

Marieke van Heugten, Christina Bergmann
and Alejandrina Cristia

*Laboratoire de Sciences Cognitives et Psycholinguistique
(ENS / EHESS / CNRS)*

*Département d'Études Cognitives, École Normale Supérieure –
PSL Research University*

The Effects of Talker Voice and Accent on Young Children's Speech Perception

Abstract: Within the first few years of life, children acquire many of the building blocks of their native language. This not only involves knowledge about the linguistic structure of spoken language, but also knowledge about the way in which this linguistic structure surfaces in their speech input. In this chapter, we review how infants and toddlers cope with differences between speakers and accents. Within the context of milestones in early speech perception, we examine how voice and accent characteristics are integrated during language processing, looking closely at the advantages and disadvantages of speaker and accent familiarity, surface-level deviation between two utterances, variability in the input, and prior speaker exposure. We conclude that although deviation from the child's standard can complicate speech perception early in life, young listeners can overcome these additional challenges. This suggests that early spoken language processing is flexible and adaptive to the listening situation at hand.

1. Introduction

Human communication appears to be effortless: Under optimal listening conditions we hardly experience difficulty understanding other people who speak our native language. Language comprehension is, however, far from trivial. Although theories differ in their implementation of the way in which words are accessed in the mental lexicon, it is clear that spoken language is often ambiguous in nature and thus triggers the simultaneous activation of multiple – partially overlapping – word candidates, all competing for recognition. Ultimately, in the case of successful language comprehension, one of the candidate words should be recognized as the target. How does this activation and selection mechanism work?

Answering this question is not as easy as one may think. This is partially due to the fact that speech perception is greatly complicated by the absence of a one-to-one correspondence between the surface forms of words and their underlying linguistic representation. That is, factors such as speech rate, the neighboring linguistic content, but also the speaker's voice or accent, can dramatically alter the pronunciation of words across utterances. Let us, for example, consider a female American English speaker from California and a male British English speaker from London, both producing the word *grass*. As adults, we immediately grasp that although the two word tokens differ on multiple dimensions (e.g., high-pitched Californian [gɹæs] versus low-pitched London [gɹɑ:s]), they nonetheless both refer to the same underlying representation of narrow green-leafed plants commonly grown on lawns and in gardens. We also understand that both pronunciations are functionally different from phonologically closely related words such as [gɹəʊs], *gross*. In order to become proficient language users, children must acquire sufficient language expertise to make both inferences when they process speech. In other words, they must learn to strike a sophisticated balance between the use of linguistic and speaker-specific cues during word recognition. This chapter deals with how young children accomplish this impressive feat.

Children learn their native language with tremendous speed. By the time they reach their first birthday, most infants will have produced their first words. But even in the preceding months, children acquire numerous aspects of their native language. By six months of age, for example, they will have developed some understanding of frequently occurring words in the input directed to them (Bergelson and Swingley, 2012; Tincoff and Jusczyk, 1999, 2012) and they will recognize these words when spoken by a speaker they have never heard before (Bergelson and Swingley, 2013; Mandel et al., 1995; Tincoff and Jusczyk, 1999, 2012). This suggests that children's lexicons develop early in life and that even the initial word representations are sufficiently robust to deal with the variability between speakers.

This does not, however, mean that young children completely disregard speaker-specific information. In fact, much like adults, children have been shown to process speaker information to engage in non-linguistic tasks. For example, a mounting body of work shows that young children's social preferences can be greatly influenced by accent information. That is, by

five to six months of age, infants prefer to look at a speaker speaking in their own native accent over a speaker speaking in a foreign accent (Kinzler et al., 2007). This early preference for native-accented speakers develops into greater trust in native speakers compared to accented speakers during the preschool period (Kinzler et al., 2011). In fact, a speaker's accent is one of the core principles children use when evaluating others. It is even more prevalent than other, perhaps visually more salient characteristics such as a person's facial morphology (Kinzler et al., 2009). This suggests that throughout early childhood, children are sensitive to and make use of the speaker-specific cues present during oral communication.

These two lines of research, showing that infants can access both the linguistic and the non-linguistic information embedded in the speech stream, suggest that in principle, children are well-equipped to take into account both types of cues. How do these cues interact during online language comprehension? Although the two types of cues originate from the same acoustic signal, it is possible that children process them separately, and that integration only takes place off-line, once each stream of information has been attended to individually. Alternatively, children may readily incorporate speaker specificities during speech perception, just like adults update their expectations about the speaker's linguistic system online (Dahan et al., 2008; Trude and Brown-Schmidt, 2012; see Cristia et al., 2012 for an overview). Distinguishing between these two possibilities has both theoretical and practical implications. Theoretically, understanding how infants contend with speaker differences allows us to establish a more complete picture of the mechanisms underlying speech perception at a young age. On a more applied level, knowing *when* typically-developing infants experience difficulty recognizing words enables us to develop strategies to overcome such difficulties. This could be particularly useful for settings in which young children encounter many different speakers (e.g., daycare or preschool). In addition, knowledge regarding children's incorporation of speaker differences could help with developing ways to identify language difficulties in children early in life.

In recent years, developmental research examining the effects of speaker variation on speech perception early in life has started to increase. In the remainder of this chapter, we will consider the ways in which infants, toddlers, and young children cope with speaker variation during language

processing, in order to address how this indexical information affects linguistic processing. For the purposes of this chapter, we consider effects of voice variability to be due to differences in the physical characteristics of speakers. This includes changes in pitch, voice clarity, and resonance from one speaker to the other. By contrast, accent variability involves changes due to differences in the phonological system across speakers of the same language who often grew up in different regions. This involves, among others, shifts in the realization of certain speech sounds and differences in intonation patterns.

To examine how speaker variation affects early spoken language processing, it is important to understand the basic research conducted in the field of infant speech perception. In Section 2, we therefore first explain the early milestones of infants' linguistic processing. Readers who are unfamiliar with this research can read this section to gain an overview of the main benchmarks of spoken language acquisition during the early years, and read through the brief explanations of procedures used for this type of research in the Appendix, whereas those who are familiar with the topic of early language development may want to continue directly to Section 3. The subsequent four sections provide an overview of empirical results testing the effects of speaker differences and variability in young children. We address four main questions:

- Section 3: What are the advantages (if any) of familiarity with a voice or accent when processing spoken language?
- Section 4: What are the effects of deviations in voice and accent, from learning to recognition, during language processing?
- Section 5: How does variability between speakers' voices and accents affect spoken language processing?
- Section 6: How can prior exposure to accented speakers help with processing accented speech?

In Section 7, we conclude by integrating these different lines of research and discussing the theoretical implications of this work.

2. A primer on infant speech processing

In the past 40 years, research in the field of infant speech perception has provided us with a refined understanding of how children take their first steps in learning their native language. In this section, we summarize the salient developmental results (see Figure 1). This allows us to establish a rough timeline of infants' discovery of linguistic structure. The experimental research discussed here can be considered as the groundwork for testing the role of voice and accent variability in subsequent sections. Note that the Appendix contains an overview of many of the behavioral procedures used in infant speech perception research. When we describe work using one of the outlined procedures, asterisks are used to indicate that more detailed information regarding this paradigm is available in the Appendix.

