubMed: 28520762]

Mullennix JW, & Pisoni DB (1990). Stimulus variability and processing dependencies in speech perception. Perception & Psychophysics, 47(4), 379–390. http://www.ncbi.nlm.nih.gov/pubmed/2345691 [PubMed: 2345691]

Narayan CR, Werker JF, & Beddor PS (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. Developmental Science, 13, 407–420. [PubMed: 20443962]

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. 10.1017/S0305000998003444 [PubMed: 9770911]

Nazzi T, Jusczyk PW, & Johnson EK (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. Journal of Memory and Language, 43(1), 1–19. 10.1006/jmla.2000.2698

Ndlovu-Gatsheni SJ (2019). Provisional notes on decolonizing research methodology and undoing its dirty history. Journal of Developing Societies, 35, 481–492.

Nittrouer S, Lowenstein JH, & Packer RR (2009). Children discover the spectral skeletons in their native language before the amplitude envelopes. Journal of Experimental Psychology: Human Perception and Performance, 35(4), 1245–1253. 10.1037/a0015020 [PubMed: 19653762]

Nober EH, & Seymour HN (1974). Speech recognition scores of white and Black student-teacher listeners for Black and white first grade speakers. In Technical report for National Inst. of Education (DHEW), Washington, D.C. (Issue 4).

Nober EH, & Seymour HN (1979). Speaker intelligibility of black and white school children for black and white adult listeners under varying listening conditions. Language and Speech, 22(3), 237–242. 10.1177/002383097902200304

Norris D, McQueen JM, & Cutler A (2003). Perceptual learning in speech. Cognitive Psychology, 47(2), 204–238. 10.1016/S0010-0285(03)00006-9 [PubMed: 12948518]

Ohde RN, & Haley KL (1997). Stop-consonant and vowel perception in 3- and 4-year-old children. Journal of the Acoustical Society of America, 102(6), 3711–3722. http://www.ncbi.nlm.nih.gov/pubmed/9407663 [PubMed: 9407663]

Patel L (2015). Decolonizing educational research. New York: Routledge.

Pater J, Stager C, & Werker J (2004). The perceptual acquisition of phonological contrasts. Language, 80(3), 384–402.

Perrachione TK, Chiao JY, & Wong PCM (2009). Asymmetric cultural effects on perceptual expertise underlie an own-race effect for voices. Cognition, 114, 42–55. [PubMed: 19782970]

Perrachione TK, Lee J, Ha LY, & Wong PC (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, 461–472. [PubMed: 21786912]

Pierrehumbert JB (2003). Phonetic diversity, statistical learning, and acquisition of phonology. Language and Speech, 46(2-3), 115–154. doi: 10.1177/00238309030460020501 [PubMed: 14748442]

Polka L, & Werker JF (1994). Developmental changes in perception of nonnative vowel contrasts. Journal of Experimental Psychology: Human Perception and Performance, 20, 421–435. [PubMed: 8189202]

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. 10.1121/1.1362689 [PubMed: 11386570]

Quam C, & Creel SC (2017). Mandarin-English bilinguals process lexical tones in newly learned words in accordance with the language context. PLOS-One, 12, e0169001. doi:10.1371/journal.pone.0169001. [PubMed: 28076400]

Quam C, & Swingley D (2007, November). Phonological knowledge trumps salient local regularity in 2-year-olds' word learning. Boston University Conference on Language Development, 32 (poster).

Quam C, & Swingley D (2010). Phonological knowledge guides 2-year-olds' and adults' interpretation of salient pitch contours in word learning. Journal of Memory and Language, 62, 135–150. [PubMed: 20161601]

Quam C, & Swingley D (2022). A protracted developmental trajectory for detection of consonant mispronunciations in newly learned words. Language Acquisition, DOI: 10.1080/10489223.2022.2069026

Quam C, & Swingley D (in preparation). English-learning children's processing of salient phonetic distinctions varying in phonological relevance for word identity.

