Developmental Review 66 (2022) 101047

Contents lists available at ScienceDirect

Developmental Review

journal homepage: www.elsevier.com/locate/dr

Understanding why infant-directed speech supports learning: A dynamic attention perspective

Mira L. Nencheva [*], Casey Lew-Williams

Princeton University, Department of Psychology, Princeton, NJ 08540, USA

ARTICLE INFO

Keywords:
Infant-directed speech
Attention
Learning
Entrainment

ABSTRACT

Infant-directed speech (IDS) refers to the suite of prosodic and structural modifications that adults use when communicating with infants, as opposed to adults. A number of theories have proposed that IDS is uniquely able to modulate infants' attention and arousal in a way that supports real-time communication and learning. However, prior research has mainly focused on infants' overall listening preference for IDS over adult-directed-speech (ADS) without providing a mechanistic account of how IDS optimizes moment-to-moment attention and learning. Here we draw on findings from adult neuroscience showing that sustained attention to a continuous stimulus like speech is supported by a process called entrainment, where neural oscillations become time-locked to key moments in an attended stimulus. Even though entrainment appears to be automatic in development, it may be more likely to occur when stimuli are tailored to infants' developing cognitive abilities. We first bring together evidence from psychology and neuroscience showing that IDS supports speech processing by optimizing neural entrainment and thus enhancing time-locked attention. Then, we discuss how moment-to-moment attentional modulations in IDS are likely to accumulate across time in a way that impacts long-term language development. This framework serves to redefine 'high-quality speech' not as a feature of speech itself, but as a dynamic interplay between behavior, attention, and the brain. With this redefinition, developmental scientists can gain traction in understanding the beginnings and high-stakes nature of young children's highly divergent learning trajectories.

Introduction

Real-time communication requires moment-to-moment processing of the unfolding dynamics of complex, multidimensional stimuli. For infants – the newest and smallest learners – breaking into the structure of speech takes time, experience, and the development of neural machinery. How do caregivers adapt their speech to help infants sustain attention? How does infants' attention to this crucial communicative signal enable them to learn? Here, we propose a new dynamic-attention framework for understanding how infant-directed speech (IDS) supports the beginnings of human learning and communication, and how individual differences in early learning trajectories may get their start. To do so, we unite research from three fields: (1) research from developmental science on infant-directed speech, (2) research from adult cognitive psychology and neuroscience on rhythmic entrainment to speech across time, and (3) research from developmental cognitive neuroscience on infants' neural processing of speech.

Multiple theories have proposed that IDS, a mode of communication that is prosodically and structurally different from adult-

---

[*] Corresponding author.
E-mail address: nencheva@princeton.edu (M.L. Nencheva).

https://doi.org/10.1016/j.dr.2022.101047
Received 27 January 2022; Received in revised form 24 June 2022;
Available online 25 August 2022

directed speech (ADS; Fernald, 1992; Gleitman et al., 1984; Grieser & Kuhl, 1988), evolved as a response to the demands of real-time communication (McMurray, 2016) and functions to modulate infants' attention and arousal at key moments (Kuhl, 2007). Discussions in developmental science about the function of IDS in capturing and sustaining infant attention based on preferential listening have provided important theoretical intuitions (Cooper & Aslin, 1990; Fernald, 1985; Kaplan, Goldstein, Huckeby, Owren, & Cooper, 1995; ManyBabies Consortium, 2020; Pegg, Werker, & McLeod, 1992; Werker & McLeod, 1989). However, they are often disconnected from recent advances in understanding (1) moment-to-moment temporal attention to speech, and (2) the neural substrates of attention and learning in development. Uniting these fields allows us to elucidate a mechanism for explaining infants' engagement with IDS and to form testable hypotheses about how caregiver-child interactions best support learning in the moment – all with greater temporal precision than in prior work. Ultimately, this approach will facilitate an understanding of how caregivers contribute to the development of neural structures and the dynamics of speech processing in infancy and early childhood.

While most research with infants has relied primarily on preferential looking techniques and summary measures of looking time to quantify attention to a dynamic signal (Bergelson et al., 2017; Cooper & Aslin, 1990; Fernald, 1985; Kaplan et al., 1995; Pegg et al., 1992; Werker & McLeod, 1989), research with adults has tested hypotheses about the internal and external factors that define precise moments of heightened attention to continuous speech. Although perception feels continuous when we sustain attention on something, increasing evidence from adult and non-human primate research suggests that in reality our brains sample information rhythmically (Fiebelkorn & Kastner, 2019; Helfrich et al., 2018; Jones et al., 1981; Large & Jones, 1999). This occurs in synchrony with neural oscillations – complex, large-scale, rhythmic patterns of neuronal firing – in several clusters of frequencies (known as frequency bands) emerging from neuronal physiology, connectivity, and responses to external input. According to dynamic attention theory (Jones, 1976; Jones & Boltz, 1989), during selective attention to a continuous stimulus like speech, neural oscillations become time-locked to the attended stimulus (see Obleser & Kayser, 2019 for a review). This process is referred to as entrainment or cortical tracking. This type of alignment ensures that the brain receives the most important moments for processing an incoming stimulus at a time that is optimal for perception (see Fig. 1). Consequently, stronger alignment predicts improvements in stimulus detection and processing (for a review see Palva & Palva, 2018). Therefore, in order to optimally process stimuli from the environment, the human brain has to detect and align to the temporal dynamics of the stimulus.

Evidence is growing, but far from complete, for the idea that infants entrain to rhythmic auditory stimuli – an ability that may scaffold processing complex dynamic auditory signals such as speech. The ability of the brain to align its activity to the dynamics of temporally predictable stimuli appears to develop early, with studies showing neural entrainment to auditory rhythmic stimuli in 7- to 15-month-old infants (Cirelli et al., 2016), as well as newborns (Háden et al., 2015). Newborns detect the rhythm of auditory stimuli (Telkemeyer et al., 2011, 2019), and show a violation-of-expectation neural response to beat omission (Winkler et al., 2009). There is some preliminary evidence that the ability to track rhythm may start to emerge in utero. Kisilevsky et al. (2004) showed that near-term fetuses showed a heart-rate response to a change in music tempo. These findings suggest that infants, from birth or even earlier, may automatically entrain to rhythmic stimuli in their environment, but critically, how do they entrain to the complexities of the units and timescales of speech?

Speech consists of many moments that are key for perception, such as the onsets of syllables and words, the ups and downs of pitch, and amplitude over the course of phrases. If these moments do not arrive at the right time for our brain to receive them, our perception can be derailed. For example, adults are better at detecting attention probes if they coincide with the onset of words (Astheimer &

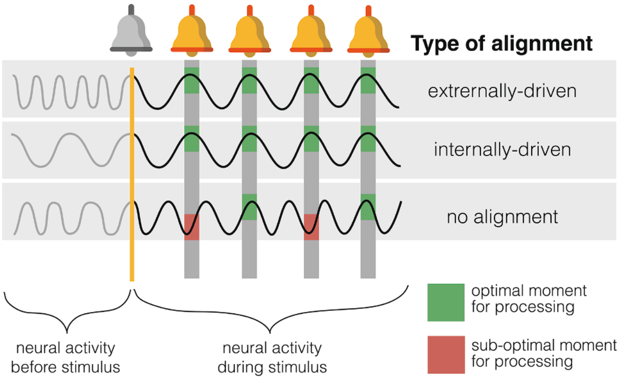

Fig. 1. In this illustration, we have depicted mock neural activity before a stimulus is presented (gray line) and during the presentation of a sequence of events, represented by bells (black line). The yellow line indicates the stimulus onset. Neural activity and stimulus dynamics become aligned when important moments in a stimulus (such as the rhythmic timing of bell sounds) coincide with a configuration of neural activity that optimizes processing. The top panel shows externally-driven alignment, where the frequency of neural activity adjusts to match that of the presented stimulus. The middle panel shows internally-driven alignment, where optimal moments for processing already align with the rhythm of the stimulus prior to its presentation. Both of these types of alignment result in perceptual information that reaches the brain at an optimal moment (shown in green). We use the terms 'entrainment' and 'cortical tracking' to refer to both of these mechanisms, given that prior research has not disentangled them. The bottom panel depicts a situation when there is no alignment. As a result, perceptual information only occasionally is received by the brain at an optimal moment (green) and frequently is received at a time that may not support encoding (red), leading to overall worse perception. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Sanders, 2009), and there is evidence that adults can track multiple nested structures simultaneously (Ding et al., 2016; Jin et al., 2018). However, the field is only beginning to understand which moments in speech are key for successful perception. Many of these prosodic events in speech have been shown to follow an approximately rhythmic pattern that matches the internal neural oscillations which coordinate activity in motor and auditory cortices (Assaneo & Poeppel, 2018; Ghazanfar & Takahashi, 2014; Shepherd et al., 2012). For example, syllable rate corresponds to the 4–7 Hz theta rhythm in adults, and lexical and phrasal-level prosodic modulations correspond to the $<$ 4 Hz delta band of neural activity (Ding et al., 2016; Wilsch et al., 2018). In adults, alignment between neural activity in these frequency bands and speech reflects selective attention to the speech stream (Ding & Simon, 2012; Horton et al., 2014; Mesgarani & Chang, 2012).

However, it is not currently clear if selective attention necessarily drives entrainment. The most direct work linking attention and entrainment has focused on top-down selective attention rather than bottom-up salience (Ding & Simon, 2012; Horton et al., 2014; Mesgarani & Chang, 2012). In addition, some entrainment to speech is observed even in participants who are asleep (Makov et al., 2017), which suggests that top-down selective attention to speech is not a prerequisite for entrainment. However, even in participants who are asleep, entrainment likely reflects processing of short-timescale auditory information (e.g., syllables and words; Makov et al., 2017). For example, sleeping adults showed neural tracking and detection of syllable and word boundaries in an artificial speech stream (Batterink & Zhang, 2022). In awake participants, there is evidence that stronger entrainment facilitates perception and segmentation of sounds and syllables (Giraud & Poeppel, 2012; Gross et al., 2013; Peelle et al., 2013) and improves comprehension and intelligibility (Dauer, 1983; Drullman, 1995). When this form of alignment becomes disrupted, there can be impairment in speech perception (Riecke et al., 2018; Vanthornhout et al., 2018; Zoefel et al., 2018). But, although better alignment is generally associated with better low-level auditory perception, it is not always a marker of comprehension or ease of processing. Evidence for this comes from studies showing stronger alignment when L2 learners are listening to their non-native as opposed to their native language (Reetzke et al., 2021; Song & Iverson, 2018), as well as studies showing that language comprehension under noisy conditions slightly suppresses cortical tracking of lower-level features of speech (Zou et al., 2019), and stronger entrainment in hearing-impaired and older adults (Palana et al., 2021). This work indicates that speech perception, attention, entrainment, and comprehension are related but not identical processes, and future research is needed to elucidate their interplay at both shorter and longer timescales in speech.

Alignment between speech and neural activity may be especially important when speech perception is made difficult by the presence of background noise (Ding & Simon, 2013), multiple speakers (Ding & Simon, 2012; Horton et al., 2014; Mesgarani & Chang, 2012), and non-native speech (Reetzke et al., 2021; Song & Iverson, 2018). Thus, in order to effectively attend to and process speech, listeners need to track its inherently complex rhythms at the level of sounds, syllables, words, and sentences, each of which play a role in enabling comprehension during natural communication. We currently know little about how these processes interact and develop to enable early language learning and, ultimately, language proficiency. While existing research has advanced knowledge of adult attentional dynamics, minimal research has examined the first years of life when speech and attention are likely to be deeply interwoven.

In three sections, we emphasize the importance of studying the moment-to-moment interplay between speech and entrainment in order to harness how speech input scaffolds development and learning at multiple timescales (from milliseconds to years). The first section combines findings from adult speech processing and neural development to discuss how IDS facilitates infants' entrainment to speech, as well as how caregivers position important information in moments when infants may be more likely to show successful perception and/or attention. The second section highlights the cumulative, developmental-timescale effects of in-the-moment attention to speech by suggesting that variation in alignment between neural activity and auditory signals may be powerful in explaining individual differences in children's language learning trajectories. The third and final section discusses future research directions for exploring how the short-timescale structures of IDS support infants' attention, communication, and learning across longer timescales. This framework provides a mechanistic understanding of what defines "high-quality speech" to infants, which has been emphasized in recent developmental science as key to successful learning outcomes (e.g., Hirsh-Pasek et al., 2015; Romeo et al., 2018; Weisleder & Fernald, 2013).