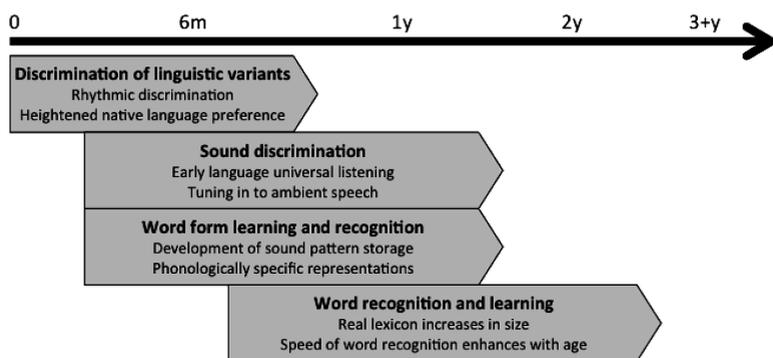


Figure 1: Infants' advances in language acquisition over the first three years of life are evident across a range of experimental tasks that have focused on certain age ranges. Presumably, development continues beyond those periods. Each of the arrows encompasses the age range typically tested in these domains, and indicates the main benchmarks infants achieve within a given area.

2.1. Discrimination and preference among languages

Languages can vary greatly in terms of their phonology. This variation not only involves differences in the use of specific sounds, but also includes differences in syllable complexity and stress. While in some languages, such as Japanese, consonants and vowels tend to alternate, other languages, such as Russian, allow for more complex syllables often containing multiple

consonants. Similarly, the temporal organization of syllables within utterances differs across languages, with some languages being described as stress-timed and others as syllable- or mora-timed. Cross-linguistic research has shown that there are acoustic correlates of such differences in phonological structure present in spoken language (Ramus et al., 1999). What does this mean for young children acquiring language? Are they able to use such surface characteristics to distinguish between languages?

Studies using **Habituation*** and **Preference* Procedures** have revealed that as early as birth, infants can discriminate between pairs of languages from different rhythmic classes (Nazzi et al., 1998), and show a preference for their native language (e. g., Mehler et al., 1988; Moon et al., 1993). This pattern of results is also observed when the speech samples are low-pass filtered, but not when they are played backwards (Mehler et al., 1988). Thus, infants' language differentiation does not appear to be based on global spectral properties, such as differences in pitch or pauses, but rather seems to be based on prosodic differences between the languages. Over time, these early language discrimination abilities are further enhanced, such that by approximately five months of age, children can distinguish their native language from another language in the same rhythmic class (Nazzi et al., 2000).

2.2. Sound discrimination

In addition to the coarse phonological differences described above, languages also differ with regard to the specific speech sounds they employ. Using a variety of paradigms (such as the **Habituation Procedure*** and the **Conditioned Head Turn Procedure***), a large body of work has examined when children tune to the sound inventory of their native language. This has typically been tested through children's abilities to discriminate specific speech sounds that either do or do not occur in the ambient language. If children's sound processing is mature, they should discriminate native-language contrasts without any problems, but should – like adults – generally discriminate non-native contrasts less well.

Studies examining sound discrimination show that infants start life with the ability to discriminate most of the linguistically relevant speech sounds employed in languages throughout the world. For instance, even though English does not have the voiceless unaspirated retroflex vs. dental

contrast (i. e. /t/ vs. /t̥/ are contrastive in other languages, such as Hindi), English-learning 6-month-olds can discriminate these two sounds (Werker and Tees, 1984). However, with more exposure to the ambient language, infants tune in to the specific phoneme contrasts relevant to their native language. This means that they not only improve their ability to discriminate native-language sounds (e. g., Kuhl et al., 2006), but also tend to lose their ability to tell apart contrasts that are not found in their native language (although there are some salient exceptions to the general decline for non-native contrasts: Best et al., 1988; Best and McRoberts, 2003). This is not to say that learners become completely insensitive to variation occurring within a native sound category. On the contrary, certain tasks reveal that toddlers can detect within-category subphonemic variation (McMurray and Aslin, 2005). This sensitivity is potentially helpful for speaker- or accent-adaptation when individual speakers differ systematically from one another at the level of subphonemic detail.

2.3. Word form learning and recognition

During the first year of life, infants not only tune in to the sound inventory of their native language, they also start learning the word forms (i. e. the sound patterns of words) that occur frequently in their input. Work using the **Word Segmentation Procedure*** reveals that infants recognize a familiarized word form embedded in fluent speech as early as six months, depending on the infant's native language, the position of the word in a sentence, and the phonological form of the target word used (Bortfeld et al., 2005; Johnson et al., 2014; see also Bosch et al., 2013, for a discussion). In the following months, this ability stabilizes (Jusczyk and Aslin, 1995; Jusczyk et al., 1999) and by eight months of age, children store long-term representations of familiarized words that are phonemically specific (Jusczyk and Hohne, 1997). This suggests that early in life, children possess the ability to encode and store (some of) the word forms they hear in the speech stream around them.

How does this ability to represent word forms help children's processing of words that occur frequently in their real-world input? Current research using the **Frequent Word Form Procedure*** suggests that as early as five months of age children prefer to listen to their own name over a matched

foil (Mandel et al., 1995), and towards the end of the first year of life they have learned many other high-frequency word forms (Hallé and De Boysson-Bardies, 1994; Swingley, 2005; Vihman et al., 2004). However, changes in the initial consonant of the frequent word form cause English-learning children to stop recognizing these items (Swingley, 2005; Vihman et al., 2004). This implies an early sensitivity to the phonemic representations of words.

2.4. Word recognition and learning

The size and content of infants' receptive lexicon is a topic of much recent work, relying mostly on measures that integrate auditory and visual information. For example, studies using the **Intermodal Preferential Looking Procedure*** have shown that infants as young as 6 months of age recognize some common nouns (Bergelson and Swingley, 2012; Tincoff and Jusczyk, 1999, 2012), although there are clear increases in both accuracy and response speed with age (Bergelson and Swingley, 2012; Fernald et al., 1998). If the word label is mispronounced, infants take longer to fixate on the correct image and show weaker preferences for that image (e.g., Mani and Plunkett, 2007, 2008; Swingley, 2009; Swingley and Aslin, 2002). These additional processing costs are proportional to the phonological distance between the mispronunciation and the target word (e.g., upon hearing *voggie* toddlers are less likely to lead to fixate on an image of a dog than upon hearing *toggie*; Mani and Plunkett, 2011; White and Morgan, 2008).

