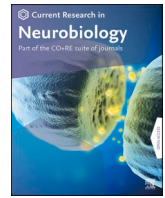


Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Current Research in Neurobiology

journal homepage: www.sciencedirect.com/journal/current-research-in-neurobiology

Unveiling the development of human voice perception: Neurobiological mechanisms and pathophysiology

Emily E. Harford^a, Lori L. Holt^b, Taylor J. Abel^{a,c,*}^a Department of Neurological Surgery, University of Pittsburgh, USA^b Department of Psychology, The University of Texas at Austin, USA^c Department of Bioengineering, University of Pittsburgh, USA

ARTICLE INFO

Keywords:

Auditory processing
Auditory cortex
Auditory neuroscience
Perception
Neurodevelopment

ABSTRACT

The human voice is a critical stimulus for the auditory system that promotes social connection, informs the listener about identity and emotion, and acts as the carrier for spoken language. Research on voice processing in adults has informed our understanding of the unique status of the human voice in the mature auditory cortex and provided potential explanations for mechanisms that underly voice selectivity and identity processing. There is evidence that voice perception undergoes developmental change starting in infancy and extending through early adolescence. While even young infants recognize the voice of their mother, there is an apparent protracted course of development to reach adult-like selectivity for human voice over other sound categories and recognition of other talkers by voice. Gaps in the literature do not allow for an exact mapping of this trajectory or an adequate description of how voice processing and its neural underpinnings abilities evolve. This review provides a comprehensive account of developmental voice processing research published to date and discusses how this evidence fits with and contributes to current theoretical models proposed in the adult literature. We discuss how factors such as cognitive development, neural plasticity, perceptual narrowing, and language acquisition may contribute to the development of voice processing and its investigation in children. We also review evidence of voice processing abilities in premature birth, autism spectrum disorder, and phonagnosia to examine where and how deviations from the typical trajectory of development may manifest.

1. Introduction

Conspecific vocalization constitutes one of the most important signals encountered by the auditory system. Sounds produced by the voice of a member of the same species hold a special status in regions of the auditory cortex, eliciting stronger neural responses compared to any other sound category in both humans (Belin et al., 2000) and other non-human animal species (Petkov et al., 2009). Vocalization constitutes a rich and complex auditory signal, composed of multiple frequency components and, importantly, providing a wealth of information to the listener about the source of a sound. While calls and vocalizations are used by a variety of species to transmit information (Gil-da-Costa et al., 2004; Seyfarth and Cheney, 2003), the human voice is particularly flexible and capable of communicating highly nuanced messages. A newborn's cry, a mother quietly soothing her child, a friend's laughter, a scream in the distance; though each is produced by the human voice, they convey distinct emotions, identities, and information that can

direct behavior and promote social connections (Gros-Louis et al., 2014; Trehub, 2017). Listeners may even use properties carried by the voice to make judgements about physical characteristics, trustworthiness, and personality of another person (Belin et al., 2017; McAleer et al., 2014; Schweinberger et al., 2014).

Despite the importance of vocalization in bonding, communication, and social perception, a timeline describing how voice perception evolves from fetal stages of development to maturity in adulthood is not fully understood. Research spanning decades has demonstrated that newborns have the ability to recognize the identity of at least some voices, namely the voice of their mother (DeCasper and Fifer, 1980), but that the processing of higher-level identity-related information does not reach maturity until late childhood or early adolescence (Mann et al., 1979). A developmental trajectory in voice processing clearly exists and will be outlined in this review. However, we simultaneously highlight the fact that there are substantial gaps in the literature base that as yet preclude precise mapping of this development and its neural

* Corresponding author. 4401 Penn Avenue, Pittsburgh, PA, USA.

E-mail address: abeltj@upmc.edu (T.J. Abel).

<https://doi.org/10.1016/j.crneur.2024.100127>

Received 6 October 2023; Received in revised form 22 February 2024; Accepted 26 February 2024

Available online 8 March 2024

2665-945X/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

underpinnings.

In this review we synthesize the existing literature on voice processing, specifically sensitivity and selectivity to human voice and voice identity perception, to provide a cohesive account of observations of developmental change from the prenatal period to early adolescence. Further, we make connections between the current developmental evidence and theoretical models proposed in adult studies to make predictions about the mechanisms underlying voice selectivity and identity perception and how these mechanisms may evolve with development. We additionally discuss factors that may modulate or contribute to the development of voice processing including neural plasticity, cerebral specialization, language acquisition, maturation of cognition and auditory processing, and other developmental processes like experience and perceptual narrowing. Finally, we consider how evidence of voice processing abilities in several pathophysiological states may further inform our understanding of typical development and underlying mechanisms. The goal of this review is to serve as a comprehensive reference of existing research, identify gaps in the literature, and provide specific suggestions that will guide future investigations.