Quam C, Clough L, Knight S, & Gerken LA (2020). Does variability impact infants' sound discrimination? Infancy. 2020;00:1–20. 10.1111/infa.12371

Quam C, Knight S, and Gerken LA (2017). The distribution of talker variability impacts infants' word learning. Laboratory Phonology: Journal of the Association for Laboratory Phonology, 8, 1–27. DOI: 10.5334/labphon.25

Richtsmeier P, Gerken LA, & Ohala D (2011). Contributions of phonetic token variability and word-type frequency to phonological representations. Journal of Child Language, 38, 951–978. [PubMed: 21126387]

Richtsmeier P, Gerken LA, Goffman L, & Hogan T (2009). Statistical frequency in perception affects children's lexical production. Cognition, 111, 372–377. [PubMed: 19338981]

Rost GC, & Mcmurray B (2009). Speaker variability augments phonological processing in early word learning. Developmental Science, 12(2), 339–349. doi: 10.1111/j.1467-7687.2008.00786.x [PubMed: 19143806]

Sadakata M, & McQueen JM (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. Frontiers in Psychology, 5, 1318. [PubMed: 25505434]

Santa Ana O, and Bayley R (2008). Chicano English: Phonology. In Schneider E (Ed.), Varieties of English, Vol. 2: The Americas and the Caribbean, 219–238. Berlin: Walter de Gruyter.

Schmale R, & Seidl A (2009). Accommodating variability in voice and foreign accent: Flexibility of early word representations. Developmental Science, 12(4), 583–601. 10.1111/j.1467-7687.2009.00809.x [PubMed: 19635085]

Schmale R, Cristià A, & Seidl A (2012). Toddlers recognize words in an unfamiliar accent after brief exposure. Developmental Science, 15(6), 732–738. 10.1111/j.1467-7687.2012.01175.x [PubMed: 23106727]

Schmale R, Cristiá A, Seidl A, & Johnson EK (2010). Developmental changes in infants' ability to cope with dialect variation in word recognition. Infancy, 15(6), 650–662. 10.1111/j.1532-7078.2010.00032.x [PubMed: 32693460]

Schmale R, Hollich G, & Seidl A (2011). Contending with foreign accent in early word learning. Journal of Child Language, 38(5), 1096–108. 10.1017/S0305000910000619 [PubMed: 21310097]

Schætz S, & Müller C (2007). A study of acoustic correlates of speaker age. In Müller C (Ed), Speaker Classification II. Lecture Notes in Computer Science, vol 4441. Springer, Berlin, Heidelberg. 10.1007/978-3-540-74122-0_1

Segbroeck MV, Travadi R, Vaz C, Kim J, Black MP, Potamianos A & Narayanan SS (2014). Classification of cognitive load from speech using an i-vector framework. In Li H, Meng HM, Ma B, Chng E & Xie L (Eds.), In the 15th Annual Conference of the International Speech Communication Association (pp. 751–755).

Seidl A, Cristià A, Bernard A, & Onishi KH (2009). Allophones and phonemes in infants' learning of sound patterns. Language Learning and Development, 5, 191–202.

Seidl A, Onishi KH, & Cristia A (2014). Talker variation aids young infants' phonotactic learning. Language Learning and Development, 10, 297–307.

Singh L (2008). Influences of high and low variability on infant word recognition. Cognition, 106, 833–870. [PubMed: 17586482]

Singh L, Morgan JL, & White KS (2004). Preference and processing: The role of speech affect in early spoken word recognition. Journal of Memory and Language, 51, 173–189.

Singh L, White KS, & Morgan JL (2008). Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. Language Learning and Development, 4, 157–178.