Other work has assessed toddlers' word learning. In one type of task (the **Switch Task***), 14-month-olds succeed at mapping novel labels onto novel objects when presented with pairs of words with little overlap (such as *lif* and *neem*; Stager and Werker, 1997), but not with pairs where only one segment mismatches (such as *bin* and *din*). Such minimal pairs are only learnable in this task by 17 to 20 months (Werker et al., 2002). Fourteen-month-olds' performance with minimal pairs, however, can be boosted by reducing task demands (such as using familiar words, referential cues, or presenting words in a sentential context rather than in isolation; Fennell and Waxman, 2010; Fennell and Werker, 2003; Yoshida et al., 2009).

2.5. Integrating speaker information

Before proceeding to the next section, we would like to point out that methods such as those described above not only enable us to study the acquisition of linguistic cues, but also make it possible to examine how children combine these cues with speaker-specific information during speech perception. Sound discrimination tasks, for example, can be used to assess children's reliance on surface-level aspects of the sounds by testing children's ability to generalize across sounds produced by different speakers. Similarly, word (form) recognition studies allow researchers to test children's reliance on speaker cues by measuring infants' recognition of frequent word forms spoken by an atypical or an accented speaker. And finally, in word learning tasks, experimenters can manipulate the familiarity of the speaker and the accent during the training and/or test phase to assess the role of speaker-specific information on lexical processing.

3. Effects of familiarity with the speaker's voice and/or accent

Infants' main source of language input comes from their primary caregivers. Starting approximately three months before birth, fetuses begin to perceive sensory stimulation. In the auditory domain, the mother's voice is one of the most salient contributors to prenatal sensory learning. As a result, the maternal voice has a privileged status. For example, shortly after birth, babies prefer to listen to their mother as compared to an unfamiliar female speaker (DeCasper and Fifer, 1980; Hepper et al., 1993; Mehler et al., 1978). In the following months, hearing the mother's voice leads to distinct neural activation compared to hearing an unknown speaker, as measured with Near Infrared Spectroscopy (Naoi et al., 2012), functional Magnetic Resonance Imaging (Dehaene-Lambertz et al., 2010), as well as Electroencephalography (Purhonen et al., 2004). This special role of the maternal voice has been observed across multiple languages and it remains present throughout the first year of life (see Chapter 5 in Kreiman and Sidtis, 2011, for an overview).

Since the mother's voice is so special during infancy, one may wonder whether and how it affects early speech perception. Researchers have started to investigate the possible interaction between the mother's voice and

linguistic processing in young children. For example, Barker and Newman (2004) examined whether the mother's voice can help infants segregate and encode speech under challenging listening conditions. In their word segmentation experiment, 7.5-month-olds were familiarized with two word forms that were both produced either by their own mother or by an unfamiliar female talker. These familiarization words were presented simultaneously with a distracter stimulus (a second unfamiliar female speaker reading a scientific article). While infants typically succeed in this task at this age under relatively advantageous listening conditions (Juszyk and Aslin, 1995; Newman and Juszyk, 1996), infants who heard the words produced by an unfamiliar speaker failed to recognize the word forms at test. By contrast, those who heard the words in their own mother's voice did recognize the trained words in the subsequent test phase. Thus, in cases of adverse listening conditions void of visual, lexical, and spatial context, familiar voices may be particularly beneficial for speech segregation (see Bergmann et al., 2015 for a discussion).

An advantage for maternal language processing is also observed in word recognition work. Specifically, in a recent study, 9-month-old infants' ability to map a label onto a referent was examined using electroencephalography. Children were presented with the name of a familiar object (e.g., *duck*), followed by a visual presentation of either a matching or a mismatching object (e.g., *duck* or *book*). In the case of a mismatching object, children displayed neural signatures indicating the detection of an incongruity, but this was only observed when it was their own mother who named the objects. When the experimenter (mis-)labeled the same objects, the mismatch went unnoticed (Parise and Csibra, 2012). This makes infants' interactions with their own mother potentially more fruitful than their interactions with strangers. Note, however, that parents in this study were allowed to gesture and speak in the way they typically speak with their children, so this advantage of speaker familiarity may be due to factors other than familiarity with the mother's voice alone. Also note that both studies providing evidence for the benefits of the maternal voice during language processing have presented children with relatively difficult listening conditions (either due to having a same-gender individual speak in the background or to the asynchronous presentation of object and label) and that studies in which these challenges are reduced do not always observe

such advantages (Bergelson and Swingley, 2013; Van Heugten and Johnson, 2012). It is thus plausible that the mother's voice may be particularly advantageous for situations where the processing demands are high. Indeed, under more optimal conditions, children start recognizing words produced by unfamiliar speakers of their native language from around the 6-months mark (Bergelson and Swingley, 2013; Tincoff and Jusczyk, 1999, 2012), suggesting a child's developing lexicon has the potential to be generalizable to novel speakers and novel situations. This can be very helpful when they encounter speakers they have never heard before.

So far, when discussing children's ability to understand unfamiliar speakers, we have assumed that these individuals pronounce words in approximately the same fashion as the children's parents. However, in today's linguistically diverse world, that assumption is not always a valid one. Many people live in environments where their language background does not match with that of the local community. At some point, infants will thus encounter speakers with different accents. How would children cope with such accent deviation? Do they hear the differences between accents? And if so, would they be able to understand speakers who have an unfamiliar accent?

As discussed in Section 2.1, 5-month-old infants possess the ability to differentiate between their native language and an unfamiliar language, even when that language belongs to the same rhythmic class (Nazzi et al., 2000). Will they extend this ability to the potentially more subtle differences between accents of the same language? Research using the habituation paradigm with both American and British English-learning children has revealed that although 5-month-olds are unable to discriminate between two unfamiliar accents of their native language (Butler et al., 2011), they can discriminate their own native accent from an unfamiliar accent (Butler et al., 2011; Nazzi et al., 2000). Moreover, around the same time, infants exhibit a preference for their own native accent over a completely unfamiliar accent (although their preference among their native and a more familiar accent has dissolved around this age; Kitamura et al., 2013).

Since children are sensitive to between-accent differences, one may wonder how this affects their recognition of words produced in an unfamiliar accent. Children growing up in Australia, for example, are used to hearing Australian English, whereas children growing up in Canada are more accustomed to Canadian English. It therefore stands to reason that different

accents be processed differently depending on the accent background of the listener and that early word comprehension is optimized to the local accent. But can children cope with unfamiliar accents at all? To examine this question, studies have built on the finding that children prefer to listen to lists of known words over lists of unknown words (Hallé and De Boysson-Bardies, 1994; Swingley, 2005; Vihman et al., 2004). If children recognize accented pronunciations of words, such a preference pattern should emerge regardless of whether the word lists are presented in their native or in an unfamiliar accent. It is not until the second half of their second year of life, however, that children display a preference for known over unknown words in an unfamiliar accent. That is, while American English-learning 19-month-olds display a known word preference both when the speaker is American-accented and when the speaker is Jamaican-accented, 15-month-olds fail to differentiate between the known and unknown words in a Jamaican accent (Best et al., 2009; see also Van Heugten and Johnson, 2014 for similar results with Canadian children listening to Australian-accented words). In addition, although both groups display successful word identification in their native accent, 19- but not 15-month-olds identify the referent of words produced in an unfamiliar accent (Mulak et al., 2013). The exact age at which this change occurs is, however, somewhat variable across tasks and accents (see Cristia et al., 2012; Mulak and Best, 2013 for overviews), with some work pointing towards a change around 20 months of age (Best et al., 2009; Mulak et al., 2013; Van Heugten and Johnson, 2014), and other work suggesting that the ability to recognize words across accents may not evolve until later (Flocchia et al., 2012; Van Heugten et al., 2015). It is thus likely that in the months preceding their second birthday, infants enter a transition period where their success in these tasks is dependent on both their linguistic maturity, potentially measured by their vocabulary size (Mulak et al., 2013; Van Heugten et al., 2015) and task demands.