2. Development of voice perception and processing

Processing specifically of human voice cannot be considered outside the context of the overall functioning of the auditory system, which undergoes a prolonged period of development relative to other sensory systems (Moore and Linthicum, 2007). The human auditory system begins to conduct signals to the cortex around the beginning of the third trimester (Goldberg et al., 2020; Moore and Linthicum, 2007) and temporal lobe responses to both internal and external auditory stimulation have been noted via functional magnetic resonance imaging (fMRI) in fetuses at 33–39 weeks gestational age (GA) (Goldberg et al., 2020; Hykin et al., 1999; Jardri et al., 2008). Therefore, the human auditory system is afforded very early experience with acoustic stimuli including human voice. Furthermore, evidence gathered via heartrate recordings suggests that fetuses can detect novelty or change within an auditory stream (Kisilevsky et al., 2009; J. P. Lecanuet et al., 1987) and recognize familiar spoken rhymes (DeCasper et al., 1994), indicating the emergence of auditory processing that supports learning. Even so, auditory cortex undergoes prolonged synaptogenesis, dendritic development, and myelination, reaching cytoarchitectural maturity by 1- to 2-years (Eggermont and Moore, 2012; Huttenlocher and Dabholkar, 1997) but continuing axonal maturation through late childhood (Moore and Guan, 2001); in fact, the superior temporal cortex has been identified as the last fully developed gray matter area, with the posterior portion of the superior temporal gyri maturing the latest (Gogtay et al., 2004). This suggests that higher-order auditory processes may also require a protracted period to reach full maturation.

In the following sections we summarize available evidence supporting the existence of a developmental trajectory for the perception of human voice. Consistent with overall development of the auditory cortex, the development of voice processing appears to begin in infancy and continues well into childhood. We focus first on evidence of voice selectivity, a response that gives primacy to the human voice in the auditory cortex and subserves early observed preferences for human voice over other auditory signals. Subsequently, we discuss voice identity processing, which represents a more complex auditory process and appears to undergo an extended period of maturation.

2.1. Voice sensitivity and selectivity

A foundational principle of voice perception as a unique auditory process is the observation that human voice holds a putatively special status in the auditory cortex. Voice selectivity describes the observation that certain auditory cortical regions respond more to human voice (or in animals, conspecific calls) compared to other sound categories. These regions are commonly referred to as temporal voice areas (TVAs) (Belin

et al., 2000; Pernet et al., 2015). Studies completed with adults over the past few decades have replicated the results first observed by Belin et al. (2000) localizing TVAs bilaterally in the upper bank of the superior temporal sulcus (STS) and the anterior superior temporal gyrus (STG) (Agus et al., 2017; Belin et al., 2000, 2002; Bodin et al., 2018; Fecteau et al., 2004; Pernet et al., 2015).