Smith CL (1997). The devoicing of /z/ in American English: Effects of local and prosodic context. Journal of Phonetics, 25(4), 471–500. 10.1006/jpho.1997.0053

Sommers MS, & Barcroft J (2013). Effects of Referent Token Variability on L2 Vocabulary Learning. Language Learning, 63(2), 186–210. 10.1111/lang.12007

Sperry DE, Sperry LL, & Miller PJ (2019). Reexamining the verbal environments of children from different socioeconomic backgrounds. Child Development, 90, 1303–1318. [PubMed: 29707767]

Stager CL, and Werker JF (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. Nature, 388(6640), 381–382. [PubMed: 9237755]

Sumner M (2015). The social weight of spoken words. Trends in Cognitive Sciences, 19(5), 238–239. 10.1016/j.tics.2015.03.007 [PubMed: 25921867]

Sumner M, and Samuel AG (2005). Perception and representation of regular variation: The case of final -/t/. Journal of Memory and Language, 52, 322–338. doi:10.1016/j.jml.2004.11.004

Sumner M, Kyung Kim S, King E, and McGowan KB (2014). The socially weighted encoding of spoken words: a dual-route approach to speech perception. Frontiers in Psychology, 4, 1015. [PubMed: 24550851]

Sundara M, Ngon C, Skoruppa K, Feldman NH, Onario GM, Morgan JL, & Peperkamp S (2018). Young infants' discrimination of subtle phonetic contrasts. Cognition, 178, 57–66. [PubMed: 29777983]

Swingley D (2016). Two-year-olds interpret novel phonological neighbors as familiar words. Developmental Psychology, 52(7), 1011–1023. [PubMed: 27337510]

Swingley D and Aslin RN (2000). Lexical neighborhoods and the word-form representations of 14-month-olds. Psychological Science, 13(5), 480–484.

Swingley D and Aslin RN (2002). Spoken word recognition and lexical representation in very young children. Cognition, 76, 147–166.

Teinonen T, Aslin RN, Alku P, & Csibra G (2008). Visual speech contributes to phonetic learning in 6-month-old infants. Cognition, 108(3), 850–855. 10.1016/j.cognition.2008.05.009 [PubMed: 18590910]

Thomas ER (2007). Phonological and phonetic characteristics of African American Vernacular English. Language and Linguistics Compass, 1(5), 450–475. 10.1111/j.1749-818x.2007.00029.x

Thompson CA, Craig HK, & Washington JA (2004). Variable production of African American English across oracy and literacy contexts. Language, Speech, and Hearing Services in Schools, 35, 269–282. [PubMed: 15248796]

Trehub SE (1976). The discrimination of foreign speech contrasts by infants and adults. Child Development, 47, 466–472.

Vallabha GK, McClelland JL, Pons F, Werker JF, & Amano S (2007). Unsupervised learning of vowel categories from infant-directed speech. Proceedings of the National Academy of Sciences, 104(33), 13273–13278. 10.1073/pnas.0705369104

van der Feest SVH, & Johnson EK (2016). Input-driven differences in toddlers' perception of a disappearing phonological contrast. Language Acquisition, 23, 89–111.

van Heugten M, & Johnson EK (2012). Infants exposed to fluent natural speech succeed at cross-gender word recognition. Journal of Speech, Language, and Hearing Research, 55(2), 554–560. 10.1044/1092-4388(2011/10-0347)

van Heugten M, & Johnson EK (2014). Learning to contend with accents in infancy: Benefits of brief speaker exposure. Journal of Experimental Psychology: General, 143(1), 340–350. 10.1037/a0032192 [PubMed: 23506084]

van Heugten M, Krieger DR, & Johnson EK (2015). The developmental trajectory of toddlers' comprehension of unfamiliar regional accents. Language Learning and Development, 11, 41–65. 10.1080/15475441.2013.879636

Washington JA and Craig HK (1994). Dialectal forms during discourse of poor urban African American preschoolers. Journal of Speech and Hearing Research, 37, 816–823. [PubMed: 7967568]

Washington JA, & Craig HK (2000). Reading performance and dialectal variation. In Harris JL et al. (Eds.), Literacy in African American Communities. Taylor & Francis Group.

Weatherhead D, & White KS (2016). He says potato, she says potahto: Young infants track talker-specific accents. Language Learning and Development, 12, 92–103.

Werker JF and Curtin S (2005). PRIMIR: A developmental framework of infant speech processing. Language Learning and Development, 1(2), 197–234.

Werker JF, & Tees RC (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior and Development, 7, 49–63.