How does infant-directed speech support speech processing in infancy?

Alignment between external stimuli and neural oscillations appears to be automatic in development, but it is much more likely when stimuli are tailored to infants' developing cognitive abilities. In this section, we show evidence that IDS optimizes alignment in infants, and we draw on work from attention and neural development to illuminate how IDS may result in enhanced time-locked attention.

As one of the most important sources of language input in early development, IDS has been hypothesized to capture infants' attention and thus facilitate processing and learning (Fernald, 1992; Grieser & Kuhl, 1988). Young children prefer listening to IDS compared to ADS (Cooper & Aslin, 1990; Fernald, 1985; Kaplan, Goldstein, Huckeby, Owren, & Cooper, 1995; ManyBabies Consortium, 2020; Pegg, Werker, & McLeod, 1992; Werker & McLeod, 1989; Bergelson et al., 2017), suggesting that IDS captures and sustains children's attention for longer than ADS. While these studies measured overall interest in IDS, infants' interest may result from activity occurring at very short timescales, i.e., from heightened time-locked attention and better in-the-moment processing. This is likely facilitated by improved neural alignment to the speech signal.

Behavioral evidence indeed suggests improved processing of IDS compared to ADS. For instance, in noisy conditions, infants detect IDS more easily than ADS (Colombo et al., 1995; Newman, 2003). Trainor and Desjardins (2002) showed increased discrimination of vowels in IDS vs ADS in infants, and suggested that these effects are likely due to heightened attention, rather than easier vowel discrimination. Additionally, when stimuli are presented using IDS prosody, infants and toddlers perform better on tasks of word segmentation (Nelson et al., 1989; Thiessen et al., 2005), familiar word recognition (Song et al., 2010) and novel word learning (Graf Estes & Hurley, 2013; Ma et al., 2011).

Are these improvements in processing due to better neural alignment to IDS? While the link between speech processing and neural alignment to speech has not been directly tested in infancy, several studies with infants have observed stronger cortical tracking of IDS compared to ADS (Kalashnikova et al., 2018a) and increased magnitude of neural response to IDS vs ADS in frontal regions (Saito et al., 2007; Santesso et al., 2007). Similarly, toddlers showed stronger time-locked pupillary responses to a story told with IDS prosody compared to a story told in ADS prosody (Nencheva et al., 2020). These findings collectively suggest that IDS optimizes time-locked attention to speech. However, what features of the speech signal may be driving this?

IDS is characterized by a number of concurrent modifications relative to ADS (Fernald, 1992; Gleitman et al., 1984; Grieser & Kuhl, 1988), such as higher and more variable pitch (Fernald & Simon, 1984), slower rhythm (Leong et al., 2017; Narayan & McDermott, 2016), vowel hyperarticulation (Kuhl et al., 1997; Werker et al., 2007), enhanced and more positive emotional content (Fernald & Kuhl, 1987; Scherer et al., 1991), shorter utterances (Martin et al., 2016), and modified timbre (Piazza et al., 2017). While each modification may serve to facilitate communication, social bonding, and/or language learning, here we focus on how these modifications affect moment-to-moment attention to speech. Research linking specific features to in-the-moment attention is still somewhat scarce, therefore we only cover a subset of the key features of IDS, as follows.

Rhythm

IDS is characterized by several modifications of the rhythms of prosodic stress and syllable rate compared to ADS. These rhythmic modifications may support processing in infants in three ways. First, its more evenly-timed rhythm (Payne et al., 2015) facilitates entrainment. Second, its slower rhythm (Narayan & McDermott, 2016) may be specifically catered to the internal neural rhythms that are most prominent in infancy. And third, alignment of the rhythms of syllables and prosodic stress (Leong et al., 2017) may help infants link information across different timescales of speech, such as lower-level acoustic processing with higher-level phrase processing.

Temporal regularity

The even timing of syllables in IDS (Payne et al., 2015) may facilitate infants' processing by providing a regular rhythm that is easier to entrain to. In adults, it is well-established that entrainment is stronger when signals have a more regular rhythm (Haegens & Zion-Golumbic, 2018). For example, when presented with beeps embedded in white noise, participants detect rhythmic (vs arrhythmic) beeps at a lower threshold, and this improvement in perception results from stronger entrainment to the rhythmic stimulus (Ten Oever et al., 2017). We know less about how temporal regularity affects infants' entrainment to speech, but there is evidence that it improves infants' speech processing. Infants are better able to segment a speech stream of unfamiliar words when the word length is kept constant, keeping the rhythm of word onsets constant (Lew-Williams & Saffran, 2012). Similarly, infants can use stress cues, an inherently rhythmic property of speech, to segment continuous speech (Jusczyk et al., 1993; Thiessen et al., 2005; Thiessen & Saffran, 2007) implying that infants' entrainment to stress cues is flexible and shaped by the input. While these studies do not directly test entrainment, infants' enhanced processing of rhythmic speech and the tighter entrainment to rhythmic stimuli in adults point to the possibility that entrainment to the stress rhythm of speech can support processing and learning. By providing a more regular rhythm, IDS supports infants' processing and learning, possibly by improving infants' neural alignment to speech.

Slower rhythm

In addition to being more regularly timed, the rhythm of IDS is also slower. This may be optimal for the organic neural rhythms present in the developing brain. Specifically, of the delta and theta frequency bands that are associated with speech processing in adults, infants show relatively stronger power in the slower delta band vs the slightly faster theta band, and this difference decreases as infants age from 2 weeks to 2 years (Sankupellay et al., 2011), up to 5 years (Marshall et al., 2002; Whitford et al., 2007) and into adulthood (Uhlhaas et al., 2010). Moreover, coherence in the theta rhythm (corresponding to faster syllabic rate) in response to auditory stimuli increases from 4 and 7 years of age (Ríos-López, 2018). The more prominent delta rhythm may also be closer in frequency to one of the most prominent auditory rhythms in utero, maternal heart rate (1–2 Hz; Bench, 1968), which has been shown to elicit event-related responses in utero (Porcaro et al., 2006). In addition, the frequency of the theta band may be slower in infants (3.6–5.6 Hz) compared to children and adults (4–8 Hz; Orekhova et al., 2006).

Because infants have increased power in slower neural oscillations and a slower theta rhythm, IDS *should* have a slower rhythm, otherwise infants' attention and processing would be frequently misaligned. Indeed, Narayan & McDermott (2016) found slower speech rate in IDS compared to ADS in Sri Lankan Tamil, Tagalog, and Korean. Similarly, Leong et al. (2017) compared the rhythmic properties of amplitude modulation in Australian English speech directed at 7- to 19-month-old infants, as well as ADS at three different timescales corresponding to the rate of phonemes, syllables, and prosodic stress. They found that IDS was characterized by a slower, stress-shifted rhythm, close in frequency to the delta band of neural oscillations, compared to the faster, syllable-dominated theta rhythm of ADS. This adaptation facilitates processing, as infants attend more to slower than to faster IDS even with prosody held constant (Panneton et al., 2006), and they are better at recognizing familiar words embedded in slower IDS (Song et al., 2010).

Alignment across timescales of speech

In addition to adapting to neuro-developmental changes, the rhythmic structure of IDS may help infants' brains link information across different timescales (e.g., across syllables and phrase-level prosodic stress) by coordinating neural activity across them

(Goswami, 2019; Peelle & Davis, 2012). That is, in addition to facilitating low-level perception, alignment between speech and neural rhythms across different frequency ranges may coordinate processing across multiple brain regions necessary for simultaneous comprehension at different timescales in speech (Giraud & Poeppel, 2012). This allows concurrent processing of the hierarchical structures of speech, spanning syllables, words, phrases and sentences (Ding et al., 2016). Leong et al. (2017) found that IDS, compared to ADS, aligned more strongly the faster rhythm of syllables with the slower rhythm of prosodic stress. This alignment resulted in more robust and hierarchically-nested rhythms compared to ADS. Such coordination of multiple neural oscillatory rhythms, often originating from different parts of the brain, has been implicated in facilitating tasks that require distributed processing of information (Palva et al., 2005; Palva & Palva, 2018), such as speech perception and comprehension (Ding et al., 2016). Thus, by coordinating distributed neural activity at different frequencies, IDS may allow infants to process shorter and longer timescales of speech in parallel.

Nursery rhymes, a specific type of IDS, may be particularly optimized for such synchronization across neural rhythms by providing even stronger coupling between the rate of syllables and prosodic stress (Leong et al., 2011). Leong et al. (2011) passed nursery rhymes through a low-frequency filter, and found that rhythms at different frequencies were time-locked (e.g., a faster rhythm corresponding to syllables was time-locked to a slower rhythm corresponding to stress; see also Goswami & Leong, 2013). This coordinated timing across timescales may support optimal entrainment and help the developing brain coordinate the processing of information at different timescales. Similarly, the heightened use of diminutives in IDS, such as “doggy” or “kitty”, may increase the temporal regularity of stress patterns (Echols et al., 1997; Jusczyk, 1997; Kempe et al., 2005). Diminutives in English (and certain other languages) have a consistent strong–weak pattern of stress, which may increase rhythmic regularity especially when these words are presented in isolation. Kempe et al. (2005) found that when English speakers listened to an unfamiliar speech stream in Dutch, word segmentation was facilitated both by this consistent stress pattern and by the consistent endings of diminutive words.

By providing rhythmic cues at the appropriate frequencies for infant perception and by scaffolding processing at multiple timescales, IDS may facilitate sustained attention to multiple features of the speech stream and allow optimal processing of the incoming signal.

Pitch

IDS is characterized by a generally higher and more variable pitch compared to ADS (Fernald, 1992; Fernald & Simon, 1984; Gleitman et al., 1984; Grieser & Kuhl, 1988). These two pitch modulations - higher frequency and increased variability - may serve different functions (Trainor & Desjardins, 2002) and impact infants’ entrainment to speech through different routes. Even though by itself the overall higher pitch does not affect the moment-to-moment dynamics of speech that could facilitate entrainment, it may provide a signal that is easier for the infant brain to detect. Infants’ perception of sounds with higher pitch matures earlier than their ability to perceive other parts of the spectrum (Schneider & Trehub, 1992). In addition, the high-pitched IDS signal may help infants tune in by providing information about emotion, as well as by signaling that the caregiver is addressing the infant, an idea we return to in later sections.

While higher pitch supports tuning in, the variable pitch of IDS may support entrainment, but via a different route. There is evidence that pitch variability is important for successful speech processing in adults; for example, speech intelligibility under noisy conditions is significantly reduced when pitch variability is flattened (Binns & Culling, 2007; Laures & Bunton, 2003). There is some evidence that pitch variability may improve speech intelligibility for infants as well. For example, infants are better at discriminating between vowels embedded in exaggerated pitch contours (Trainor & Desjardins, 2002). Even though the link between entrainment to speech and pitch variability has not been directly tested, previous work has shown that disruptions to intelligibility are typically associated with disruptions to entrainment (Riecke et al., 2018; Vanthornhout et al., 2018; Zoefel et al., 2018).

Räsänen et al. (2018) found that IDS has less *predictable* pitch, compared to ADS. They hypothesized that this unpredictability in pitch throughout the speech stream may serve to capture and sustain infants’ attention, much as prosodic stress or emphasis do in ADS. Thus, while a predictable temporal rhythm of amplitude modulations and syllable rate optimizes processing, *un*predictable pitch changes increase the salience of speech. Chang et al. (2016) found that when adult participants listened to a series of rhythmically presented tones that varied in pitch, power in neural oscillations associated with pitch perception increased in response to tones with unpredictable or surprising pitch. Similarly, participants showed stronger entrainment to music with unpredictable vs predictable pitch (achieved by scrambling notes; Kumagai et al., 2017; Meltzer et al., 2015). Behaviorally, the unpredictability of pitch in speech has been associated with perceptual salience (Kakouros et al., 2018; Kakouros & Räsänen, 2016). Together, these adult findings on the increased intelligibility and salience of speech with variable pitch, as well as the stronger neural tracking of unpredictable tones, suggest that variable pitch in IDS is very likely to strengthen infants’ entrainment and facilitate speech processing.