4. Effects of deviation in speakers' voices and accents

The previous section dealt with the effects of infants' familiarity with a given voice and a given accent. We have seen that listening to the maternal voice (rather than an unknown voice) can have processing advantages for language comprehension. We have also presented evidence suggesting that

listening to a native accented speaker (rather than listening to someone with an unfamiliar accent) can be beneficial for word recognition. We now turn to the effects of what we call *deviation*, namely the presence of discrepancy in the speaker or accent between an initial learning phase and a later test phase. Please note that we wish to keep this notion strictly distinct from that of *variability*, which involves the presence of multiple speakers and accents during the initial learning phase, and to which we will turn in the next section.

Research examining children's ability to cope with differences in the speaker's voice and affect has suggested this type of speaker-related deviation may at first be challenging. That is, even though young children have no problem recognizing word forms after only limited exposure to these items when the speaker remains unchanged (e. g., Jusczyk and Aslin, 1995) or when the speaker changes to a similar-sounding speaker (Houston and Jusczyk, 2000), they do appear to initially experience greater difficulty recognizing word forms when the speaker's voice at test is clearly different from that during familiarization (Houston and Jusczyk, 2000; Singh et al., 2004). By 7.5 months of age, for example, children familiarized with word forms such as *dog* or *feet* in a female voice later recognize these words when they are spoken by another female speaker, but not when these words are subsequently spoken by a very distinct male speaker. Only a few months later, when the child is around nine months of age, such difficulties related to voice deviation have mostly disappeared (Houston and Jusczyk, 2000). Difficulties due to accent changes are somewhat more persistent. Specifically, only by 12–13 months of age will infants generalize familiarized word forms from one accent to the other (Schmale et al., 2010; Schmale and Seidl, 2009). This decline in reliance on the exact accent-induced phonetic detail thus appears to lag a few months behind the development to learn to better cope with voice (or affect) deviation.

The finding that children are able to contend with voices before they are able to contend with accents raises an important question that we have not touched on so far. In particular, one may wonder whether children's initial difficulty to cope with voices and accents is proportional to the distance between familiarization and test items. It could, for example, be possible that children are hindered more by accent than by voice deviation simply because accents may affect the relevant acoustic-phonetic cues

that children use to recognize words to a greater extent than voices do. The previously described studies suggest that while dissimilarity among voices predicts difficulty (generalization across similar voices occurs earlier than generalization across dissimilar voices), the picture is more complex when it comes to dissimilarity among accents. In word segmentation tasks, children's abilities to generalize accents emerges around the same time regardless of whether a distinct Spanish accent (Schmale and Seidl, 2009) or a much closer Canadian accent (Schmale et al., 2010) is used as the deviating accent for learners of North Midland American English. By contrast, amount of acoustic-phonetic mismatch may be more important for early word recognition. For example, while 15-month-olds have been found to reliably learn minimal pairs such as *deet* and *dit* in a word learning task (e.g., Curtin et al., 2009), success at this task only holds when the vowels of these two words are acoustically distinct in the speaker's accent (Escudero et al., 2014). When the vowels differ less on the acoustically relevant dimension, the two words are not reliably distinguished at test (Curtin et al., 2009; Escudero et al., 2014), likely because learning minimal pairs differing in just vowel quality can be challenging for young children (e.g., Nazzi, 2005; Havy and Nazzi, 2009). Thus, the generalization cost as a function of the strength of acoustic-phonetic deviation is clearly a matter for further work.

Of course, the findings that acoustic-phonetic deviation in the pronunciation of linguistic material can be challenging for infants in certain tasks does not imply that children cannot deal with any form of deviation early in life. While deviation may make linguistic processing more effortful, there is also evidence that children possess the basic capacity to deal with speaker differences early in life, both at the level of word forms (Johnson et al., 2014; Van Heugten and Johnson, 2012) and at the level of speech sounds (Kuhl, 1979, 1983). When presented with word forms in sentences, for example, rather than with isolated words in the initial familiarization phase – more similar to the way in which speech is typically heard outside the lab – children do recognize word forms in a male voice even if they had only heard them in a female voice prior to test (Van Heugten and Johnson, 2012). This suggests that while acoustic deviation can complicate word recognition, young children are, at least to some extent, equipped to deal with this challenge from early on.

5. Effects of variation in speakers' voices and accents

If acoustic *deviation* negatively affects the recognition of what is learned, does this imply that *variation* also impedes the learning process? This does not necessarily have to be the case. It seems plausible that, in contrast to what happens when learners have to generalize one speaker's pronunciation of linguistic units to a novel speaker, hearing multiple distinct speakers may help listeners construct the invariant structure (i. e. what remains the same across speakers and utterances), and would, as such, not hinder learning. If this were true, then we should observe an asymmetry where speech processing difficulties associated with multiple voices may be restricted to deviation and will not be observed for variability. Evidence for this view has been found at different levels of processing. That is, infants are able to successfully build and access linguistic representations despite (or perhaps by virtue of) variability. At the sound level, infants maintain their ability to discriminate phonemic contrasts in the face of speaker variability (Jusczyk et al., 1992; Kuhl, 1979). Similarly, word form encoding remains robust when the speaker varies during the initial learning phase (Houston, 1999; also see Singh, 2008 for similar results with affective variation). In addition to evidence for the idea that variability may not harm linguistic processing and encoding, *positive* effects of variability are observed in phonotactic learning studies: Infants presented with an artificial sound pattern grammar, in which plosive consonants are followed by lax vowels whereas fricative consonants are followed by tense vowels, are better able to learn these rules when this made-up language is uttered by multiple speakers as opposed to when it is uttered by a single speaker (Seidl, Onishi, and Cristia, 2014). Such facilitative effects are present from very early on as infants in this study benefitted from hearing multiple voices as early as four months of age. This demonstrates that voice variability can help shape the phonological patterns in the native language early on in the course of language development.