Of primary interest to this topic from a developmental perspective is whether selectivity for human voice is evident at birth or if it gradually emerges at a later point in development. To date, there are limited studies establishing the existence and possible localization of voice-selective cortical responses early in development. While research has provided convincing evidence that infants (Grossmann et al., 2010; Vouloumanos and Werker, 2004), neonates (Cheng et al., 2012; Simon et al., 2009), and even fetuses (Granier-Deferre et al., 2011; Jardri et al., 2012) respond differently to human voice than to other classes of auditory stimuli, it is difficult to ascertain whether these findings represent true voice selectivity as it is observed in adults or merely a preference for voice. Studies using both functional magnetic resonance imaging (fMRI), functional near-infrared spectroscopy (fNIRS), and scalp electroencephalography (EEG) have broadly identified the superior temporal cortex as a region that responds more strongly to voice than other acoustic categories in fetuses (Jardri et al., 2012), neonates (Simon et al., 2009), and infants 3-7 months-old (Blasi et al., 2011; Calce et al., 2023; Lloyd-Fox et al., 2012). Importantly, however, the nonvoice stimuli used in these studies span a wide range of sound classes including pure tones (Jardri et al., 2012), noise (Simon et al., 2009), and environmental sounds with (Calce et al., 2023; Lloyd-Fox et al., 2012) and without (Blasi et al., 2011) animal vocalizations, making comparisons to the adult literature difficult to reconcile. Two studies with 4–7-month-old infants have investigated voice selectivity using fNIRS along with the standard voice localizer paradigm employed in the adult literature (Belin et al., 2000) but have yielded inconsistent results. While some studies have found evidence of voice selectivity across infants in this age range (Calce et al., 2023; Lloyd-Fox et al., 2012), Grossmann et al. (2010) found that infants at 7- but not 4-month-olds exhibited voice-selective responses. Despite a lack of consensus surrounding definitive voice selectivity in the developing auditory cortex, studies using behavioral paradigms have contributed additional evidence in favor of, at minimum, an emerging preference for human voice over music in fetuses (Granier-Deferre et al., 2011) and over synthetic sounds in infants as young as a few days old (Vouloumanos et al., 2010; Vouloumanos and Werker, 2004). Moreover, the apparent early specialization in temporal cortical regions for responding to human voice seems to undergo a period of development over the first year of life with increases in activation to human voice observed between 3-4 and 6–7 months (Blasi et al., 2011; Grossmann et al., 2010; Lloyd-Fox et al., 2012; McDonald et al., 2019). Gaps in the literature provide an incomplete picture of exactly how voice selectivity evolves throughout early childhood, though adult-like voice-selective responses have been noted via fMRI in children 5-to-8 years-old (Abrams et al., 2019; Bonte et al., 2013; Raschle et al., 2014; Rupp et al., 2022).

Though limited in number, results of studies with infants and young children may still contribute to our understanding of how human voice is represented in the auditory cortex and mechanisms that drive voice processing. Findings of differential responses to vocal stimuli compared to synthesized, speech-like sounds in neonates (Cheng et al., 2012) and 2- to 7-month old infants (Calce et al., 2023; Vouloumanos et al., 2010; Vouloumanos and Werker, 2004) lend credence to models proposed in the adult literature suggesting that voice processing in auditory association cortex is not primarily driven by acoustic feature analysis, but rather categorical encoding of a higher-order representation of human voice as an auditory object (Agus et al., 2017; Barbero et al., 2021; Bodin et al., 2021; Leaver and Rauschecker, 2010; Norman-Haignere and McDermott, 2018; Rupp et al., 2022). Despite presentation of stimuli with low-level acoustic similarities, including matched pitch contour (Cheng et al., 2012) and energy peaks (Vouloumanos et al., 2010;

Vouloumanos and Werker, 2004), the developing auditory cortex still appears to respond differently to human voice, pointing to a more abstracted representation than can be explained by acoustic structure alone. Though studies with infants and young children have not explicitly tested this hypothesis with the same granularity that it has been investigated in adults, electrophysiological studies have shown that children as young as 9 years-old demonstrate neural responses indicative of categorical encoding of human voice (Rupp et al., 2022). If voice is encoded as a sound category, it is likely that this representation is not quite as robust in the developing auditory cortex as it is in adults. Acoustic stimuli that are substantially similar in spectrotemporal complexity and structure to human voice (i.e., monkey vocalizations) (Joly et al., 2012), produce similar behavioral (Vouloumanos et al., 2010) and physiological (Minagawa-Kawai et al., 2011) responses in infants up to 4-months old as those observed for voice. Certain lines of research in adults suggest that voice-selective cortical responses are driven by a lower-level sensitivity to a preferred constellation of acoustic features typical of voice but not voice *per se* (Agus et al., 2017; Leaver and Rauschecker, 2010; Staib and Frühholz, 2021, 2022). This model predicts that listeners will respond similarly to human voice and other auditory stimuli that share a common set of properties, even when these sounds are not perceived as human voice (Staib and Frühholz, 2021). While this set of preferred properties may initially include both human and monkey vocalizations in the first few months of life, there is an apparent narrowing in sensitivity that excludes animal vocalizations by 7-months old (Vouloumanos et al., 2010) and ultimately results in the consistent findings of species-specific voice selectivity observed in adulthood (Bodin et al., 2021; Fecteau et al., 2004; Joly et al., 2012b).