Formant exaggeration

In addition to providing distinctive prosodic cues, caregivers often produce formant-exaggerated speech by hyper-articulating vowels (Eaves et al., 2016; Kuhl et al., 1997; Werker et al., 2007). This adjustment continues in early childhood and decreases with age (Liu et al., 2009). While prior work has primarily focused on the role of formant expansion in the formation of separable vowel categories (e.g., Vallabha et al., 2007), it may additionally serve to facilitate in-the-moment attention to speech by increasing phase-locking between neural oscillations and speech dynamics. Evidence for this comes from a study of 6- to 12-month-old infants who showed increased synchronization in delta and theta neural rhythms in response to formant-exaggerated syllable streams (Zhang et al., 2011). In this study, the authors controlled for overall pitch and pitch variability, thus isolating the effects of formant exaggeration. While it is unknown how formant-exaggeration facilitates time-locked neural responses to speech in infants, Zhang et al. (2011)

speculate that this effect may result from increased attentional salience and phonetic distinctiveness of formant-exaggerated speech. In addition, because intelligible speech is associated with stronger entrainment compared to unintelligible speech (Peelle et al., 2013; Riecke et al., 2018; Vanthornhout et al., 2018; Zoefel et al., 2018), hyper-articulation may serve to simultaneously strengthen time-locked entrainment and increase intelligibility. This idea is supported by research showing that hyper-articulated IDS helps infants recognize familiar words (e.g., Song et al., 2010), suggesting improved intelligibility. Infants whose Mandarin-speaking mothers used more formant expansion had better speech discrimination abilities (Liu et al., 2003). Similarly, infants (7 and 10 months) at risk for dyslexia heard less hyperarticulated speech from their mothers and showed poorer auditory discrimination abilities, compared to infants who were not at risk for dyslexia (Kalashnikova et al., 2018b).

Emotional content

Another key component of IDS that has been implicated in capturing and sustaining infant attention is its emotional content. Caregivers use IDS in part to regulate infants' emotional states (Fernald, 1989; Smith & Trainor, 2008; Spinelli & Mesman, 2018). In particular, many researchers have pointed out the striking acoustic similarities between IDS and positively valenced (or "happy") ADS (Fernald, 1989, 1992). In fact, Trainor et al. (2000) argued that the prosodic differences between IDS and ADS are explained mainly by differences in expressed emotion. For instance, IDS utterances conveying positive affect are more prevalent than those conveying negative affect during the first year of life (Kitamura & Burnham, 2003), and the more positive tone of IDS may serve to capture and sustain infant attention. Young infants prefer speech with comforting or approving intent over speech with directive or disapproving intent (Kitamura & Lam, 2009; Papoušek et al., 1990). Similarly, even when controlling for prosodic characteristics, infants prefer high- vs low-affect speech (Kitamura & Burnham, 1998), and happy over neutral speech both in IDS and ADS (Panneton et al., 2006; Singh et al., 2002). Further, infants are better at recognizing words embedded in happy vs neutral speech (Singh, 2008).

The higher emotional content of IDS is likely to promote time-locked responses to speech. Adult emotional processing points to increases in alignment of neural rhythms during emotional processing, such as empathy (Mu et al., 2008), detection of emotion change (Chen et al., 2012), emotional memory (Headley & Paré, 2013) and processing emotional stories (Kang & Wheatley, 2017). This effect is not limited to adults. Infants too show increased synchronization of neural oscillations and increased power in the theta band during emotional behaviors (Futagi et al., 1998; Maulsby, 1971; Posikera et al., 1986) and when listening to emotional speech (Orekhova et al., 2006; Santesso et al., 2007).

IDS not only provides prosodic cues to emotion, but is often accompanied by concurrent facial and gestural cues to emotion (Brand et al., 2002; Kosie & Lew-Williams, 2022). Such multimodal cues have been found to enhance entrainment across multiple frequency bands (for a review see Symons et al., 2016). In fact, according to some scientists, the striking parallels in frequency between the syllable and stress rates in speech and the rhythms of facial expressions and movements in non-human primates suggest an even deeper evolutionary link between affective and speech cues (Ghazanfar et al., 2013; Ghazanfar & Poeppel, 2014; Ghazanfar & Takahashi, 2014; Shepherd et al., 2012). By increasing selective attention and synchronizing activity across multiple neural oscillation bands, the emotional content of IDS supports infants' real-time processing of speech.

Effects of familiarity and informativeness of speech

Infants may also prefer listening to IDS over ADS because of its own history of being familiar, engaging, and/or informative. Over time, the increased familiarity and relevance of a speech register directed at them may lead infants to learn its prosodic characteristics and selectively attend to IDS. There is evidence that early experience with a specific speaker can lead to a preference for their familiar voice vs unfamiliar voices (Barker & Newman, 2004; DeCasper & Fifer, 1980; Friedlander, 1968). Infants show higher entrainment to familiar vs unfamiliar speakers (Purhonen et al., 2004) and increased activation in frontal and temporal regions (Naoi et al., 2012), suggesting that heightened attention to a stimulus that the child has experienced can increase time-locked responses to the speech stream. Cirelli et al. (2016) directly tested the role of prior experience on entrainment by comparing infants who attended music classes to those who did not; infants in music classes showed increased entrainment to musical stimuli. Similar effects of musical experience have been shown in adults (Pallesen et al., 2015). The evidence that experience with a specific stimulus increases entrainment suggests that, as infants experience IDS across time, they should entrain to it more strongly and therefore process it more effectively. Similarly, infants who hear more IDS in their household or community environment should entrain to it more strongly. No research has directly tested the effect of experience on preference for or learning from IDS vs ADS (e.g., by relating an infant's preference for IDS to the amount or features of IDS they were exposed to); however, it is likely that infants would most optimally entrain to and learn from speech that matches the characteristics of the IDS they hear at home, akin to effects related to music and to specific speakers (Cirelli et al., 2016; Purhonen et al., 2004). As infants become better at entraining to IDS, this signal may in turn become more informative, further supporting their attention to its important features in a self-reinforcing manner.

Optimal placement of important information in IDS

Because of its rhythmic nature, attention fluctuates from one moment to the next during the perception of continuous stimuli (Fiebelkorn & Kastner, 2019). Even if IDS optimizes the timing of these fluctuations, there are still, by definition, moments of stronger vs weaker attention. Here we present evidence that caregivers not only globally facilitate infants' sustained attention by using the IDS register, but also strategically position important information at optimal moments in the ebb-and-flow of rhythmic changes in attention.

Prosodic emphasis

New or important labels in IDS often coincide with pitch and amplitude peaks, which signal prosodic emphasis (Aslin, 1993; Fernald & Mazzie, 1991; Messer, 1981). There is evidence that prosodic emphasis of words is associated with increased salience (Kakouros et al., 2018; Kakouros & Räsänen, 2016) and enhanced processing (Cutler & Foss, 1977; Li & Ren, 2012; Wang et al., 2011). It is likely that such improved processing stems from increased attention at amplitude peaks due to infants' rhythmic alignment to the stress patterns of speech (Kalashnikova et al., 2018a), as well as phase resetting in response to unpredictable pitch changes (Chang et al., 2016). Räsänen et al. (2018) argue that the same mechanisms guiding sustained attention to unpredictable pitch variations in IDS also define increased attention to and memory for emphasized words. Indeed, Nencheva et al. (2020) found increased time-locked pupillary responses and better learning of novel word labels during pitch peaks compared to other prosodic contours. Similarly, Grassmann & Tomasello (2007) showed that toddlers are better at learning novel labels positioned in a moment with (vs without) prosodic emphasis. This suggests that caregivers optimize children's learning by aligning new information to moments that elicit heightened attention.

Repetition

Even without strategically positioning labels in the most attended-to moments in an utterance, caregivers often repeat words (Brodsky & Waterfall, 2007; Hills, 2013; Schwab et al., 2018), which increases the likelihood of a novel label being received at a moment of heightened attention. Indeed, repetition in close succession supports learning (Schwab & Lew-Williams, 2016a). When controlling for prosodic characteristics, McRoberts et al. (2009) found that 6-month-olds preferred listening to IDS that contains verbal repetitions. Repetition may be especially important for young infants, whose ability to entrain to the dynamics of speech may still be developing. Caregivers' use of repetition decreases as children's vocabulary grows, which allows for a greater diversity of word tokens (Schwab et al., 2018).

Overall, this section has reviewed several mechanisms through which IDS may facilitate processing and learning. Specific features of IDS, such as its slower and more consistent rhythm, variable pitch, formant exaggeration, and emotional content, likely drive alignment between its dynamics and neural activity, which in turn allows infants to optimally process and learn from the speech stream. Infants' experience with this informative signal – both its specific features and overall – may shape a preference for selectively attending and entraining to IDS, enabling them to optimize their learning of new information positioned at salient moments. By facilitating infants' in-the-moment processing of IDS across milliseconds and seconds, caregivers' modifications may support learning over the course of many short interactions.

Effects of entrainment on young children's language learning

Entrainment to speech optimizes perception in the moment. In this section, we draw from research on individual differences to showcase how the impact of entrainment on in-the-moment processing aggregates over time and affects long-term language development. Additionally, we propose that the quality and quantity of infants' language input iteratively shapes their internal neural rhythms over time, which helps them become increasingly efficient in processing speech.

Stronger alignment shapes young children's learning in the moment

Consistent with long-standing evidence that attention improves learning (Posner et al., 2004), research shows that greater entrainment to speech – a result of increased attention – supports both learning and memory. To explore this relationship, researchers have quantified time-locked attention as alignment *between participants* in neural or pupillary responses during a continuous stimulus. For adult participants, stronger inter-subject alignment in neural activity during a story predicts better memory for story details (Simony et al., 2016). Nencheva et al. (2020) recorded the pupil responses of 24- to 30-month-old toddlers listening to sentence-final novel words with varied prosodic contours. Toddlers learned a novel word better if they had been closely aligned to other participants in their pupillary responses when hearing the word in the first place. This suggests that higher time-locked attention to speech resulted in better learning. Piazza et al. (2021) found similar results in 3- and 4-year-olds using functional near-infrared spectroscopy (fNIRS); children who exhibited higher inter-subject neural alignment during story-book reading both learned and remembered the novel words better. These findings suggest that time-locked attention to speech predicts young children's learning of novel information.

Individual differences in alignment strength correlate with long-term language outcomes

Language learning occurs through learning over many moments, spread across many months and years. Therefore, optimized alignment of attention across small moments is likely to accumulate over the course of development and have long-term effects on language development. Regardless of the underlying processes resulting in impaired entrainment to speech, it has been hypothesized that imperfect time-locked attention to speech may have compounding consequences for language development and, in some populations, result in language delays. In this section, we will review the existing (but in some cases, sparse) evidence linking individual differences in entrainment to language outcomes in typically and atypically developing populations. To date, more research has been conducted with the latter.

While individual differences in entrainment to speech have been rarely studied in typically developing infants, there is existing evidence that individual differences in infants' rhythmic processing, which relies on entrainment, predict language development. Infant's processing of prosodic changes associated with speech rhythms predicts vocabulary development at three years of age

(Kalashnikova et al., 2019), and rhythm processing predicts children's grammar skills (Gordon et al., 2015). In addition, musical training, which is associated with improved rhythmic perception, facilitates phonological processing (Tierney & Kraus, 2014), and the ability to behaviorally and neurally follow a rhythm predicts literacy in early readers (Bonacina et al., 2018). Further, children's cortical tracking of speech rhythm predicts their later reading performance (Ríos-López et al., 2022). Young children's ability to entrain to external stimuli could plausibly mediate the link between rhythmic processing and language outcomes, however, further research is needed to understand the extent to which individual differences in entrainment in typically developing populations track with individual differences in language outcomes.

More powerful evidence for the long-term importance of successful entrainment for language development comes from the study of atypical populations with language delay. In particular, language delay symptoms in children with dyslexia and with autism spectrum disorder (ASD) have been linked to difficulties in entrainment to speech. The cause of impaired entrainment may differ across children with different developmental delays and disorders, but the resulting language delays and difficulties may nonetheless present in comparable ways.