The advantage of variability can also be observed at the word level. Previous work has revealed that 14-month-old children experience difficulty learning to map two phonologically similar words (e.g., *bin* and *din*) onto different visual items, even though they successfully learned to map two phonologically unrelated words (e.g., *lif* and *neem*; Stager

and Werker, 1997) onto different items. In that study, however, infants only heard a single speaker pronounce the words. To examine whether variability could help children learn phonologically similar words, researchers increased talker variation by introducing multiple voices (Rost and McMurray, 2009). When only one speaker utters a single token of each word, infants appear to conflate /buk/ and /tuk/ tokens and display no evidence of learning the mapping between word form and referent. However, when many different speakers provide the learning material, infants successfully distinguish between the two very similar forms at test. Follow-up work has furthermore revealed that it was likely the acoustic variability in linguistically irrelevant dimensions rather than the variability in the realization of the contrastive phonemes (i. e. voice onset time) that may have driven this boost in performance (Galle et al., 2015; Rost and McMurray, 2010).

Taken together, these findings reveal that exposure to speaker variability can be helpful for learning sounds and words during infancy, at least in a laboratory setting. Whether being exposed to variation in accents in everyday life can be useful in a similar fashion has not yet been examined with young children (see Levi, 2015, however for work with school-age children). A recent study revealing greater sensitivity to phonemic detail in monolingual children who hear a single accent in their language input as compared to their age-matched monolingual peers with routine exposure to multiple accents in the home environment may, however, suggest that daily exposure to accent variability could lead to less precise representations (Durrant et al., 2015; though see Van der Feest and Johnson, in press, for evidence suggesting that children with mixed accent input may simply be more flexible, rather than less precise, in their signal-to-word mapping strategies). If such results of accent variability are also observed when phonological detail is potentially more important (i. e. in cases where two phonologically similar words are learned), this could indicate that exposure to large variability might induce greater tolerance of deviation rather than greater attention to phonetic detail. Independent of the outcome of such a test case, however, the findings to date demonstrate that non-linguistic factors can alter infants' linguistic performance. This suggests that linguistic and non-linguistic information are rapidly integrated during language processing early in life.

6. Effects of brief and long-term exposure to accents

As reviewed above, understanding speech produced by someone with an unfamiliar accent is more challenging than understanding speech produced by a speaker of the listener's own accent. This holds both for children and for adults, although adults have been shown to readily adapt to unfamiliar pronunciations of words after some experience with the accent at hand (Bradlow and Bent, 2008; Clarke and Garrett, 2004; Dahan et al., 2008; Floccia, et al., 2006; Maye et al., 2008). Would brief exposure to an accented speaker also enhance children's ability to contend with accents? Studies exploring children's abilities to cope with unfamiliar accents have recently begun to look at the effects of brief exposure to a speaker. In these studies, children are first presented with a sample spoken by an accented speaker. This allows them to build a representation of the accent that can be used to understand similarly-accented input in the future. Following this initial exposure phase, children are tested on their recognition of familiar words in the exposure accent. In a first study investigating this issue, White and Aslin (2011) tested 19-month-olds on a variant of English that involved a single segment change, where low mid-front vowels were raised (leading to *dog* being pronounced as *dag*, for instance). Such exposure changed children's perception of words, such that children who had previously heard the speaker produce *dog* like *dag*, later recognized the same speaker's *battle* as *bottle* (even though they had never heard the speaker pronounce *bottle/battle* before). By contrast, children without exposure to the change did not recognize the shifted variants, and neither group tolerated *bittle* as an instance of *bottle*. This suggests that toddlers' word recognition abilities are sufficiently flexible to deal with speaker-specific differences in pronunciation, without them being too broad to accept any deviation from the native-accented form.

Although segment shifts may play a prominent role in distinguishing certain North-American English accents, dialectal differences can be much greater. Consider, for example, North-American-, Australian-, Jamaican-, Scottish-, and Spanish-accented English. Listening to only a short excerpt in each of these accents quickly reveals that they differ on more than just a single dimension. This greater deviation potentially makes accommodation harder. To examine whether children can also accommodate more

distinct accents after gaining experience with that accent, a recent study tested Canadian English learners on their recognition of known words in an unfamiliar Australian accent. Without any exposure prior to test, children do not recognize the Australian-accented words until around their second birthday (Van Heugten and Johnson, 2014; Van Heugten et al., 2015). After exposure to the Australian English speaker reading a familiar story, however, 15-month-olds did recognize the Australian-accented test words (Van Heugten and Johnson, 2014). This suggests that brief exposure to the speaker may be beneficial for the recognition of familiar words in accents that are phonetically dissimilar from the child's own accent. Similarly, in a word learning study, North-American English-learning 2-year-olds were taught a novel word by a speaker of their own accent following brief exposure to either Spanish-accented speakers or to speakers of their own native accent. All children were subsequently tested on their recognition of the newly learned word spoken in a Spanish accent. Only the group previously exposed to Spanish-accented speech succeeded (Schmale, Cristia, and Seidl, 2012; see Schmale, et al., 2015 for benefits of exposure to variability more generally). This speaks to the continued use of accent experience throughout toddlerhood, at least when listening conditions are sufficiently challenging. This benefit of accent exposure is further exemplified by findings that routine exposure to a minority accent at home enables children to acquire phonological contrasts of that minority accent that do not surface in the regionally dominant accent. The contrast can then be flexibly used where necessary during online language processing depending on the speaker at hand (Van der Feest and Johnson, in press). Future work is necessary to examine how generalizable such adaptation is. Will exposure to a speaker in a given accent also allow children to better understand another speaker of that accent, perhaps even speakers of closely related accents?

Note that the finding that speaker exposure can help children contextualize their input does not mean that *any* form of experience with the speaker's accent *always* enables children to accommodate that accent. Neither short-term nor life-long exposure prevents young language listeners from experiencing difficulty understanding accented speakers in all situations. Routine exposure to a certain accent feature through one of the parents, for example, may not always be sufficient for the child to recognize words pronounced with such features in the lab. That is, 20-month-old children

growing up in a rhotic accent area in the UK (where /r/ is generally preserved in all positions), but who are exposed to a non-rhotic accent (where /r/ tends to be unpronounced in postvocalic position) at home through at least one parent, experienced difficulty recognizing words in which the /r/ is not produced (Floccia et al., 2012). Prior accent experience may also be less beneficial in situations where, in the absence of a mature vocabulary, the accented pronunciations of words cannot be easily mapped onto their corresponding native-accented word forms (Van Heugten and Johnson, 2014) or when children's ability to cope with accent variability on the fly has become sufficiently robust to contend with accented speech even in the absence of exposure (Van Heugten et al., 2015). Future research will have to examine the exact conditions necessary for children to make use of speaker experience and how this relates to understanding accented speakers in the real world.

7. Conclusions

The speech signal is highly complex: In addition to linguistic data, it also conveys speaker-related information, signaling factors such as the speaker's age, sex, and regional origin. To efficiently process spoken language, listeners need to take into account these indexical factors. In the developmental literature, speaker variation has most frequently been studied using voice quality, likely because effects of voice familiarity have been observed so shortly after birth (DeCasper and Fifer, 1980; Hepper et al., 1993; Mehler et al., 1978), and because this research started to emerge before much was known about how infants perceive spoken language. In recent years, research on the integration of voice information during speech perception has been complemented by research examining the consequences of hearing speech produced by speakers of unfamiliar accents. With increasing globalization, a growing number of people move to new areas where the language background differs from what they are used to. Speakers may sound accented to members of their new community, either because of differences in the ways words are pronounced across regions or because their first language affects the pronunciation of words in their second language. In addition, global media has increased the potential for exposure to accents from different regions in the world. In this chapter, we have described how

infants, toddlers, and young children cope with such variation in voice and accent during language processing.