Taken together, the results of studies indicating an early similar response to human and monkey vocalization (Minagawa-Kawai et al., 2011; Vouloumanos et al., 2010) but not to other complex non-vocal sounds (Blasi et al., 2011; Cheng et al., 2012; Lloyd-Fox et al., 2012; Vouloumanos and Werker, 2004) could imply that there is an initially broad tuning for a particular profile of acoustic features that gradually refines and becomes more sensitive and species-specific to human voice across development. Mechanistically, this may represent a transition in auditory voice processing from reliance on low-level feature analysis to higher-order categorical representation. This hypothesis suggests that, while neonates and young infants are sensitive to human voice to a degree, they have not yet formed a robust auditory category representing human voice on a higher level. Studies that fail to find voice-selective responses early in development may be explained by less robust higher-order representations. For example, acquisition of and expertise with novel auditory categories has been shown to increase activation in a posterior region of the STS (Leech et al., 2009; Liebenthal et al., 2010). In using non-speech complex stimuli, Leech et al. (2009) further demonstrated that activity in this region was not simply explained by speech-sensitivity but instead driven by a more general expertise in categorizing sounds. This finding has been echoed in the visual domain by Gauthier et al. (1999) who found that learning of novel visual categories led to increased activity in a classically proposed face-sensitive region. In other words, the act of learning and establishing novel perceptual categories may evoke responses traditionally associated with category selective cortical responses. The implication of the posterior STS (pSTS) in the formation of novel auditory categories is particularly interesting given that studies with infants have localized voice-sensitive responses to a slightly more posterior region than those identified in adults (Belin and Grosbras, 2010). Should future research establish that there is indeed a more posterior bias in voice-sensitive responses during infancy, it may be prudent to consider whether this increased activity in the pSTS is evidence of the development of a categorical representation of human voice. Accordingly, we may predict that acquisition and establishment of human voice as an auditory category over the first year of life would result in a gradually decreasing role of the pSTS, thereby shifting responses more anteriorly toward voice-selective regions traditionally observed in older children and

adults.

Future research with infants and children requires a more sensitive and intricate probing of responses in the developing auditory cortex to better understand if and how they evolve with maturation and experience. Replication of studies in infants and young children that utilize acoustic stimuli along a spectrum of acoustic and perceptual similarity to human voice (Agus et al., 2017; Staib and Frühholz, 2021, 2022) would fill in current gaps surrounding what set of acoustic properties or sound categories the auditory cortex is sensitive to. In turn, identification of developmental changes in voice sensitivity/selectivity and localization of these responses may provide valuable insight into the neural computations underlying voice-selective responses and how representations of human voice are developed.

2.2. Voice identity processing

One of the most remarkable feats of voice processing is the ability to recognize individuals by their voices. For the purposes of this manuscript, we use the terms talker recognition and talker identification to broadly refer to the ability to distinguish individual voices from one another and to utilize formed representations to explicitly name or match voice to identity. This ability partially subserves the formation of social bonds and has been demonstrated in both non-human primate (Kojima et al., 2003; Rendall et al., 1996) and other non-primate (Insley, 2000; Seyfarth and Cheney, 2003) species. Voice recognition is an especially complex auditory task because listeners must not only process the ways in which a particular voice differs from all other voices (inter-talker variability) but also maintain identity representation in the face of slight variations in acoustic patterns produced by the same speaker (intra-talker variability) (Lavan et al., 2019a). To provide a concrete example of this complexity, a child must distinguish their own mother's voice versus a stranger's voice telling them to "come here" (i.e., process inter-talker variability) and recognize that on different occasions their mother may produce the same message in slightly different ways if she is yelling, whispering, or sick (i.e., process intra-talker variability).

One of the earliest – and most consistently replicated – observations in the developmental voice processing literature is recognition of maternal voice. Researchers have shown that both near-term fetuses (Jardri et al., 2012; Kisilevsky et al., 2003, 2009) and neonates (Beauchemin et al., 2011; DeCasper et al., 1994; Fifer and Moon, 1994; Lee and Kisilevsky, 2014) exhibit differential cardiac and neural response patterns when presented with their own mother's voice compared to other voices, pointing to an early ability to process identity-related information in the voice signal. Recognition of maternal voice does not, however, translate to an ability to successfully extract identity-related information from all voices. Thus far, studies using habituation procedures along with measures of novelty and orienting responses have suggested that near-term fetuses (J.-P. Lecanuet et al., 1993), newborns (Flocchia et al., 2000), and infants from 6- to 16-months old (Fecher et al., 2019; Fecher and Johnson, 2018b; Friendly et al., 2014; Johnson et al., 2011) can discriminate between talkers beyond maternal voice. This finding may not always be replicated, though, if stimuli are not presented in a familiar language (Johnson et al., 2011) or in a manner that sufficiently captures attention (Flocchia et al., 2000). Overall, current evidence is supportive of an early emerging capability to perform basic contrastive perceptual analyses that allow for the detection of differences between talkers. Investigations of infants' ability to detect differences among talkers are relatively well-served by preferential gaze or habituation paradigms because of their ability to index discrimination of stimuli. On the other hand, examining explicit voice identity recognition in infants, that is, their ability to map and match voice to a particular individual, poses methodological challenges and has been less well investigated. Talker discrimination and talker recognition, likely place slightly different demands on both auditory and cognitive processing (Creel and Jimenez, 2012; Fecher et al., 2019), and at the very least