Children with dyslexia show early and persistent deficits in entrainment to speech, measured as the alignment in neural oscillations in the corresponding frequencies to the rate of syllables and words in natural speech (Abrams et al., 2009; Leong & Goswami, 2014; Power et al., 2016; Soltész et al., 2013). The impairment in entrainment to speech in developmental dyslexia has been hypothesized to result from impaired processing of subtle changes in prosody (Hämäläinen et al., 2005) and rhythm (Thomson & Goswami, 2008). This imperfect alignment to speech in children with dyslexia may lead to worse processing and segmentation of phonemes (Leong & Goswami, 2015) and, over time, give rise to characteristic deficits such as phonological processing and reading or spelling difficulties (Goswami, 2002, 2011). Indeed, the ability of children with dyslexia to process prosodic changes predicts later symptoms (Goswami et al., 2013), and the extent of entrainment to speech in children with dyslexia is positively associated with literacy (Power et al., 2016).

ASD is linked to deficits in prosodic processing (Brock et al., 2002; Simon & Wallace, 2016), auditory processing (Otto-Meyer et al., 2018), entrainment to audiovisual streams of speech (Jochaut et al., 2015), and long-range synchronization of neural networks (Dinstein et al., 2011). For example, one study presented an audio-visual scientific documentary to adolescents and adults with an ASD diagnosis and measured neural activity using combined fMRI and EEG (Jochaut et al., 2015). Participants in the ASD group showed impaired cortical tracking of speech in the theta band, as well as altered connectivity between lower-level auditory regions and other regions involved in language processing (e.g., Broca's area). Benítez-Burraco & Murphy (2016) propose that impaired entrainment in ASD may be causally linked to language difficulties. Together, these findings from typical and atypical populations provide evidence that, by optimizing in-the-moment processing, entrainment to speech facilitates language acquisition and can have long-term consequences for development.

How can a dynamic attention framework guide the future of research on early communication and learning?

Exploring IDS at the sub-second scale can bring a new perspective on the types of 'high-quality' input that support language learning (e.g., Hirsh-Pasek et al., 2015; Romeo et al., 2018). While short-timescale effects on attention are small and fleeting, there is evidence that they accumulate over time and have global effects on entrainment to the speech signal, both in typical and atypical populations. The key ideas are that (1) the efficiency of processing speech in the moment, facilitated by IDS, builds up across time, allowing infants to optimally process new information, including new words, and (2) by supporting the neural connections necessary for processing complex input at multiple timescales, IDS may prepare the developing brain for later processing of ADS. This dynamic-attention framework may allow us to understand the shape of language development – and the nature of individual differences – over increasingly broad windows of learning.

This poses two broad questions for future research: What features of natural language input and everyday caregiver-child interactions support optimal in-the-moment attention and learning? And how do the effects of moment-to-moment processing shape long-term language development? Understanding how different properties of speech affect entrainment can allow us to go beyond categorical comparisons of learning and processing between IDS and ADS, and to make predictions about what 'high-quality' and/or optimized learning moments may look like during natural communication – with higher temporal precision than has been possible in recent decades. The field has uncovered evidence that appropriately timed labels and attention to objects and people shape what is encoded and retained in memory (Goldstein & Schwade, 2009; Suarez-Rivera et al., 2019; Yu & Smith, 2012). As we continue to advance our understanding of the interplay between visual and auditory attention, we will begin to uncover how short-timescale dynamics build up across developmental time to realize the full value of caregiver-infant interactions in the real world.

To bring the field closer to this goal, we think five research directions will be important. Below we elaborate on each in turn. First, we call for experimental studies that examine how specific features of IDS drive entrainment, and we outline different data collection methods that could be used to accomplish this. Second, we highlight the need for experiments that test entrainment-based predictions about which moments in speech may be optimal for learning. Third, we frame the importance of understanding the cumulative effects of these moment-to-moment processes on long-term language development. Fourth, we call for studies that go beyond speech and harness the dynamics of the multimodal cues of natural caregiver-child interactions to understand how auditory and visual attention dynamically support learning. Finally, we emphasize the importance of understanding how caregivers and infants dynamically co-create signals that support real-time communication.

What features of IDS drive entrainment?

We have outlined what is currently known about the effects of various features of IDS on infants' entrainment. However, there are still many features we know little about. For example, how is entrainment affected by modifications in timbre (Piazza et al., 2017), changes in pitch (Fernald, 1992; Gleitman et al., 1984; Grieser & Kuhl, 1988), and different combinations of isolated words and longer utterances (Brent & Siskind, 2001; Phillips, 1973)? Because most of the work on entrainment to speech is based on adults, we know little about the changes in entrainment and the features of speech that optimize it across different ages. There is limited work on how specific features of IDS may be perceived differently by the developing brain at different ages, making it hard to determine what counts as a key feature of speech for successful entrainment. There is growing evidence that the extent to which caregivers use various features of IDS, such as prosodic exaggeration (Vosoughi & Roy, 2012), vowel and acoustic exaggeration (Lui et al., 2009), emotion and pitch (Kitamura & Burnham, 2003), and repetition (Schwab et al., 2018) and speech rate (Narayan & McDermott, 2016), changes from infancy through early childhood, and some evidence that children's preference for features of IDS changes with age (ManyBabies Consortium, 2020; Segal & Newman, 2015; Hayashi et al., 2001). However, there are still gaps in our understanding of the links between these processes and neural development. In order to study the development of entrainment to speech, developmental scientists and language scientists can use several techniques.

First, scientists can *directly compare alignment between speech rhythms and frequency bands of neural activity* using an electroencephalogram (EEG – e.g., Kalashnikova et al., 2018a). EEG currently provides the most temporally precise measure of entrainment. Second, researchers can measure *inter-subject alignment* (e.g., Nencheva et al., 2020) using less temporally granular measures of processing (on the order of 1–2 s) such as fNIRS, functional magnetic-resonance imaging (fMRI) and pupillometry. The main methodological insight in this approach is that when participants are entrained to the same stimulus, their neural activity becomes increasingly aligned to that of other participants. This leads to a similar physiological 'fingerprint' of summary neural activity over both shorter and longer intervals (provided by hemodynamic and pupillary responses). This technique is useful for studying effects of different stimulus features on processing, even in infants, because it reduces individual-level noise and highlights the common response driven by the unfolding dynamics of a stimulus. Third, even in the absence of direct neurophysiological measures of entrainment, researchers may indirectly quantify evidence for stronger or weaker entrainment by exploring the *intelligibility of speech in real-world noise*. This approach is based on the tight correlation between entrainment and intelligibility in noise observed in prior research (Vanthornhout et al., 2018). Real-world speech input to infants is so often accompanied by noise (McMillan & Saffran, 2016), and signals like IDS are more intelligible under noise compared to ADS (Colombo et al., 1995). Measuring the intelligibility of speech with varying features under noisy conditions can be used as a proxy for the ease of entrainment to the speech signal and reveal how these features enhance speech processing in a more naturalistic setting.

By varying the features of IDS in experimental settings, exploring their effects on alignment of neural and physiological activity across participants, and examining speech intelligibility in natural, messy conditions, we can further uncover how IDS drives optimal time-locked attention, especially across milliseconds and seconds.

How does entrainment to IDS shape learning?

Beyond identifying the features of IDS that drive entrainment, the next step will be to understand how entrainment at different moments in speech shapes the ability to encode new information. What local parameters of natural speech enhance infants' encoding and learning? To answer this, we can design experiments that probe attention *and* the ability to learn at different moments in the unfolding speech stream. This will involve drawing on natural IDS to understand the parameter space of various features of speech and the (meaningful) moment-to-moment patterns of variation, as well as manipulating the local features of speech surrounding a novel label in experimental conditions. These types of studies can both create new ideas about which moments are optimal for learning as well as elucidate the mechanism for why certain caregiver behaviors are effective.

What kinds of structures in speech are key for processing at different ages and how do infants navigate incoming speech structures at multiple timescales? One promising approach would be to understand how different kinds of prosodic cues (e.g., peaks or troughs of intensity or pitch) interact to shape optimal moments for learning (e.g., a new word) over the course of development. For instance, caregivers often place novel labels at the end of an utterance and this appears to facilitate retention of the label (Aslin, 1993; Aslin et al., 1996; Fernald & Mazzie, 1991; Messer, 1981). While the positioning of labels at the end of an utterance has been argued to support word segmentation, the role of attentional processes in children's extraction of information is not well-understood. For example, the accumulation of rhythmic cues during the start of the sentence may prepare infants to optimally tune into the structure of the speech signal by the time the new label is introduced. There is evidence that entrainment to a preceding rhythm leads to better processing (Cason & Schön, 2012), suggesting that longer exposure to a congruent rhythm preceding a target could facilitate its perception. Since perfect alignment between neural oscillations and speech is not instantaneous, it is possible that positioning information later in a sentence would make use of better alignment achieved over the first portion of the sentence. Additionally, infants' processing of labels in this position may be independently shaped by the patterns of pitch variation over the course of the utterance. Räsänen et al. (2018) showed that in IDS (compared to ADS) changes in pitch are *less* predictable at the end of the utterance, which, as discussed in earlier sections, may increase the salience of key words, and support their learning via entrainment processes. An experimental study could disentangle the effects of these different prosodic features on shaping the learning of labels in an utterance-final position by systematically varying utterance-level prosodic features with the position of the label. This is just one example of how examining entrainment can provide a mechanistic explanation for past findings in developmental science and generate new predictions.

Long-term effects of language input on children's entrainment

The ultimate goal of examining the moment-to-moment dynamics of learning from speech input is to provide a mechanistic account of children's long-term language-learning outcomes over many such moments, spanning years of development. Here we propose two future directions that examine how patterns of short-timescale dynamics of early speech input can cumulatively shape children's long-term development.

An increasingly robust body of research has tackled the question of how variation in the quantity and quality of input children receive from their caregivers shapes their language learning outcomes (Romeo et al., 2018; Rowe, 2008, 2012; Schwab & Lew-Williams, 2016b; Weisleder & Fernald, 2013). Children who hear more speech on average produce more words (Rowe, 2012), but not all speech contributes equally to children's vocabulary growth. The amount of speech directed *to* the child is more predictive of language development than overheard speech (Weisleder & Fernald, 2013). Quality of input in these studies has been examined categorically (for example, by comparing IDS vs ADS), without much attention to the features that caregivers use within IDS or where they position new labels within the speech stream. This is partly because examining these questions even at a very coarse temporal granularity is time-intensive and costly. Even to study the quantity of speech, researchers need large amounts of data from the child's home environment and laborious manual and/or automated coding to tag IDS vs ADS; the demands of transcribing and analyzing acoustic characteristics of speech input under the noisy and messy conditions is an added burden. But this effort will be worth it. Moment-to-moment patterns in caregiver speech input are likely to add up to have significant long-term consequences on learning. Even controlling for overall amount of IDS, a child whose caregiver uses features that facilitate entrainment and positions labels in moments that are optimal for learning over time may learn more words than a child whose caregiver uses IDS in ways that are less optimal for learning.

This is likely to be the case in cultures and communities where caregivers often speak to infants, but it may also be true in cultures where caregivers do not usually speak directly to infants, as infants must nonetheless use ambient speech to break into the structures of sounds, words, and sentences (e.g., Casillas et al., 2020). Further, as discussed in earlier sections, the effects of IDS features on attention and learning may be in part due to the child's own experience and familiarity with the signal. This idea is reflected in the finding that infants' preference for IDS over ADS is predicted by the quality of their caregivers' IDS (Outters et al., 2020). Therefore, the features of IDS that are most optimal may result from a dynamic interplay between experience with specific features of the language environment and individual differences in neural processing. Greater understanding of this interplay at the level of individual children can push the field to increasingly precise, individualized interventions that are sensitive to the child's cultural environment. We hope that advances in automatic transcription and machine learning will reduce the need for manual coding of the timing and content of utterances and words, and thereby enable researchers to examine and define 'high-quality' input in ways that are both culturally specific and temporally fine-grained.