Although effects of voice and accent could both be captured under the umbrella term “speaker-related differences”, the two types of variation may in fact be different in nature. Specifically, differences in surface form due to voice quality may be considered to be acoustic, whereas differences in surface form due to accents may be thought of as being phonetic or phonological (though, of course, to examine cross-accent differences, voice differences are typically conflated with accent changes). In addition, the amount of exposure to voice variation can differ dramatically from the amount of exposure to accent variability, at least for monolingual children growing up in households in which both parents originate from the same region. Nonetheless, many of the effects of voice and accents on speech perception are convergent. For both voices and accents, listening to what is familiar has advantages for language processing (although children learn to cope with unfamiliar voices long before they learn to cope with unfamiliar accents). When learning new word forms, deviation (i.e., hearing a word spoken in a new voice or accent) furthermore tends to increase difficulty, regardless of whether the differences are due to voices or accents. Variation (i.e., hearing the same word uttered by multiple speakers), by contrast, is useful for word learning, and prior knowledge of how a speaker pronounces sounds can be helpful when contending with unfamiliar accents.

Studies examining the effects of voice and accent on infants’ linguistic processing have important implications for theories of early speech perception. On the one hand, the surface form of words has been shown to play an essential role during early language processing. Infants and young children appear to be better able to recognize words and word forms when the acoustic-phonetic characteristics resemble those of previously heard instances (Best et al., 2009; Houston and Jusczyk, 2000; Mulak et al., 2013; Schmale et al., 2010; Schmale et al., 2011; Schmale and Seidl, 2009; Singh et al., 2004). This may be indicative of an exemplar-based storage system of words early in life, where speaker information is retained in the mental lexicon (Goldinger, 1996, 1998). On the other hand, young children overcome difficulties due to speaker-related discrepancies after only brief exposure to the speaker (Schmale et al., 2012; Van Heugten and Johnson, 2012, 2014; White and Aslin, 2011). This would imply

that successful word recognition is not only dependent on the amount of acoustic-phonetic overlap between word tokens, but also on children's opportunity to adapt to the speaker. Moreover, this enhanced ability to recognize accented words following brief accent experience generalizes to words that have not been previously heard in the unfamiliar accent, suggesting that exposure allows children to learn the phonetic-to-phonemic mappings. Despite the emphasis on episodic storage in current models of infant speech perception (such as WRAPSA and PRIMIR; Jusczyk, 1997; Werker and Curtin, 2005, respectively), abstraction processes evidently play a significant role during word recognition. Thus, even at the early stages of spoken language processing, word representations contain an abstract component. Of course, this does not rule out the possibility that early word representations also contain exemplar information. In fact, research on adult speech perception is increasingly turning to hybrid models of spoken language processing that incorporate both exemplar theory and abstraction (e.g., Goldinger, 2007; Luce and McLellan, 2005; Pierrehumbert, 2006). In the future, this combination of episodic and abstract information in the storage of word representations should be implemented in models of infant language comprehension.

Taken together, the research on early speech perception outlined in this chapter reveals that processing spoken language that deviates from the typical language input (in terms of the speaker's voice or accent) is undeniably much more complex than processing familiar voices and accents. Nonetheless, infants and toddlers are surprisingly capable to contend with voice and accent deviation. With only brief speaker exposure, for example, children can overcome the additional processing costs associated with listening to unfamiliar accents. Moreover, infants seemingly use surface-level variability in speakers' voices to access the underlying invariant structure. Differences in the way individuals speak can thus serve as a frame of reference to help infants accommodate variation. This makes children's early spoken language processing extremely sophisticated in nature.

Appendix: Infant behavioral techniques

Procedures employed for language and sound discrimination

Conditioned Head Turn Procedure. The goal of this procedure is to train infants to make a head turn each time they detect a sound change. This is implemented by presenting children with a repeating set of sounds (e.g., the sequence /ba ba ba.../), regardless of the infant's response. When a linguistically relevant change occurs (for example, the presentation of /pa/ instead of /ba/), a head turn towards a toy on the child's side is rewarded by the toy lighting up. To help children along in this task, the sound change can, at first, be accompanied by an increase in volume. Over time, this cue fades out, such that the only information signaling the change is the phonetic difference between the speech sounds.

Habituation Procedure. In this procedure, sound presentation is dependent on infants' behavioral responses (looking at the source of the sound, or sucking on a special pacifier). There are two phases to habituation studies. During the initial habituation phase, infants are presented with one or multiple stimuli drawn from a category (e.g. different tokens of the same vowel, or different sentences spoken in the same language) until their interest (measured as their looks at the source of the sound or number of sucks on the pacifier) declines. This is taken to indicate that they have encoded the key features common to the stimuli (e.g., the phonological structure of the language present in the sentences), and are ready to process new information. In a second phase, infants are presented with tokens that belong to a new category. If they increase their attention, and hence dishabituate, this indicates that they have noticed the difference between the two types of presented tokens and can distinguish them. Studies using looking time often have a within-participant design, measuring both responses to new tokens of the habituated category and responses to tokens of a new category in different test trials. They sometimes also contain visual information of the speaker pronouncing the stimuli. Studies relying on sucking responses, by contrast, tend to use between-subject comparisons, whereby one group of infants experiences no change and thus acts as control. They do not have a visual component. Typically, infants in all of these implementations show a novelty preference, reacting more strongly to tokens of a new vowel

or passages in a new linguistic variety than to tokens of the habituated category.

Preference Procedure. In this procedure, sound presentation is dependent on infants' behavioral responses (looking at the source of the sound or sucking on a special pacifier). There is typically only one phase to preference studies, although the test phase can be preceded by a familiarization phase. During the test phase, infants are presented with alternating trials that each contain tokens of one type; for example, sentences in the infant's native language versus an unfamiliar language. Sometimes, visual information is available as well. A significant difference in listening times (measured as infants' looks at the source of the sound or number of sucks on the pacifier), revealing a preference for one variant over the other, is interpreted as a sign of discrimination.

Procedures employed for word form recognition and learning

Frequent Word Form Procedure. Infants tested in this paradigm are presented with two types of trials: In familiar word trials, children hear lists of words that occur frequently in speech directed to infants (e. g., *ball*, *diaper*). By contrast, in unfamiliar word trials, lists of phonotactically legal non-words (e. g., *dimma*) or real, but rarely occurring words in infant-directed speech (e. g., *feline*) are presented. Alternatively, the list of unfamiliar words may consist of mispronunciations of the likely-known words. Frequently occurring implementations involve either a central fixation screen or a head turn preference set-up with lights positioned in front of the infant as well as on each of the infant's sides. In all cases, children's attention to these items is assessed through orientation times towards the sound source (coming from the direction of the screen or from a blinking side light).