require different response types that infants may not be able to provide such as pointing, pressing a particular button on a response box, or verbally labeling an identity. Use of the above-mentioned traditional infant behavioral methodologies to index genuine recognition of voice identity may initially have led to an overestimation of infant voice processing abilities (see Fecher et al., 2019 for review). One study that attempted to contend with these methodological issues found that 16.5-month-old children were only able to map voice to visual identity (i.e., form identity representations based on voice) when pairs of voices were highly acoustically distinct from one another (Fecher et al., 2019). This study contributed to the now widely accepted “Protracted Tuning Hypothesis,” which suggests that proficiency in learning, recognizing, and identifying voices gradually and continuously increases throughout childhood (Creel and Jimenez, 2012).

Evidence gathered from studies of preschool-age, school-age, and adolescent children supports the notion of gradual developmental improvements in the ability to discriminate among talkers and identify individual voices. Adults consistently out-perform children on a variety of tasks (Creel and Jimenez, 2012; Fecher and Johnson, 2018a; Levi and Schwartz, 2013; Zaltz, 2023) until roughly 10- to 14-years-old (Mann et al., 1979; Zaltz et al., 2020). Like observations of talker recognition in children at just over a year old, children from 3- to 5-years-old demonstrate the most accuracy when learning and discriminating between new voices presented in pairs that differ substantially from one another acoustically; for example, voices of distinctly different genders or ages (Creel and Jimenez, 2012; Mann et al., 1979). Children appear to become more adept at distinguishing between and recognizing talkers by 5- to 6-years old (Fecher and Johnson, 2018a, 2021) and continue to show improvements in accuracy from 6- to 9-years and further from 10- to 12-years (Levi, 2018; Levi and Schwartz, 2013; Mann et al., 1979). These age-related improvements in identity processing are observed across a variety of task designs including “same-different” judgements (Levi, 2018; Levi and Schwartz, 2013; Mann et al., 1979) and identifying a voice associated with an image or a name from a lineup of learned identities (e.g., alternative forced-choice tasks) (Levi, 2015, 2018; Zaltz, 2023). This is not to say, however, that methodology may not influence study findings. Task components including response modality (e.g., pointing to a picture versus verbally identifying a talker) (Bartholomeus, 1973) and design of stimulus presentation (Zaltz, 2023) appear to impact accuracy. Additionally, the content of stimuli themselves may have an effect on discrimination and identification tasks across ages, specifically whether listeners hear the same utterances across speakers (Mann et al., 1979) and whether stimuli are presented in the native language (Fecher and Johnson, 2018a, 2021). Overall, both infant and child literature have produced mixed results that seem at least partially dependent on methodological variables. When interpreting the results of studies examining voice identity processing in children, it is important to simultaneously consider the role that experimental design and stimulus selection may play in their results and any conclusions drawn from their findings.

One stimulus-related variable that has relevance for both task performance and mechanisms underlying voice identity encoding is the degree of familiarity of the voices presented. While the studies mentioned previously indicate that young children may have difficulty discriminating identity between two novel speakers, children from 4- to 5-years-old consistently perform well above chance level when identifying personally familiar voices including teachers, classmates, and well-known cartoon characters (Bartholomeus, 1973; Jeffries, 2015; Spence et al., 2002). Children around this age also demonstrate high accuracy in identifying recordings of their own voices (Strömbergsson, 2013). Accuracy in the identification of familiar voices increases with both level of familiarity and age, with steady improvements noted in children from 2- to 5-years-old (Jeffries, 2015; Spence et al., 2002). There is some debate as to whether familiar and unfamiliar voices involve slightly different processing demands (Kriegstein and Giraud, 2004; Maguinness et al., 2018; Stevenage, 2018) and, therefore, to what