Exposure to IDS may not only facilitate infants' ability to process speech input in the moment, but also contribute to the development of neural connectivity that sets the stage for processing ADS later in life. While typically developing 7-month-old infants entrain to IDS with ease, they are not yet able to entrain to ADS (Kalashnikova et al., 2018a), and interactive auditory experience can sharpen infants' cortical tracking of auditory signals in the theta band (Musacchia et al., 2017). These findings point to a developmental change in infants' ability to process speech, presumably enabled by maturation of the neural structures and connections that allow for entrainment to speech and integration across its timescales. Even though nobody has studied how the amount of IDS input shapes entrainment, there is evidence that toddlers whose mothers provided more input are more efficient at processing speech in real-time (Hurtado et al., 2008; Weisleder & Fernald, 2013). Further, other aspects of the quality of early caregiver speech input (e.g., conversational turn-taking) have been shown to shape children's later neural responses to speech (Romeo et al., 2018). By synchronizing activity across neural oscillation frequencies that track different timescales (Leong et al., 2017), it is plausible that IDS scaffolds the requisite cross-region connections. Infants who are exposed to more and/or 'higher-quality' IDS may in turn be better equipped to process the different – and longer – timescales of ADS later in life. In addition, specific types of infant-directed communication may be especially beneficial; for instance, nursery rhymes found in most children's books provide high synchronization between syllable rate and prosodic stress (Leong et al., 2011). Differences in the amount of IDS (Hart & Risley, 1995; Schwab & Lew-Williams, 2016b) and shared book reading (Bradley et al., 2001) in different households therefore may have important implications not only for the child's early vocabulary growth (Farrant & Zubrick, 2012; Montag et al., 2015), but also for the neural connections responsible for optimal real-time processing of speech. This may even have a gradual snowballing effect on the child's ability to communicate, learn, and retain higher-level information from speech in school settings.

Co-creating input in caregiver-child interactions

In this review, we have examined how infants entrain to IDS. However, infants' input is highly dynamic and tailored to their changing needs (e.g., Schwab et al., 2018). We have so far overlooked the fact that infants experience IDS in the context of highly dynamic interactions, in which they actively co-create their input with their caregivers. This occurs over the long timescale of development but also during the short timescale of caregiver-child interactions.

Prior research has uncovered ways in which caregivers adapt their speech input to reflect young children's developing language skills over the first years of life (see Saint-Georges et al., 2013 for a review). For example, in Australian English and in Thai, there are changes in the average pitch, pitch range, and affect in caregiver speech during the first year (Kitamura et al., 2001; Kitamura & Burnham, 2003), and the most exaggerated modifications in these features peak around 6 months. Between 7 and 19 months, at least in Australian English, there are no notable acoustic changes in child-directed speech (Kalashnikova & Burnham, 2018). Between 18 and

24 months, Liu et al. (2009) and Han et al. (2018) showed a decrease in average pitch and pitch range of Mandarin-speaking caregivers, and Ramírez-Esparza et al. (2017) found a decrease in caregiver use of “parentese” between 11 and 24 months. Similar patterns were observed in a longitudinal study of Japanese children between 0 and 60 months (Amano et al., 2006). Over time, these adaptations may serve to create a signal that is fine-tuned to the child’s developing brain. For example, caregiver speech rate increases between the first and second year of life (Narayan & McDermott, 2016), which may parallel developmental changes in infants’ neural rhythms during a similar span (Sankupellay et al., 2011).

At shorter timescales, caregivers adapt their speech to infants from one moment to the next during single interactions. For example, caregivers shift their pitch contours depending on infants’ gaze being directed toward vs away from them (Stern et al., 1982) and respond differently to different kinds of infant vocalizations (Gros-Louis et al., 2006). Relatedly, caregivers show stronger real-time adaptations (including vowel hyper-articulation and prosodic adjustments) when they can see their child’s contingent responses (Braarud & Stormark, 2008; Lam and Kitamura, 2012; Snow, 1972).

What processes may allow caregivers to fine-tune their input to their infants’ needs, both in the moment and across longer periods of time? One possibility is that caregivers are attuned to their child’s increasingly sophisticated ability to communicate. This is supported by research showing that caregivers tend to repeat words more often when their child is young, but reduce word repetition over time to make room for more diverse words; the trajectory of this shift in word repetition is coupled to their child’s vocabulary (Schwab et al., 2018).

A second possibility is that caregivers repeat behaviors that elicit a positive (or desired) response from infants. Such reinforcing behavioral signals from the infant include smiles, babbles, gestures, and looks to the caregiver or to an object. Prior work shows that infants play an active role in shaping their input by vocalizing (Gros-Louis et al., 2016), gesturing (Lucca & Wilbourn, 2018; 2019), and directing attention (Smith et al., 2018). These behaviors may shape both short-term shifts in caregiver input as well as long-term adaptations related to maturation or perceived language skills. One clear example of short-term adaptation comes from a study by Smith and Trainor (2006), who showed that when signals of engagement from the infant aligned with moments of an increase in mothers’ pitch, mothers tended to use increasingly high pitch over time (Smith & Trainor, 2008). This work suggests that infant behaviors may work in tandem with caregiver behaviors to foster optimal attention and entrainment. There are also examples of long-term adaptations that are correlated with infant behavior. For example, in one study, caregivers’ use of different intonation contours over the first year of the infant’s life correlated with infants’ responsiveness to these contours at different ages. Specifically, caregivers used more falling contours with 3- and 5-month-old Japanese learning infants, and more rising contours with 7- and 9-month-old infants (Niwano & Sugai, 2002). This shift mirrored age-related changes in infants’ responsiveness: younger infants responded more to falling contours and older infants responded more to rising ones. Although it is difficult to know if caregivers or infants drove this change, it suggests bidirectional adaptation. That is, there is an iterative interplay between caregiver input and infant attention to various communicative cues (see Van Dijk et al., 2013). These studies highlight the active role of child feedback as a ‘reward’ for caregivers in the moment, which may accumulate over time to give rise to longer-term adaptations.

A third possible explanation for such adaptation lies in neural and behavioral synchrony, such that caregivers become aligned to the temporal dynamics of infants’ brains and behaviors. Recent research shows that during live interactions, neural activity becomes aligned between adults and young children (Nguyen et al., 2020; Nguyen et al., 2021; Piazza et al., 2020; Piazza et al., 2021). Both neural and behavioral synchrony could serve to support the co-creation of optimal input for entrainment. As an example of behavioral synchrony, it has been shown that caregivers implicitly match the pitch of infant vocalizations during live interactions (McRoberts & Best, 1997; Shimura & Yamanoucho, 1992), which may create input at an optimal frequency for infants’ perception. According to the logic of dynamic attention theory, by shifting their neural rhythms to match those of the infant, caregivers produce speech input that may elicit optimal entrainment from their infants. Although the exact mechanism for this alignment is currently unknown, synchronized neural activity may result in coordinated input and perception between the caregiver and the infant (Piazza et al., 2021; Wass et al., 2020).

Uncovering how infants and caregivers co-create input for optimal entrainment is an exciting direction for future research. A wealth of research has advanced what is known about infants and young children as active learners (e.g. Zettersten & Saffran, 2021), but the next phase of research on this topic could examine whether or how active learning helps to create *optimal* input that could in turn support neural entrainment and learning. What kinds of dyadic processes give rise to infant-directed input that is fine-tuned to the child’s neural and cognitive development, both in the moment and across time? Exploring this question will enable new insights into the origins of individual differences in young children’s learning trajectories.

Beyond speech: Understanding the many dimensions of early communication

In each section above, the focus was on how individual speech cues may impact infants’ readiness to process and learn from speech. That is, each section generally stayed within the domain of speech. However, in natural interaction between caregivers and infants, a breadth of multimodal cues come together to support communication and learning (Brand et al., 2002; Kosie & Lew-Williams, 2022; Suarez-Rivera et al., 2019; Yu & Smith, 2012). Eye gaze, touch, gesture, and facial expressions have been shown to shape children’s attention across moments in time and locations in space. For example, a caregiver may smile, touch a child’s cheek, and use exaggerated happy prosody to say “Is that a smile?”; this simultaneously draws attention to the speech signal at a key word, highlights emotional content, offers a physical cue about the location of the smile, and provides a visual of the caregiver smiling. At another moment, the caregiver may use gesture or touch to highlight the rhythm of the children’s song ‘Wheels on the Bus’, while holding a toy bus to draw attention to the target object. Last but not least, it is crucial to point out that learning occurs far beyond language. As one minor example, a caregiver may use their prosody to guide the child through the steps of learning how to tie a shoe or place a triangle in the triangle slot.

Research on attention across time, space, and communicative modality has been largely disconnected. Several prominent word learning theories have focused on how infants match word labels to their physical referents in the world (e.g., Trueswell et al., 2013; Yu & Smith, 2011), a process that may be facilitated by spatial attention to parts of the visual scene. Other theoretical frameworks, such as PRIMIR (Processing Rich Information in Multidimensional Interactive Representations; Werker & Curtin, 2005), have focused on speech processing, focusing specifically on how infants process different levels of information in the speech stream depending on their developmental level, perceptual biases, and task demands. This framework dovetails nicely with our entrainment-based approach given the mutual emphasis on speech as a deep, complex, and engaging source of information. However, developmental scientists have yet to wed the dynamics of attention and processing in children's environments in a way that encompasses visual, auditory, and tactile sensations that unfold over time. This will be a crucial step toward expanding our understanding of links between perception, attention, entrainment, and learning.

There is limited existing experimental evidence that multimodal cues work in tandem with the speech signal to direct attention and enhance entrainment at specific moments in time (Peelle & Sommers, 2015). But, to our knowledge, developmental scientists have not examined real-time attention dynamics during natural, multimodal communication. Ultimately, in order to understand how caregivers' multimodal cues work together to support learning in real-world interactions, researchers will need to merge findings from different fields of attention research. Traditional experimental and statistical designs that examine one or at most two features at a time are limited in elucidating how a multitude of interacting shapes attention. Large-scale multimodal observations of naturalistic caregiver-child interactions will be integral for mapping out the parameter space of caregiver cues in which the child's attention and learning are situated. Computational approaches, such as machine learning, can identify patterns of variation in such multivariate measures in a data-driven way and ultimately help us understand when different cues occur together or diverge. In addition to characterizing the multimodal learning environment, researchers will need to develop robust measures of child attention that reflect the complexity of processing information across time, space, and modalities. Prior work on child attention commonly uses the child's gaze direction and touch as a measure of their attention, with recent advances in head-mounted eye trackers that allow for an accurate first-person view (Slone et al., 2018). Neural measures such as EEG and fNIRS can supplement this technique to quantify children's attention across other modalities and provide finer-grained estimates of attention modulations across time.

Conclusion

This review integrated research on the temporal dynamics of IDS and the neural substrates of sustained attention to better understand theoretical claims about the role of IDS in children's learning. We first emphasized that the sub-second timescale of IDS is crucial for understanding how this dynamic signal shapes real-time communication. Second, we used intuitions from neural research on entrainment to explore which moments within IDS are optimal for processing new information. Third, we explored how auditory and visual attention during real-time communication interact to support infants' learning.

The key proposal is that infants do not simply prefer IDS over ADS, but rather they entrain particularly well to its moment-to-moment dynamics. Entrainment, in turn, enhances communication and learning both in the moment and across much longer timescales and may guide the development of neural structures that process speech. We can use these insights to redefine 'high-quality' speech not as a feature of speech itself, but as a dynamic interplay between behavior, attention, and the brain. With this redefinition, developmental scientists can gain traction in understanding the beginnings and high-stakes nature of young children's highly divergent learning trajectories.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by NIH R01HD095912 (C.L.W.), and ACM SIGHPC Computational & Data Science Fellowship (M.L.N.). We would like to thank Sabine Kastner and Xiaofang Yang for their input and comments on early drafts of this paper, as well as the reviewers of this manuscript for their input on later drafts. This work was supported by a grant from the ACM SIGHPC Computational & Data Science Fellowship to MLN and by the National Institutes of Health to CLW [R01HD095912].

References

Abrams, D. A., Nicol, T., & Zecker, S. (2009). Abnormal cortical processing of the syllable rate of speech in poor readers. *Journal of Neuroscience.* https://www.jneurosci.org/content/29/24/7686.short.

Amano, S., Nakatani, T., & Kondo, T. (2006). Fundamental frequency of infants' and parents' utterances in longitudinal recordings. *The Journal of the Acoustical Society of America, 119*(3), 1636–1647.

Aslin, R. N. (1993). Segmentation of Fluent Speech into Words: Learning Models and the Role of Maternal Input. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (pp. 305–315). Netherlands: Springer.