Word Segmentation Procedure. This procedure consists of two phases both of which only present sounds when the infant attends to the source of the sound (a flashing light in case of a head turn implementation or an abstract image shown on a central screen). In the familiarization phase, infants typically hear two repeating word forms presented in isolation. Once children have accumulated a preset amount of listening time, they proceed to the subsequent test phase. In this test phase, children are presented with passages that either do or do not contain the familiarized word forms. Sometimes,

the order of words in isolation and words in passages is switched, such that children are familiarized with word forms presented in sentence context and tested on familiarized and unfamiliarized word forms in isolation.

Procedures employed for word-to-image mapping and learning

Intermodal Preferential Looking Procedure (IPLP). This procedure tests word recognition. In the typical IPLP, two images are shown side-by-side on a screen in front of the child. During the presentation of the images, one of them – the target – is named. In many studies, the words and their referents are selected to have a high probability of being well-known to the children tested in the procedure. Sometimes, words are purposely mispronounced. This allows researchers to study the phonetic specification in children’s lexical representations. Competitor images may also be well-known words, but sometimes unknown objects (e. g., a rare tool) are used. A greater proportion of looks towards the labeled picture as opposed to the unlabeled one is taken as evidence for the child knowing the word. By examining the time course of children’s looking patterns, researchers can furthermore compare the efficiency of word recognition in different conditions. In word learning tasks, the word recognition phase is preceded by a training phase where infants are taught a novel word. This teaching phase can take the form of labeling trials, where the word form is played and a single (or at least unambiguous) image is shown on the screen or it can be conducted in-person.

Switch task. In this procedure, children are first presented with two types of alternating trials. In each trial, the image of a given object displayed on a central screen is paired with a label (e. g., one novel object is paired with *lif* and the other with *neem*). Children are presented with these two word-object pairs until their interest, measured by their looks toward the screen, drops significantly (i. e. they *habituate*). In the subsequent test phase, a single object is projected on the screen, either accompanied by the same label as before (e. g., *lif* with the *lif*-object) or by the other label (e. g., *neem* with the *lif*-object). If children have successfully encoded the two words, looking times should be longer (children should be surprised) when the label and object mismatch compared to when they are matched. Versions where just a single word-object pair is used are possible as well.

References

- Barker, B., and Newman, R. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, 94(2), B45–B53.
- Bergelson, E., and Swingle, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, 109(9), 3253–3258.
- Bergelson, E., and Swingle, D. (2013). Infant word comprehension: Robust to speaker differences but sensitive to single phoneme changes. Talk presented at the Workshop on Infant Language Development, Donostia – San Sebastian, Spain.
- Bergmann, C., ten Bosch, L., Fikkert, P., and Boves, L. (2015). Modelling the noise-robustness of infants' word representations: The impact of previous experience. *PLoS ONE* 10(7): e0132245.
- Best, C. T., and McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, 46(2-3), 183–216.
- Best, C. T., McRoberts, G. W., and Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3), 345–360.
- Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., and Quann, C. A. (2009). Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words. *Psychological Science*, 20(5), 539–542.
- Bortfeld, H., Morgan, J. L., Golinkoff, R. M., and Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science*, 16(4), 298–304.
- Bosch, L., Figueras, M., Teixidó, M., and Ramon-Casas, M. (2013). Rapid gains in segmenting fluent speech when words match the rhythmic unit: evidence from infants acquiring syllable-timed languages. *Frontiers in Psychology*, 4, 106.
- Bradlow, A. R., and Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707–729.

- Butler, J., Floccia, C., Goslin, J., and Panneton, R. (2011). Infants' discrimination of familiar and unfamiliar accents in speech. *Infancy*, 16(4), 392–417.
- Clarke, C. M., and Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, 116(6), 3647–3658.
- Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., and Floccia, C. (2012). Linguistic processing of accented speech across the lifespan. *Frontiers in Cognition*, 3, 479.
- Curtin, S. A., Fennell, C., and Escudero, P. (2009). Weighting of vowel cues explains patterns of word-object associative learning. *Developmental Science*, 12(5), 725–731.
- Dahan, D., Drucker, S. J., and Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*, 108(3), 710–718.
- DeCasper, A. J., and Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, 208(4448), 1174–1176.
- Dehaene-Lambertz, G., Montavont, A., Jobert, A., Alliol, L., Dubois, J., Hertz-Pannier, L., and Dehaene, S. (2010). Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain and Language*, 114(2), 53–65.
- Durrant, S., Delle Luche, C., Cattani, A., and Floccia, C. (2015). Mono-dialectal and multidialectal infants' representation of familiar words. *Journal of Child Language*, 42(2), 447–465.
- Escudero, P., Best, C. T., Kitamura, C., and Mulak, K. E. (2014). Magnitude of phonetic distinction predicts success at early word learning in native and non-native accents. *Frontiers in Psychology*, 5, 1059.
- Fennell, C. T., and Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development*, 81(5), 1376–1383.
- Fennell, C. T., and Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*, 46(2-3), 245–264.
- Fernald, A., Pinto, J. P., Swingle, D., Weinberg, A., and McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, 9(3), 228–231.

- Floccia, C., Delle Luche, C., Durrant, S., Butler, J., and 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.
- Floccia, C., Goslin, J., Girard, F., and Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance*, 32(5), 1276–1293.
- Galle, M. E., Apfelbaum, K. S., and McMurray, B. (2015). The role of single talker acoustic variation in early word learning. *Language Learning and Development*, 11(1), 66–79.
- Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1166–1183.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279.
- Goldinger, S. D. (2007). A complementary-systems approach to abstract and episodic speech perception. In J. Trouvain and W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 49–54). Dudweiler, Germany: Pirrot.
- Hallé, P. A., and De Boysson-Bardies, B. (1994). Emergence of an early receptive lexicon: Infants' recognition of words. *Infant Behavior and Development*, 17(2), 119–129.
- Havy, M., and Nazzi, T. (2009). Better processing of consonantal over vocalic information in word learning at 16 months of age. *Infancy*, 14(4), 439–456.
- Hepper, P. G., Scott, D., and Shahidullah, S. (1993). Newborn and fetal response to maternal voice. *Journal of Reproductive and Infant Psychology*, 11(3), 147–153.
- Houston, D. M. (1999). *The role of talker variability in infant word representations* (Unpublished doctoral dissertation). The Johns Hopkins University, Baltimore, MD.
- Houston, D. M., and Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, 26(5), 1570–1582.
- Johnson, E. K., Seidl, A., and Tyler, M. D. (2014). The edge factor in early word segmentation: utterance-level prosody enables word form extraction by 6-month-olds. *PLoS One*, 9(1), e83546.

- Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT press.
- Jusczyk, P. W., and Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29(1), 1–23.
- Jusczyk, P. W., and Hohne, E. A. (1997). Infants' memory for spoken words. *Science*, 277(5334), 1984–1986.
- Jusczyk, P. W., Houston, D. M., and Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39(3), 159–207.
- Jusczyk, P. W., Pisoni, D. B., and Mullennix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition*, 43(3), 253–291.
- Kinzler, K. D., Corriveau, K. H., and Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Developmental Science*, 14(1), 106–111.
- Kinzler, K. D., Dupoux, E., and Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, 104(30), 12577–12580.
- Kinzler, K. D., Shutts, K., DeJesus, J., and Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition*, 27(4), 623–634.
- Kitamura, C., Panneton, R., and Best, C. T. (2013). The development of language constancy: Attention to native versus nonnative accents. *Child Development*, 84(5), 1686–1700.
- Kreiman, J., and Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. Malden, MA: John Wiley & Sons.
- Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *The Journal of the Acoustical Society of America*, 66(6), 1668–1679.
- Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, 6(2), 263–285.
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., and Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9(2), F13–F21.

- Levi, S. V. (2015). Talker familiarity and spoken word recognition in school-age children. *Journal of Child Language* 42(4), 843–872.
- Luce, P. and McLennan, C. (2005). Spoken word recognition: The challenge of variation. In D. Pisoni, D. and R. Remez (Eds.), *The Handbook of Speech Perception* (pp. 591–609). Malden, MA: Blackwell.
- Mandel, D. R., Jusczyk, P. W., and Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science*, 6(5), 314–317.
- Mani, N., and Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, 57(2), 252–272.
- Mani, N., and Plunkett, K. (2008). Fourteen-month-olds pay attention to vowels in novel words. *Developmental Science*, 11(1), 53–59.
- Mani, N., and Plunkett, K. (2011). Does size matter? Subsegmental cues to vowel mispronunciation detection. *Journal of Child Language*, 38(3), 606–627.
- Maye, J., Aslin, R., and Tanenhaus, M. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, 32(3), 543–562.
- McMurray, B., and Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition*, 95(2), B15–26.
- Mehler, J., Bertoncini, J., Barrière, M., and Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. *Perception*, 7(5), 491–497.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., and Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29(2), 143–178.
- Moon, C., Cooper, R. P., and Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, 16(4), 495–500.
- Mulak, K. E., and Best, C. T. (2013). Development of word recognition across speakers and accents. In L. J. Gogate and G. Hollich (Eds.), *Theoretical and computational models of word learning: Trends in psychology and artificial intelligence* (pp. 242–269). Hershey: IGI Global: Robotics Division.
- Mulak, K. E., Best, C. T., Tyler, M. D., Kitamura, C., and Irwin, J. R. (2013). Development of phonological constancy: 19-month-olds, but

- not 15-month-olds, identify words in a non-native regional accent. *Child Development*, 84(6), 2064–2078.
- Naoi, N., Minagawa-Kawai, Y., Kobayashi, A., Takeuchi, K., Nakamura, K., Yamamoto, J., and Kojima, S. (2012). Cerebral responses to infant-directed speech and the effect of talker familiarity. *NeuroImage*, 59(2), 1735–1744.
- Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition*, 98(1), 13–30.
- Nazzi, T., Bertocini, J., and Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 756
- Nazzi, T., Jusczyk, P. W., and Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, 43(1), 1–19.
- Newman, R. S. and Jusczyk, P. W. (1996). The cocktail party effect in infants. *Perception and Psychophysics*, 58(8), 1145–1156.
- Parise, E., and Csibra, G. (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychological Science*, 728–733.
- Pierrehumbert, J. (2006). The next toolkit. *Journal of Phonetics*, 34(6), 516–530.
- Purhonen, M., Kilpeläinen-Lees, R., Valkonen-Korhonen, M., Karhu, J., and Lehtonen, J. (2004). Cerebral processing of mother's voice compared to unfamiliar voice in 4-month-old infants. *International Journal of Psychophysiology*, 52(3), 257–266.
- Ramus, F., Nespore, M., and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265–292.
- Rost, G. C., and McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, 12(2), 339–349.
- Rost, G. C., and McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, 15(6), 608–635.

- Schmale, R., Cristia, A., and Seidl, A. (2012). Toddlers recognize words in an unfamiliar accent after brief exposure. *Developmental Science*, 15(6), 732–738.
- Schmale, R., Cristia, A., Seidl, A., and Johnson, E. K. (2010). Developmental changes in infants' ability to cope with dialect variation in word recognition. *Infancy*, 15(6), 650–662.
- Schmale, R., Hollich, G., and Seidl, A. (2011). Contending with foreign accent in early word learning. *Journal of Child Language*, 38(5), 1096–1108.
- Schmale, R., and Seidl, A. (2009). Accommodating variability in voice and foreign accent: flexibility of early word representations. *Developmental Science*, 12(4), 583–601.
- Schmale, R., Seidl, A. and Cristia, A., (2015). Mechanisms underlying accent accommodation in early word learning: Evidence for general expansion. *Developmental Science* 18(4), 664–670.
- Seidl, A., Onishi, K. H., and Cristia, A. (2014). Talker variation aids young infants' phonotactic learning. *Language Learning and Development*, 10(4), 297–307.
- Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, 106(2), 833–870.
- Singh, L., Morgan, J. L., and White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*, 51(2), 173–189.
- Stager, C. L., and Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388(6640), 381–382.
- Swingle, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, 8(5), 432–443.
- Swingle, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language*, 60(2), 252–269.
- Swingle, D., and Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13(5), 480–484.
- Tincoff, R., and Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, 10(2), 172–175.

- Tincoff, R., and Jusczyk, P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy*, 17(4), 432–444.
- Trude, A. M., and Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes*, 27(7-8), 979–1001.
- Van der Feest, S. V. H., and Johnson, E. K. (in press). Input driven differences in toddler's perception of a disappearing phonological contrast. *Language Acquisition*.
- Van Heugten, M., and Johnson, E. K. (2012). Infants exposed to fluent natural speech succeed at cross-gender word recognition. *Journal of Speech, Language and Hearing Research*, 55(2), 554–560.
- Van Heugten, M., and Johnson, E. K. (2014). Learning to contend with accents in infancy: Benefits of brief speaker exposure. *Journal of Experimental Psychology: General*, 143(1), 340–350.
- Van Heugten, M., Krieger, D. R., and Johnson, E. K. (2015). The developmental trajectory of toddlers' comprehension of unfamiliar regional accents. *Language Learning and Development*, 11(1), 41–65.
- Vihman, M. M., Nakai, S., DePaolis, R. A., and Hallé, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, 50(3), 336–353.
- Werker, J. F., and Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, 1(2), 197–234.
- Werker, J. F., Fennell, C. T., Corcoran, K. M., and Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, 3(1), 1–30.
- Werker, J. F., and Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1), 49–63.
- White, K. S., and Aslin, R. N. (2011). Adaptation to novel accents by toddlers. *Developmental Science*, 14(2), 372–384.
- White, K. S., and Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, 59(1), 114–132.
- Yoshida, K. A., Fennell, C. T., Swingle, D., and Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science*, 12(3), 412–418.