extent studies using these two types of stimuli can be compared to one another. Furthermore, not all “familiar” voices are necessarily represented or encoded in the same manner, specifically when comparing “trained-to-familiar” voices (Kreiman and Sidtis, 2011) learned in experimental tasks and personally familiar voices (Fontaine et al., 2017; Kanber et al., 2022; Mathias and von Kriegstein, 2019). Although both unfamiliar and familiar voices likely undergo some shared perceptual analysis, representation of familiar voices, especially those that are personally familiar, is nonetheless multi-faceted involving activation of semantic and episodic memory and emotional processing (Lavan and McGettigan, 2023). It may be questioned, then, whether voice recognition performance observed in children listening to personally familiar voices is purely a measurement of voice processing or the product of shared processing systems. Although we cannot draw any conclusions on this matter, we can at least use evidence of differences in performance between identification of personally familiar versus trained-to-familiar voices to support the idea that representation of voice identity exists on a spectrum of familiarity even in children as young as 2- to 3-year-old. Nevertheless, the fact that children perform worse than adults in recognition of trained-to-familiar voices through at least late childhood suggests that fully developed voice processing involves more efficient extraction and encoding of identity-related information. Determining the mechanisms that subservise voice identity encoding is, therefore, crucial in explaining the developmental improvements noted in the literature.

Borrowing from prototype-based models proposed in the face processing literature (Leopold et al., 2001; Rhodes and Jeffery, 2006), the leading theory explaining voice identity perception in the adult literature suggests that voice identification is achieved by analyzing a particular voice’s relative acoustic deviation from a template prototypical voice (Latinus and Belin, 2011; Lavner et al., 2001). This voice prototype is thought to represent the acoustic average of all voices encountered by the listener (Lavner et al., 2001; Maguinness et al., 2018), though multiple prototypes or perceptual categories may exist that assist in the processing of voice identity (Latinus et al., 2013; Lavan and McGettigan, 2023). Some studies also suggest that, given sufficient exposure, individual voice identities are encoded as average representations of a particular talker’s overall variability (Andics et al., 2013; Fontaine et al., 2017; Lavan et al., 2019b). Prototype models predict that those voices which deviate furthest, and are therefore the most acoustically distinct, from the average voice will be most easily recognized (Latinus et al., 2013; Lavner et al., 2001). A few child studies have demonstrated that voices with the most extreme fundamental frequencies (f_0) or most “atypical” acoustic features of those voices used in the study were the most accurately recognized across participants ranging from 3- to 11-years old (Jeffries, 2015; Levi, 2015). This may provide tentative support for the idea that children within this age range are referencing an average voice template when encountering new voices. There is additional evidence in the visual domain that infants as young as 3- to 6-months form prototypic representation of shapes (Bomba and Siqueland, 1983) and faces (de Haan et al., 2001; Rubenstein et al., 1999). Why, then, does there appear to be such a protracted development of adult-like voice identification performance? Unfortunately, no studies with infants or young children have specifically tested prototype-based coding of voice identity, though one explanation may be that children require more prolonged exposure to a variety of talkers and experience with analyzing differences between and within talkers during the first years of life to form the average-based voice representations proposed in adult studies.

There is substantial acoustic variation both between (Kreiman et al., 2015; Mathias and von Kriegstein, 2019) and within (Lavan et al., 2019a) talkers that may make learning of the distributions that define both the overall voice space and a particular talker’s identity especially difficult. Further complicating the analysis of voice identity, it has also been argued that individual voices are defined not by one common set of parameters, but by patterns of overall variation or “gestalt” unique to

each voice (Lavan et al., 2019a; Lavner et al., 2001). Though learning the overall variability of an individual voice appears to be key to identity representation, until sufficient exposure has been attained these patterns of variability may be a detriment to recognition (Lavan et al., 2019b). Given the clear complexity of the perceptual demands required by voice identity processing, it would not be surprising if the development of voice discrimination and recognition rely on a somewhat prolonged period of auditory experience with a wide variety of talkers and multiple experiences with individual talkers. Perhaps developmental improvements in voice discrimination and recognition reflect the gradual development of an average voice percept and increased efficiency in prototype-based analysis of inter- and intra-talker differences.