Werker JF, Cohen LB, Lloyd VL, Casasola M, & Stager CL (1998). Acquisition of word-object associations by 14-month-old infants. Developmental Psychology, 34(6), 1289–1309. [PubMed: 9823513]

Werker JF, Fennell CT, Corcoran KM, & Stager CL (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. Infancy, 3, 1–30.

White J, & Sundara M (2014). Biased generalization of newly learned phonological alternations by 12-month-old infants. Cognition, 133(1), 85–90. 10.1016/j.cognition.2014.05.020 [PubMed: 24973627]

White KS, & Aslin RN (2011). Adaptation to novel accents by toddlers. Developmental Science, 14(2), 372–384. [PubMed: 21479106]

White KS, & Morgan JL (2008). Sub-segmental detail in early lexical representations. Journal of Memory and Language, 59(1), 114–132. 10.1016/j.jml.2008.03.001

White KS, Peperkamp S, Kirk C, & Morgan JL (2008). Rapid acquisition of phonological alternations by infants. Cognition, 107(1), 238–65. 10.1016/j.cognition.2007.11.012 [PubMed: 18191826]

Wolfram W (2007). Sociolinguistic folklore in the study of African American English. Language and Linguistics Compass, 1(4), 292–313.

Yap TF, Epps J, Choi EH, & Ambikairajah E (2010). Glottal features for speech-based cognitive load classification. In Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing, 5234–5237.

Yu B (2013). Issues in bilingualism and heritage language maintenance: Perspectives of minority-language mothers of children with autism spectrum disorders. American Journal of Speech-Language Pathology, 22(1), 10–24. [PubMed: 23071196]
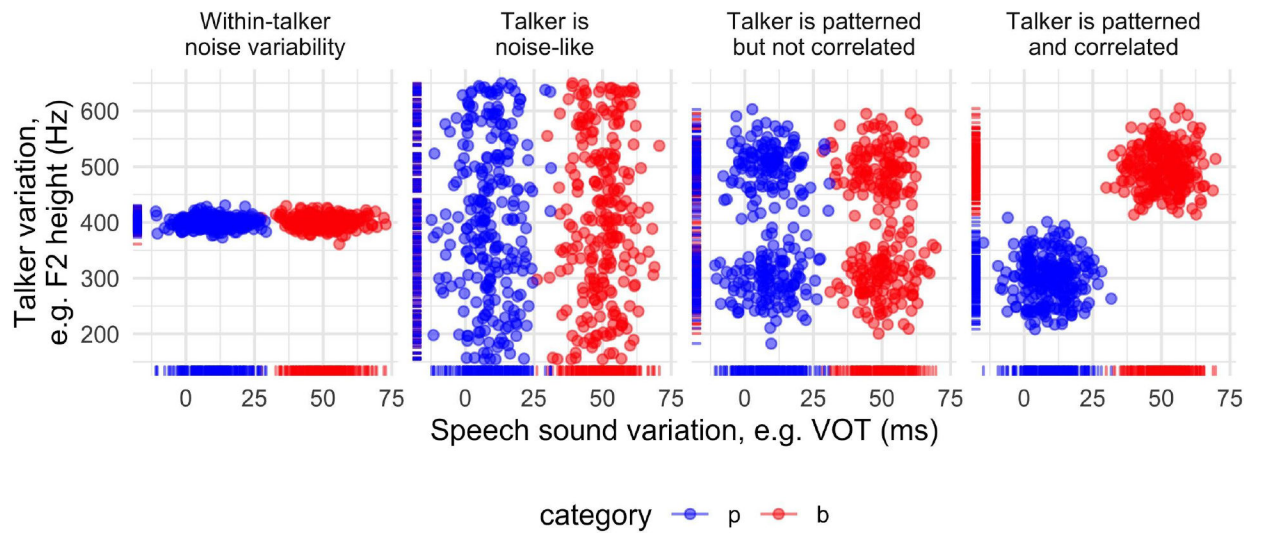
Figure 1: Possible relationships between speech and nonspeech variability in laboratory studies and natural settings.