Assaneo, M. F., & Poeppel, D. (2018). The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Science Advances, 4*(2), eaao3842.

Astheimer, L. B., & Sanders, L. D. (2009). Listeners modulate temporally selective attention during natural speech processing. *Biological psychology, 80*(1), 23–34.

Braarud, H. C., & Stormark, K. M. (2008). Prosodic modification and vocal adjustments in mothers' speech during face-to-face interaction with their two-to four-month-old infants: A double video study. *Social Development, 17*(4), 1074–1084.

Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition, 94*(2), B45–B53.

Batterink, L. J., & Zhang, S. (2022). Simple statistical regularities presented during sleep are detected but not retained. *Neuropsychologia, 164*, Article 108106.

Bench, J. (1968). Sound transmission to the human foetus through the maternal abdominal wall. *The Journal of Genetic Psychology*, *113*(1st Half), 85–87.

Benítez-Burraco, A., & Murphy, E. (2016). The Oscillopathic Nature of Language Deficits in Autism: From Genes to Language Evolution. *Frontiers in Human Neuroscience, 10*, 120.

Bergelson, E., Bergmann, C., Byers-Heinlein, K., Cristia, A., Cusack, R., Dyck, K., Floccia, C., Frank, M. C., Gervain, J., Gonzalez, N., & al., E. (2017). *Quantifying sources of variability in infancy research using the infant-directed speech preference*. https://doi.org/10.31234/osf.io/s98ab.

Binns, C., & Culling, J. F. (2007). The role of fundamental frequency contours in the perception of speech against interfering speech. *The Journal of the Acoustical Society of America, 122*(3), 1765.

Bonacina, S., Krizman, J., White-Schwoch, T., & Kraus, N. (2018). Clapping in time parallels literacy and calls upon overlapping neural mechanisms in early readers. *Annals of the New York Academy of Sciences*. https://doi.org/10.1111/nyas.13704

Bradley, R. H., Corwyn, R. F., McAdoo, H. P., & Coll, C. G. (2001). The home environments of children in the United States part I: Variations by age, ethnicity, and poverty status. *Child Development, 72*(6), 1844–1867.

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for "motionese": Modifications in mothers' infant-directed action. *Developmental Science, 5*(1), 72–83.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81*(2), B33–B44.

Brock, J., Brown, C. C., Boucher, J., & Rippon, G. (2002). The temporal binding deficit hypothesis of autism. *Development and Psychopathology, 14*(2), 209–224.

Brodsky, P., & Waterfall, H. (2007). Characterizing motherese: On the computational structure of child-directed language. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *29*. https://cloudfront.escholarship.org/dist/prd/content/qt54k371nk/qt54k371nk.pdf.

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early Language Experience in a Tseltal Mayan Village. *Child Development, 91*(5), 1819–1835.

Chang, A., Bosnyak, D. J., & Trainor, L. J. (2016). Unpredicted Pitch Modulates Beta Oscillatory Power during Rhythmic Entrainment to a Tone Sequence. *Frontiers in Psychology, 7*, 327.

Chen, X., Yang, J., Gan, S., & Yang, Y. (2012). The contribution of sound intensity in vocal emotion perception: Behavioral and electrophysiological evidence. *PloS One, 7*(1), Article e30278.

Cirelli, L. K., Spinelli, C., Nozaradan, S., & Trainor, L. J. (2016). Measuring Neural Entrainment to Beat and Meter in Infants: Effects of Music Background. *Frontiers in Neuroscience, 10*, 229.

Colombo, J., Frick, J. E., Ryther, J. S., Coldren, J. T., & Mitchell, D. W. (1995). Infants' detection of analogs of" motherese" in noise. *Merrill-Palmer Quarterly, 41*, 104.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61*(5), 1584–1595.

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech, 20*(1), 1–10.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11*(1), 51–62.

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

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience, 19*(1), 158–164.

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences of the United States of America, 109*(29), 11854–11859.

Ding, N., & Simon, J. Z. (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 33*(13), 5728–5735.

Dinstein, I., Pierce, K., Eyler, L., Solso, S., Malach, R., Behrmann, M., et al. (2011). Disrupted neural synchronization in toddlers with autism. *Neuron, 70*(6), 1218–1225.

Drullman, R. (1995). Temporal envelope and fine structure cues for speech intelligibility. *The Journal of the Acoustical Society of America, 97*(1), 585–592.

Eaves, B. S., Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant-directed speech is consistent with teaching. *Psychological Review, 123*(6), 758–771.

Echols, C. H., Crowhurst, M. J., & Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. *Journal of memory and language, 36*(2), 202–225.

Farrant, B. M., & Zubrick, S. R. (2012). Early vocabulary development: The importance of joint attention and parent-child book reading. *First Language, 32*(3), 343–364.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior & Development, 8*(2), 181–195.

Fernald, A. (1989). Intonation and Communicative Intent in Mothers' Speech to Infants: Is the Melody the Message? *Child Development, 60*(6), 1497–1510.

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. https://ci.nii.ac.jp/naid/20000152521/.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior & Development, 10*(3), 279–293.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27*(2), 209–221.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20*(1), 104.

Fiebelkorn, I. C., & Kastner, S. (2019). A Rhythmic Theory of Attention. *Trends in Cognitive Sciences, 23*(2), 87–101.

Friedlander, B. Z. (1968). The effect of speaker identity, voice inflection, vocabulary, and message redundancy on infants' selection of vocal reinforcement. *Journal of Experimental Child Psychology, 6*(3), 443–459.

Futagi, Y., Ishihara, T., Tsuda, K., Suzuki, Y., & Goto, M. (1998). Theta rhythms associated with sucking, crying, gazing and handling in infants. *Electroencephalography and Clinical Neurophysiology, 106*(5), 392–399.

Ghazanfar, A. A., Morrill, R. J., & Kayser, C. (2013). Monkeys are perceptually tuned to facial expressions that exhibit a theta-like speech rhythm. *Proceedings of the National Academy of Sciences of the United States of America, 110*(5), 1959–1963.

Ghazanfar, A. A., & Poeppel, D. (2014). The neurophysiology and evolution of the speech rhythm. *The Cognitive Neurosciences., 5th Ed., 5*(2014), 629–638.

Ghazanfar, A. A., & Takahashi, D. Y. (2014). Facial expressions and the evolution of the speech rhythm. *Journal of Cognitive Neuroscience, 26*(6), 1196–1207.

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience, 15*(4), 511–517.

Gleitman, L. R., Newport, E. L., & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language, 11*(1), 43–79.

Goldstein, M. H., & Schwade, J. A. (2009). From Birds to Words. In M. S. Blumberg, J. H. Freeman, & S. R. Robinson (Eds.), *Oxford Handbook of Developmental Behavioral Neuroscience*. Oxford University Press.

Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., & Devin McAuley, J. (2015). Musical rhythm discrimination explains individual differences in grammar skills in children. *Developmental Science, 18*(4), 635–644.

Goswami, U. (2002). Phonology, reading development, and dyslexia: A cross-linguistic perspective. *Annals of Dyslexia, 52*(1), 139–163.

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences, 15*(1), 3–10.

Goswami, U. (2019). Speech rhythm and language acquisition: An amplitude modulation phase hierarchy perspective. *Annals of the New York Academy of Sciences, 1453*(1), 67–78.

Goswami, U., & Leong, V. (2013). Speech rhythm and temporal structure: Converging perspectives? *Laboratory Phonology, 4*(1), 67–92.

Goswami, U., Mead, N., Fosker, T., Huss, M., Barnes, L., & Leong, V. (2013). Impaired perception of syllable stress in children with dyslexia: A longitudinal study. *Journal of Memory and Language, 69*(1), 1–17.

Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy: The Official Journal of the International Society on Infant Studies, 18*(5), 797–824.

Grassmann, S., & Tomasello, M. (2007). Two-year-olds use primary sentence accent to learn new words. *Journal of Child Language, 34*(3), 677–687.

Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology, 24*(1), 14.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., et al. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology, 11*(12), Article e1001752.

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development, 30*(6), 509–516.

Gros-Louis, J., West, M. J., & King, A. P. (2016). The influence of interactive context on prelinguistic vocalizations and maternal responses. *Language Learning and Development, 12*(3), 280–294.

Háden, G. P., Honing, H., Török, M., & Winkler, I. (2015). Detecting the temporal structure of sound sequences in newborn infants. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology, 96*(1), 23–28.

Haegens, S., & Zion-Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience and Biobehavioral Reviews, 86*, 150–165.

Han, M., De Jong, N. H., & Kager, R. (2018). Lexical tones in Mandarin Chinese infant-directed speech: Age-related changes in the second year of life. *Frontiers in psychology, 9*, 434.

Hämäläinen, J., Leppänen, P. H. T., Torppa, M., Müller, K., & Lyytinen, H. (2005). Detection of sound rise time by adults with dyslexia. *Brain and Language, 94*(1), 32–42.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children. 268*. https://psycnet.apa.org/fulltext/1995-98021-000.pdf.

Hayashi, A., Tamekawa, Y., & Kiritani, S. (2001). Developmental change in auditory preferences for speech stimuli in Japanese infants.

Headley, D. B., & Paré, D. (2013). In sync: Gamma oscillations and emotional memory. *Frontiers in Behavioral Neuroscience, 7*, 170.

Helfrich, R. F., Fiebelkorn, I. C., Szczepanski, S. M., Lin, J. J., Parvizi, J., Knight, R. T., et al. (2018). Neural Mechanisms of Sustained Attention Are Rhythmic. *Neuron, 99*(4), 854–865.e5.

Hills, T. (2013). The company that words keep: Comparing the statistical structure of child- versus adult-directed language. *Journal of Child Language, 40*(3), 586–604.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., et al. (2015). The Contribution of Early Communication Quality to Low-Income Children's Language Success. *Psychological Science, 26*(7), 1071–1083.

Horton, C., Srinivasan, R., & D'Zmura, M. (2014). Envelope responses in single-trial EEG indicate attended speaker in a "cocktail party". *Journal of Neural Engineering, 11*(4), Article 046015.

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science, 11*(6), F31–F39.

Jin, P., Zou, J., Zhou, T., & Ding, N. (2018). Eye activity tracks task-relevant structures during speech and auditory sequence perception. *Nature communications, 9*(1), 1–15.

Jochaut, D., Lehongre, K., Saitovitch, A., Devauchelle, A.-D., Olasagasti, I., Chabane, N., et al. (2015). Atypical coordination of cortical oscillations in response to speech in autism. *Frontiers in Human Neuroscience, 9*, 171.

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review, 83*(5), 323–355.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96*(3), 459–491.

Jones, M. R., Kidd, G., & Wetzel, R. (1981). Evidence for rhythmic attention. *Journal of Experimental Psychology. Human Perception and Performance, 7*(5), 1059–1073.

Jusczyk, P. W. (1997). The discovery of spoken language.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*(3), 675–687.

Kakouros, S., & Räsänen, O. (2016). Perception of Sentence Stress in Speech Correlates With the Temporal Unpredictability of Prosodic Features. *Cognitive Science, 40*(7), 1739–1774.

Kakouros, S., Salminen, N., & Räsänen, O. (2018). Making predictable unpredictable with style–Behavioral and electrophysiological evidence for the critical role of prosodic expectations in the perception of prominence in speech. *Neuropsychologia, 109*, 181–199.

Kalashnikova, M., & Burnham, D. (2018). Infant-directed speech from seven to nineteen months has similar acoustic properties but different functions. *Journal of child language, 45*(5), 1035–1053.

Kalashnikova, M., Goswami, U., & Burnham, D. (2019). Sensitivity to amplitude envelope rise time in infancy and vocabulary development at 3 years: A significant relationship. *Developmental Science, 22*(6), Article e12836.

Kalashnikova, M., Peter, V., Di Liberto, G. M., & Lalor, E. C. (2018a). Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Scientific Reports*. https://www.nature.com/articles/s41598-018-32150-6.pdf?origin=ppub.

Kalashnikova, M., Goswami, U., & Burnham, D. (2018b). Mothers speak differently to infants at-risk for dyslexia. *Developmental science, 21*(1), Article e12487.