Though adult studies provide convincing evidence, there is one obvious question regarding the compatibility prototype-based voice identity models in children: If initial analysis of voices relies on comparison to an average voice representation formed via experience, and decreased performance in voice recognition in children is due to a lack of prototypical representation, how can we explain the recognition and discrimination of maternal voice so much earlier in development? If representation of the prototypical voice is built over time, this would predict that children should not be able to build any identity representations until the average voice prototype is formed. In other words, adult models of voice identity processing assume that unfamiliarity begets familiarity. Certain researchers have reconciled this conflict by suggesting that representations of highly familiar voices, particularly maternal voice, are formed first (Kreiman and Sidtis, 2011). In this sense, it could be that early voice identity processing is based not on reference to a veridical average voice, but on representation of highly familiar voice identities like parents and caregivers (Creel and Jimenez, 2012). Though self-voice perception has been largely under-investigated it may be interesting to consider how a child's experience with and representation of their own voice may fit with this idea. One theory that may accommodate findings in the infant literature and prototype-based models proposed in the adult literature is that a highly familiar voice serves as the initial prototypical voice representation. This representation may then morph and develop into the "average" prototype voice proposed in the adult literature given gradual exposure to other talkers. Another possibility is a shift from exemplar-based to prototype-based coding of voices. Under this assumption, babies may be able to discriminate and recognize maternal voice not due to a higher-order identity representation but simply because they have acquired sufficient exemplars. This exemplar-based coding may drive identification of familiar voices until a prototypical voice representation is developed and processing becomes dependent on norm-based coding. Though debates surrounding the theoretical processing hierarchy of voice and voice identity are beyond the scope of this review, the dissonance between early maternal voice recognition and currently proposed prototype-based encoding models is in accordance with recent questioning of whether prototypes serve an obligatory functional role in voice recognition (Lavan and McGettigan, 2023). Longitudinal investigations with infants and children using methodologies employed in the adult literature may provide us with evidence of whether a prototypical voice representation is evident in infancy, where this prototype "sits" in the overall acoustic space, and how it changes over the course of development.

It is, of course, difficult to ignore the potential contributions of general auditory processing and cognition when considering developmental trends in this research. Are improvements noted in voice discrimination and identification due not to refinement of voice processing specifically but rather to more generalized improvements in central auditory processing or cognitive mechanisms? Overall cognitive development may at least partially explain increased performance on certain voice processing tasks. As alluded to previously, differences in task design on auditory tasks can produce varying results by placing increasing demand on cognitive processes such as memory and attention (Rose et al., 2018; Zaltz, 2023). In fact, several studies have explicitly

demonstrated that auditory-related working memory and other cognitive scores are associated with better performance in voice recognition and discrimination (Levi, 2015; Zaltz et al., 2020). In the same vein, developmental improvements in voice identity processing may be related to maturation of wholesale auditory processing rather than refinement of a particular voice processing mechanism. A generally increased ability to analyze changes in the dimensions comprising an acoustic signal may support the ability to process the inter- and intra-talker characteristics that define voice identity. Components of complex sound processing, including temporal integration/resolution and spectral analysis, continue to evolve and improve from roughly 4- to 10-years-old and in some cases may not be fully mature until early adolescence (Allen-Mearns and Wightman, 1992; Buss et al., 2012; Moore and Linthicum, 2007), echoing the relative trajectory observed in voice identity processing. As overall sensitivity to acoustic feature analysis relates to voice processing specifically, children may show elevated thresholds compared to adults in detecting changes in fundamental frequency (f_0) and cues related to vocal tract length (VTL) (Nagels et al., 2020), both of which are considered crucial in differentiating and identifying voices (Darwin et al., 2003). The age at which these thresholds reach adult-like levels may be anywhere from 6- to 8-years-old (Nagels et al., 2020; Zaltz et al., 2020). Voice identity processing requires the simultaneous analysis of both inter- and intra-talker characteristics, involving an intricate balance of attending to talker-specific information and disregarding those acoustic features that do not contribute to identity perception. It is therefore interesting to consider the interplay of cognitive and auditory processing and how development in either domain may benefit the other. General cognitive development may underly the development of increasingly efficient perceptual strategies (e.g., cue integration and weighting) used in auditory processing tasks like speech or voice perception (Nagels et al., 2020; Nittrouer et al., 1993; Petrini and Tagliapietra, 2008).