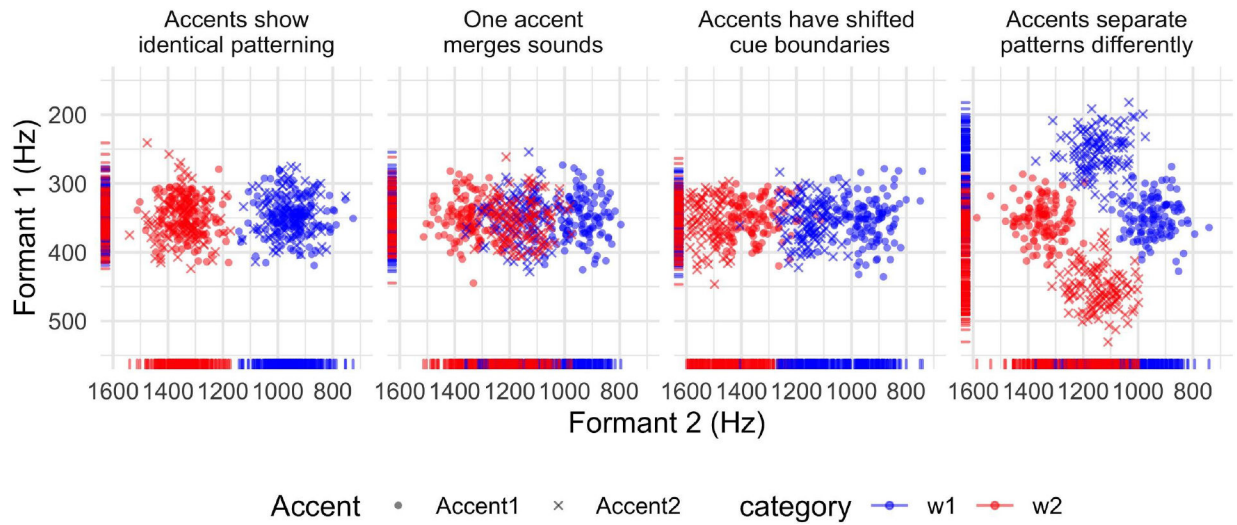Colors refer to speech-sound category. Rugs indicate frequency of each value on each axis.

Figure 2: Some of the possible forms of between-accent variability in two speech-sound categories.

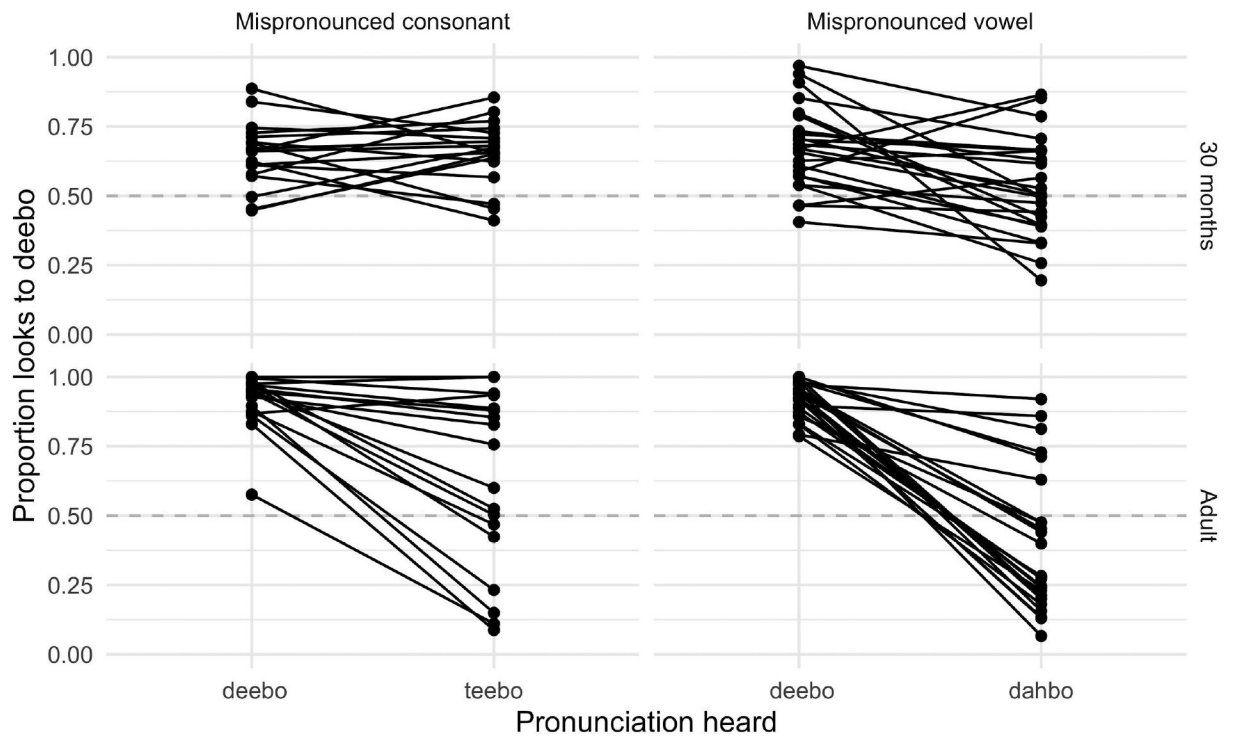Rugs indicate frequency of each value on each axis.