Kang, O., & Wheatley, T. (2017). Pupil dilation patterns spontaneously synchronize across individuals during shared attention. *Journal of Experimental Psychology. General, 146*(4), 569–576.

Kaplan, P. S., Goldstein, M. H., Huckeby, E. R., Owren, M. J., & Cooper, R. P. (1995). Dishabituation of visual attention by infant- versus adult-directed speech: Effects of frequency modulation and spectral composition. *Infant Behavior & Development, 18*(2), 209–223.

Kempe, V., Brooks, P. J., & Gillis, S. (2005). Diminutives in child-directed speech supplement metric with distributional word segmentation cues. *Psychonomic Bulletin & Review, 12*(1), 145–151.

Kisilevsky, S., Hains, S. M. J., Jacquet, A. Y., Granier-Deferre, C., & Lecanuet, J. P. (2004). Maturation of fetal responses to music. *Developmental Science, 7*(5), 550–559.

Kitamura, C., & Burnham, D. (1998). The infant's response to maternal vocal affect. *Advances in Infancy Research, 12*, 221–236.

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2001). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant behavior and development, 24*(4), 372–392.

Kitamura, C., & Burnham, D. (2003). Pitch and Communicative Intent in Mother's Speech: Adjustments for Age and Sex in the First Year. *Infancy: The Official Journal of the International Society on Infant Studies, 4*(1), 85–110.

Kitamura, C., & Lam, C. (2009). Age-Specific Preferences for Infant-Directed Affective Intent. *Infancy: The Official Journal of the International Society on Infant Studies, 14*(1), 77–100.

Kosie, J.E., & Lew-Williams, C. (2022). *Infant-Directed Communication: Examining the many overlapping dimensions of communication beyond speech.* Talk presented at the 23rd Biennial International Congress of Infant Studies, Ottawa, Ontario, Canada.

Kuhl, P. K. (2007). Is speech learning "gated"by the social brain? *Developmental Science, 10*(1), 110–120.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., et al. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*(5326), 684–686.

Kumagai, Y., Arvaneh, M., & Tanaka, T. (2017). Familiarity Affects Entrainment of EEG in Music Listening. *Frontiers in Human Neuroscience, 11*, 384.

Lam, C., & Kitamura, C. (2012). Mommy, speak clearly: Induced hearing loss shapes vowel hyperarticulation. *Developmental Science, 15*(2), 212–221.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*(1), 119–159.

Laures, J. S., & Bunton, K. (2003). Perceptual effects of a flattened fundamental frequency at the sentence level under different listening conditions. *Journal of Communication Disorders, 36*(6), 449–464.

Leong, V., & Goswami, U. (2014). Assessment of rhythmic entrainment at multiple timescales in dyslexia: Evidence for disruption to syllable timing. *Hearing Research, 308*, 141–161.

Leong, V., & Goswami, U. (2015). Acoustic-Emergent Phonology in the Amplitude Envelope of Child-Directed Speech. *PloS One, 10*(12), Article e0144411.

Leong, V., Kalashnikova, M., Burnham, D., & Goswami, U. (2017). The Temporal Modulation Structure of Infant-Directed Speech. *Open Mind, 1*(2), 78–90.

Leong, V., Turner, R., Stone, M., & Goswani, U. (2011). Spoken Nursery Rhymes Have a Fractal Rhythmic Structure-Evidence from Patterns of Slow Amplitude Modulation (AM). *Proceedings of the Annual Meeting of the Cognitive Science Society, 33*. https://cloudfront.escholarship.org/dist/prd/content/qt3cv587w6/qt3cv587w6.pdf.

Lew-Williams, C., & Saffran, J. R. (2012). All words are not created equal: Expectations about word length guide infant statistical learning. *Cognition, 122*(2), 241–246.

Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6*(3), F1–F10.

Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2009). Age-related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: A longitudinal study. *Journal of Child Language, 36*(4), 909–922.

Li, X.-Q., & Ren, G.-Q. (2012). How and when accentuation influences temporally selective attention and subsequent semantic processing during on-line spoken language comprehension: An ERP study. *Neuropsychologia, 50*(8), 1882–1894.

Lucca, K., & Wilbourn, M. P. (2018). Communicating to learn: Infants' pointing gestures result in optimal learning. *Child development, 89*(3), 941–960.

Lucca, K., & Wilbourn, M. P. (2019). The what and the how: Information-seeking pointing gestures facilitate learning labels and functions. *Journal of experimental child psychology, 178*, 417–436.

ManyBabies Consortium. (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science, 3*(1), 24–52.

Marshall, P. J., Bar-Haim, Y., & Fox, N. A. (2002). Development of the EEG from 5 months to 4 years of age. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 113*(8), 1199–1208.

Martin, A., Igarashi, Y., Jincho, N., & Mazuka, R. (2016). Utterances in infant-directed speech are shorter, not slower. *Cognition, 156*, 52–59.

Makov, S., Sharon, O., Ding, N., Ben-Shachar, M., Nir, Y., & Golumbic, E. Z. (2017). Sleep disrupts high-level speech parsing despite significant basic auditory processing. *Journal of Neuroscience, 37*(32), 7772–7781.

Maulsby, R. L. (1971). An illustration of emotionally evoked theta rhythm in infancy: Hedonic hypersynchrony. *Electroencephalography and Clinical Neurophysiology, 31*(2), 157–165.

Ma, W., Golinkoff, R. M., Houston, D., & Hirsh-Pasek, K. (2011). Word Learning in Infant- and Adult-Directed Speech. *Language Learning and Development: The Official Journal of the Society for Language Development, 7*(3), 185–201.

McMillan, B. T. M., & Saffran, J. R. (2016). Learning in Complex Environments: The Effects of Background Speech on Early Word Learning. *Child Development, 87*(6), 1841–1855.

McMurray, B. (2016). Language at Three Timescales: The Role of Real-Time Processes in Language Development and Evolution. *Topics in Cognitive Science, 8*(2), 393–407.

McRoberts, G. W., & Best, C. T. (1997). Accommodation in mean f0 during mother–infant and father–infant vocal interactions: A longitudinal case study. *Journal of Child Language, 24*(3), 719–736.

McRoberts, G. W., McDonough, C., & Lakusta, L. (2009). The role of verbal repetition in the development of infant speech preferences from 4 to 14 months of age. *Infancy: The Official Journal of the International Society on Infant Studies, 14*(2), 162–194.

Meltzer, B., Reichenbach, C. S., Braiman, C., Schiff, N. D., Hudspeth, A. J., & Reichenbach, T. (2015). The steady-state response of the cerebral cortex to the beat of music reflects both the comprehension of music and attention. *Frontiers in Human Neuroscience, 9*, 436.

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature, 485*(7397), 233–236.

Messer, D. J. (1981). The identification of names in maternal speech to infants. *Journal of Psycholinguistic Research, 10*(1), 69–77.

Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The Words Children Hear: Picture Books and the Statistics for Language Learning. *Psychological Science, 26*(9), 1489–1496.

Mu, Y., Fan, Y., Mao, L., & Han, S. (2008). Event-related theta and alpha oscillations mediate empathy for pain. *Brain Research, 1234*, 128–136.

Musacchia, G., Ortiz-Mantilla, S., Choudhury, N., Realpe-Bonilla, T., Roesler, C., & Benasich, A. A. (2017). Active auditory experience in infancy promotes brain plasticity in theta and gamma oscillations. *Developmental cognitive neuroscience, 26*, 9–19.

Naoi, N., Minagawa-Kawai, Y., Kobayashi, A., Takeuchi, K., Nakamura, K., Yamamoto, J.-I., et al. (2012). Cerebral responses to infant-directed speech and the effect of talker familiarity. *NeuroImage, 59*(2), 1735–1744.

Narayan, C. R., & McDermott, L. C. (2016). Speech rate and pitch characteristics of infant-directed speech: Longitudinal and cross-linguistic observations. *The Journal of the Acoustical Society of America, 139*(3), 1272–1281.

Nelson, D. G. K., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language, 16*(1), 55–68.

Nencheva, M. L., Piazza, E. A., & Lew-Williams, C. (2020). The moment-to-moment pitch dynamics of child-directed speech shape toddlers' attention and learning. *Developmental Science, e12997*.

Newman, R. S. (2003). Prosodic differences in mothers' speech to toddlers in quiet and noisy environments. *Applied Psycholinguistics, 24*(4), 539–560.

Nguyen, T., Schleihauf, H., Kayhan, E., Matthes, D., Vrtička, P., & Hoehl, S. (2020). The effects of interaction quality on neural synchrony during mother-child problem solving. *cortex*, *124*, 235-249.

Nguyen, T., Schleihauf, H., Kayhan, E., Matthes, D., Vrtička, P., & Hoehl, S. (2021). Neural synchrony in mother–child conversation: Exploring the role of conversation patterns. *Social Cognitive and Affective Neuroscience, 16*(1–2), 93–102.

Niwano, K., & Sugai, K. (2002). Intonation contour of Japanese maternal infant-directed speech and infant vocal response. *The Japanese Journal of Special Education, 39*(6), 59–68.

Obleser, J., & Kayser, C. (2019). Neural Entrainment and Attentional Selection in the Listening Brain. *Trends in Cognitive Sciences, 23*(11), 913–926.

Orekhova, E. V., Stroganova, T. A., Posikera, I. N., & Elam, M. (2006). EEG theta rhythm in infants and preschool children. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 117*(5), 1047–1062.

Otto-Meyer, S., Krizman, J., White-Schwoch, T., & Kraus, N. (2018). Children with autism spectrum disorder have unstable neural responses to sound. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale, 236*(3), 733–743.

Outters, V., Schreiner, M. S., Behne, T., & Mani, N. (2020). Maternal input and infants' response to infant-directed speech. *Infancy, 25*(4), 478–499.

Palana, J., Schwartz, S., & Tager-Flusberg, H. (2021). Evaluating the use of cortical entrainment to measure atypical speech processing: A systematic review. *Neuroscience & Biobehavioral Reviews*.

Pallesen, K. J., Bailey, C. J., Brattico, E., Gjedde, A., Palva, J. M., & Palva, S. (2015). Experience Drives Synchronization: The phase and Amplitude Dynamics of Neural Oscillations to Musical Chords Are Differentially Modulated by Musical Expertise. *PloS One, 10*(8), Article e0134211.

Palva, J. M., Palva, S., & Kaila, K. (2005). Phase synchrony among neuronal oscillations in the human cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 25*(15), 3962–3972.

Palva, S., & Palva, J. M. (2018). Roles of Brain Criticality and Multiscale Oscillations in Temporal Predictions for Sensorimotor Processing. *Trends in Neurosciences, 41*(10), 729–743.

Panneton, R., Kitamura, C., Mattock, K., & Burnham, D. (2006). Slow Speech Enhances Younger But Not Older Infants' Perception of Vocal Emotion. *Research in Human Development, 3*(1), 7–19.

Papoušek, M., Bornstein, M. H., Nuzzo, C., Papoušek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior & Development, 13*(4), 539–545.

Payne, E. M., Post, B., Astruc, L., Prieto, P., & Vanrell, M. (2015). *Rhythmic modification in child directed speech.* https://ora.ox.ac.uk/objects/uuid:0248005f-c604-49ce-9a07-33b1fb9fd8f2.

Peelle, J. E., & Davis, M. H. (2012). Neural Oscillations Carry Speech Rhythm through to Comprehension. *Frontiers in Psychology, 3*, 320.

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex, 23*(6), 1378–1387.

Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 68*, 169–181.

Pegg, J. E., Werker, J. F., & McLeod, P. J. (1992). Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior & Development, 15*(3), 325–345.

Phillips, J. R. (1973). Syntax and Vocabulary of Mothers' Speech to Young Children: Age and Sex Comparisons. *Child Development, 44*(1), 182–185.

Piazza, E. A., Cohen, A., Trach, J., & Lew-Williams, C. (2021). Neural synchrony predicts children's learning of novel words. *Cognition, 214*, Article 104752.

Piazza, E. A., Hasenfratz, L., Hasson, U., & Lew-Williams, C. (2020). Infant and adult brains are coupled to the dynamics of natural communication. *Psychological Science, 31*(1), 6–17.

Piazza, E. A., Iordan, M. C., & Lew-Williams, C. (2017). Mothers Consistently Alter Their Unique Vocal Fingerprints When Communicating with Infants. *Current Biology: CB, 27*(20), 3162–3167.e3.