Figure 3: Response to mispronunciations of a newly learned word ("deebo") is related to degree of mispronunciation.

Adults and 30-month-olds showed a larger target-fixation decrement to a two-feature change in vowel height and front/backness (/i/ to /a/ in "dahbo") than to a single-feature change in consonant voicing (/d/ to /t/ in "teebo").

Table 1:

A small number of studies and real-world scenarios where each type of variability depicted in Figure 1 is observed.

Within-category noise variability on contrastive dimension	Non-contrastive variability is noise-like	Non-contrastive variability is patterned but not correlated	Non-contrastive variability is patterned + correlated
Galle, Apfelbaum, & McMurray, 2015	Rost & McMurray, 2009, 2010; Quam et al., 2017, Exp. 1	Barcroft & Sommers 2005, Exp. 2, moderate variability condition (3 speakers) *	Ø Quam et al. 2017, Exp. 2 Creel 2014a
Language input predominantly from one caregiver	Learning from a large number of talkers with similar speech cues (e.g., VOTs)	Learning from small number of caregivers with distinct voices but similar speech cues (e.g., VOTs)	Certain talkers use certain speech forms (e.g., Foulkes, Docherty, & Watt, 1999)

Note. Top row: Experiments where: variability helps; Ø no effect of variability. Bottom row: possible situations where such variability might occur.

*
This study was conducted with adults learning second-language vocabulary. Three speakers facilitated learning relative to one speaker.

Table 2:

A small number of studies and real-world scenarios, experienced by child listeners, where each type of accent variability depicted in Figure 2 is observed.

Accents show identical patterning	One accent merges sounds	Accents have shifted cue boundaries	Accents separate patterns differently
German-accented and US-accented English /i/ vs. /ɪ/ (Flege et al., 1997)	The typical depiction of AAE coda devoicing as merger (of, e.g., *had* and *hat*)	White & Aslin, 2011; Weatherhead & White, 2016	AAE vs. WAE coda voicing (Y. Holt et al., 2016)
Many similar vowel contrasts in Los Angeles Chicano English and WAE (Santa Ana & Bayley, 2008)	caught/cot merger in California WAE but not northeastern WAE; salary-celery /æ/-/e/ merger before /l/ in East LA Chicano English (García, 1984)	Lax front vowels in Californian Vowel Shift (Eckert, 2008)	AAE > WAE child listeners at AAE and cross-dialect comprehension (Baran & Seymour, 1976)

Note. Unshaded cells contain experimental or quasi-experimental tests of effects of variability on recognition or comprehension; shaded cells contain real-language cases where such variability has been described or measured. Because there is relatively little evidence of facilitation vs. inhibition based on particular accent sound-pattern relationships, we do not mark facilitation/inhibition, but discuss it in the text.