Piazza, E. A., Nencheva, M. L., & Lew-Williams, C. (2021). The Development of Communication Across Timescales. *Current Directions in Psychological Science, 30*(6), 459–467.

Porcaro, C., Zappasodi, F., Barbati, G., Salustri, C., Pizzella, V., Rossini, P. M., et al. (2006). Fetal auditory responses to external sounds and mother's heart beat: Detection improved by Independent Component Analysis. *Brain Research, 1101*(1), 51–58.

Posikera, I. N., Stroganova, T. A., Zhurba, L. T., & Timonina, O. V. (1986). Electroencephalographic correlates of positive emotional reactions in children in the first year of life. *Zhurnal Nevropatologii I Psikhiatrii Imeni S.S Korsakova, 86*(10), 1485–1491.

Posner, M. I., Snyder, C. R., & Solso, R. (2004). Attention and cognitive control. *Cognitive Psychology: Key Readings, 205*. https://books.google.com/books?hl=en&lr=&id=DcJAqyCK6T8C&oi=fnd&pg=PA205&dq=posner+2004&ots=q4zh05gfIN&sig=Ppx24RC_NTrhDgYhDimreMHwlvw.

Power, A. J., Colling, L. J., Mead, N., Barnes, L., & Goswami, U. (2016). Neural encoding of the speech envelope by children with developmental dyslexia. *Brain and Language, 160*, 1–10.

Purhonen, M., Kilpeläinen-Lees, R., Valkonen-Korhonen, M., Karhu, J., & Lehtonen, J. (2004). Cerebral processing of mother's voice compared to unfamiliar voice in 4-month-old infants. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology, 52*(3), 257–266.

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2017). Look who's talking NOW! Parentese speech, social context, and language development across time. *Frontiers in psychology, 8*, 1008.

Räsänen, O., Kakouros, S., & Soderstrom, M. (2018). Is infant-directed speech interesting because it is surprising?–Linking properties of IDS to statistical learning and attention at the prosodic level. *Cognition, 178*, 193–206.

Reetzke, R., Gnanateja, G. N., & Chandrasekaran, B. (2021). Neural tracking of the speech envelope is differentially modulated by attention and language experience. *Brain and Language, 213*, Article 104891.

Riecke, L., Formisano, E., Sorger, B., Başkent, D., & Gaudrain, E. (2018). Neural Entrainment to Speech Modulates Speech Intelligibility. *Current Biology: CB, 28*(2), 161–169.e5.

Ríos López, P. (2018). Investigation of the development of neural and behavioural auditory rhythmic sensitivity and of its contribution to reading acquisition.

Ríos-López, P., Molinaro, N., Bourguignon, M., & Lallier, M. (2022). Right-hemisphere coherence to speech at pre-reading stages predicts reading performance one year later. *Journal of Cognitive Psychology, 34*(2), 179–193.

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., et al. (2018). Beyond the 30-Million-Word Gap: Children's Conversational Exposure Is Associated With Language-Related Brain Function. *Psychological Science, 29*(5), 700–710.

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language, 35*(1), 185–205.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development, 83*(5), 1762–1774.

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., et al. (2013). Motherese in Interaction: At the Cross-Road of Emotion and Cognition? (A Systematic Review). *PLOS ONE, 8*(10), Article e78103. https://doi.org/10.1371/journal.pone.0078103

Saito, Y., Aoyama, S., Kondo, T., Fukumoto, R., Konishi, N., Nakamura, K., et al. (2007). Frontal cerebral blood flow change associated with infant-directed speech. *Archives of Disease in Childhood. Fetal and Neonatal Edition, 92*(2), F113–F116.

Sankupellay, M., Wilson, S., Heussler, H. S., Parsley, C., Yuill, M., & Dakin, C. (2011). Characteristics of sleep EEG power spectra in healthy infants in the first two years of life. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 122*(2), 236–243.

Santesso, D. L., Schmidt, L. A., & Trainor, L. J. (2007). Frontal brain electrical activity (EEG) and heart rate in response to affective infant-directed (ID) speech in 9-month-old infants. *Brain and Cognition, 65*(1), 14–21.

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15*(2), 123–148.

Schneider, B. A., & Trehub, S. E. (1992). *Sources of developmental change in auditory sensitivity*. https://psycnet.apa.org/record/1992-98622-001.

Schwab, J. F., & Lew-Williams, C. (2016a). Repetition across successive sentences facilitates young children's word learning. *Developmental Psychology*. https://psycnet.apa.org/journals/dev/52/6/879.html?uid=2016-22438-001.

Schwab, J. F., & Lew-Williams, C. (2016b). Language learning, socioeconomic status, and child-directed speech. *Wiley Interdisciplinary Reviews. Cognitive Science, 7*(4), 264–275.

Schwab, J. F., Rowe, M. L., Cabrera, N., & Lew-Williams, C. (2018). Fathers' repetition of words is coupled with children's vocabularies. In. *Journal of Experimental Child Psychology, 166*, 437–450. https://doi.org/10.1016/j.jecp.2017.09.012

Segal, J., & Newman, R. S. (2015). Infant preferences for structural and prosodic properties of infant-directed speech in the second year of life. *Infancy, 20*(3), 339–351.

Shepherd, S. V., Lanzilotto, M., & Ghazanfar, A. A. (2012). Facial muscle coordination in monkeys during rhythmic facial expressions and ingestive movements. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 32*(18), 6105–6116.

Shimura, Y., & Yamanoucho, I. (1992). Sound spectrographic studies on the relation between motherese and pleasure vocalization in early infancy. *Pediatrics International, 34*(3), 259–266.

Simon, D. M., & Wallace, M. T. (2016). Dysfunction of sensory oscillations in Autism Spectrum Disorder. *Neuroscience and Biobehavioral Reviews, 68*, 848–861.

Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., et al. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nature Communications, 7*, 12141.

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition, 106*(2), 833–870.

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' Listening Preferences: Baby Talk or Happy Talk? *Infancy: The Official Journal of the International Society on Infant Studies, 3*(3), 365–394.

Slone, L. K., Abney, D. H., Borjon, J. I., Chen, C.-H., Franchak, J. M., Pearcy, D., et al. (2018). Gaze in Action: Head-mounted Eye Tracking of Children's Dynamic Visual Attention During Naturalistic Behavior. *Journal of Visualized Experiments: JoVE, 141*. https://doi.org/10.3791/58496

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in cognitive sciences, 22*(4), 325–336.

Smith, N. A., & Trainor, L. J. (2008). Infant-Directed Speech Is Modulated by Infant Feedback. *Infancy: The Official Journal of the International Society on Infant Studies, 13*(4), 410–420.

Snow, C. E. (1972). Mothers' speech to children learning language. *Child development*, 549–565.

Soltész, F., Szűcs, D., Leong, V., White, S., & Goswami, U. (2013). Differential entrainment of neuroelectric delta oscillations in developmental dyslexia. *PloS One, 8*(10), Article e76608.

Song, J., & Iverson, P. (2018). Listening effort during speech perception enhances auditory and lexical processing for non-native listeners and accents. *Cognition, 179*, 163–170.

Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *The Journal of the Acoustical Society of America, 128*(1), 389–400.

Spinelli, M., & Mesman, J. (2018). The regulation of infant negative emotions: The role of maternal sensitivity and infant-directed speech prosody. *Infancy: The Official Journal of the International Society on Infant Studies, 23*(4), 502–518.

Stern, D. N., Spieker, S., & MacKain, K. (1982). Intonation contours as signals in maternal speech to prelinguistic infants. *Developmental Psychology, 18*(5), 727.

Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental Psychology, 55*(1), 96–109.

Symons, A. E., El-Deredy, W., Schwartze, M., & Kotz, S. A. (2016). The Functional Role of Neural Oscillations in Non-Verbal Emotional Communication. *Frontiers in Human Neuroscience, 10*, 239.

Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., et al. (2009). Sensitivity of newborn auditory cortex to the temporal structure of sounds. *Journal of Neuroscience, 29*(47), 14726–14733.

Telkemeyer, S., Rossi, S., Nierhaus, T., Steinbrink, J., Obrig, H., & Wartenburger, I. (2011). Acoustic processing of temporally modulated sounds in infants: Evidence from a combined near-infrared spectroscopy and EEG study. *Frontiers in psychology, 2*, 62.

Ten Oever, S., Schroeder, C. E., Poeppel, D., van Atteveldt, N., Mehta, A. D., Mégevand, P., et al. (2017). Low-Frequency Cortical Oscillations Entrain to Subthreshold Rhythmic Auditory Stimuli. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 37*(19), 4903–4912.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy: The Official Journal of the International Society on Infant Studies, 7*(1), 53–71.

Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development: The Official Journal of the Society for Language Development, 3*(1), 73–100.

Thomson, J. M., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: Auditory and motor rhythms link to reading and spelling. *Journal of Physiology, Paris, 102*(1–3), 120–129.

Tierney, A., & Kraus, N. (2014). Auditory-motor entrainment and phonological skills: Precise auditory timing hypothesis (PATH). *Frontiers in Human Neuroscience, 8*, 949.

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science, 11*(3), 188–195.

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review, 9*(2), 335–340.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive psychology, 66*(1), 126–156.

Uhlhaas, P. J., Roux, F., Rodriguez, E., Rotarska-Jagiela, A., & Singer, W. (2010). Neural synchrony and the development of cortical networks. *Trends in Cognitive Sciences, 14*(2), 72–80.

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences, 104*(33), 13273–13278.

Van Dijk, M., van Geert, P., Korecky-Kröll, K., Maillochon, I., Laaha, S., Dressler, W. U., et al. (2013). Dynamic adaptation in child–adult language interaction. *Language Learning, 63*(2), 243–270.

Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z., & Francart, T. (2018). Speech Intelligibility Predicted from Neural Entrainment of the Speech Envelope. *Journal of the Association for Research in Otolaryngology: JARO, 19*(2), 181–191.

Vosoughi, S., Roy, D. K. (2012). A longitudinal study of prosodic exaggeration in child-directed speech. *Speech Prosody, 6th International Conference. Speech Prosody Special Interest Group (SProSIG)*.

Wang, L., Bastiaansen, M., Yang, Y., & Hagoort, P. (2011). The influence of information structure on the depth of semantic processing: How focus and pitch accent determine the size of the N400 effect. *Neuropsychologia, 49*(5), 813–820.

Wass, S. V., Whitehorn, M., Haresign, I. M., Phillips, E., & Leong, V. (2020). Interpersonal neural entrainment during early social interaction. *Trends in cognitive sciences, 24*(4), 329–342.

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

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language learning and development, 1*(2), 197–234.

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology, 43*(2), 230–246.

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition, 103*(1), 147–162.

Whitford, T. J., Rennie, C. J., Grieve, S. M., Clark, C. R., Gordon, E., & Williams, L. M. (2007). Brain maturation in adolescence: Concurrent changes in neuroanatomy and neurophysiology. *Human Brain Mapping, 28*(3), 228–237.

Wilsch, A., Neuling, T., Obleser, J., & Herrmann, C. S. (2018). Transcranial alternating current stimulation with speech envelopes modulates speech comprehension. *NeuroImage, 172*, 766–774.

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences of the United States of America, 106*(7), 2468–2471.

Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental science, 14*(2), 165–180.

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition, 125*(2), 244–262.

Zettersten, M., & Saffran, J. R. (2021). Sampling to learn words: Adults and children sample words that reduce referential ambiguity. *Developmental science, 24*(3), Article e13064.

Zhang, Y., Koerner, T., Miller, S., Grice-Patil, Z., Svec, A., Akbari, D., et al. (2011). Neural coding of formant-exaggerated speech in the infant brain. *Developmental Science, 14*(3), 566–581.

Zoefel, B., Archer-Boyd, A., & Davis, M. H. (2018). Phase Entrainment of Brain Oscillations Causally Modulates Neural Responses to Intelligible Speech. *Current Biology: CB, 28*(3), 401–408.e5.

Zou, J., Feng, J., Xu, T., Jin, P., Luo, C., Zhang, J., et al. (2019). Auditory and language contributions to neural encoding of speech features in noisy environments. *NeuroImage, 192*, 66–75.