Despite several caveats in the interpretation of results, there is clearly an improvement in the ability to process voice identity that begins with recognition of maternal voice in infancy and continues well into childhood with learning and identification of novel voices. Given the possible influence of experimental design, cognitive development, and improvements in general auditory processing, we are presently unable to generalize findings across studies and describe how voice identity processing and its underlying mechanisms evolve. To adequately answer this question, future research will require careful selection of stimuli to probe a range of familiar and unfamiliar voices, use of control tasks or measures to examine associations between cognitive domains (e.g., memory and attention) and general auditory processing, and inspiration from methods developed in the adult literature to investigate how coding and processing mechanisms differ or shift over time.

3. Factors modulating development of voice processing

Evidence across listeners from prenatal development to adolescence indicates that voice perception does indeed go through developmental change, therefore further study of what drives or modulates this development is warranted. As already pointed out in the discussion of how overall maturation of auditory processing and cognition may impact voice processing, we cannot assume that the development of voice perception proceeds in isolation from other developmental processes. The following sections address factors that may influence the development of voice processing and assist in explaining the developmental changes observed in the literature. We focus specifically on processes that begin to impact overall neurological development and auditory perceptual mechanisms within the first months of life, thereby having implications for voice processing from its onset.

3.1. Neural plasticity

Observations across development indicate that voice perception is not fully developed at birth, and its maturation likely involves an interaction of both internal and external factors. In this context, whether voice perception is an experience-expectant or experience-dependent process is of interest. The terms experience-dependent and experience-expectant refer to types of neural plasticity and provide explanations for what drives and induces developmental changes in various processes. Briefly, processes that are experience-expectant are those to which an individual is biologically predisposed to develop while processes that are experience-dependent require the input of specific stimuli or experiences to develop (Galván, 2010; Greenough et al., 1987).

In the absence of pathophysiology, voice perception develops without explicit instruction. In this case, it is compelling to label it as a strictly experience-expectant mechanism. However, there are several developmental and cognitive processes that demonstrate both experience-expectant and -dependent features, developing innately but optimized and modulated by experience. The most obvious example of this in the domain of voice processing is familiar voice identity formation and identification; only those voices with which a listener has sufficient experience will be maintained as long-term representations. Differences in representations of average voices, especially prototypes for average male and female voices, may differ slightly cross-linguistically (Andrianopoulos et al., 2001; Pépiot, 2015), though cross-linguistic and cross-cultural variation in voice acoustic characteristics has not been documented extensively enough to draw firm conclusions (Yamauchi et al., 2022). Another potential optimizer of experience-expectant voice processing is maternal voice. Several studies indicate that it may act as a powerful modulatory auditory stimulus for the development of the auditory cortex, language networks, and extended reward circuitry at least up until adolescence (Abrams et al., 2016, 2019, 2022; Beauchemin et al., 2011; Dehaene-Lambertz et al., 2010; Goldberg et al., 2020; Liu et al., 2019; Uchida-Ota et al., 2019; Webb et al., 2015).

3.2. Response to biological relevance and perceptual narrowing

One theory explaining an experience-expectant basis of voice (vocalization) processing not just in humans but across a wide range of species suggests an ecological benefit in applying privilege to conspecific signals. In other words, neural and behavioral responses are driven by the degree of biological relevance contained in a signal. From an evolutionary standpoint it is logical that a species would prioritize responding to cues that promote survival, like indicators of potential danger and reproductive fitness. This idea bears out in animal studies showing that many species recognize and respond preferentially to vocal signals and cues from conspecifics (Bodin and Belin, 2019; Ortiz-Rios et al., 2015; Petkov et al., 2008, 2009; Sidtis and Kreiman, 2012).

Existing research suggests that human infants may initially demonstrate a broad tuning to a variety of visual and auditory stimuli. For example, infants appear to be initially sensitive to the faces (de Haan et al., 2002, 2002, 2002; Pascalis et al., 2002) and voices (Minagawa-Kawai et al., 2011; Shultz et al., 2014; Vouloumanos et al., 2010) of both humans and monkeys. There is, however, an apparent shift toward selectivity for human voice between 3- and 7-months and a possible correlation between age and degree of voice selectivity in portions of the left STS (Blasi et al., 2011; Grossmann et al., 2010; Lloyd-Fox et al., 2012; McDonald et al., 2019). While responses in voice-selective cortex are stronger when speech stimuli are used (Belin et al., 2002), selectivity for conspecific vocalization is evident in humans even when stimuli contain no linguistic information (Rupp et al., 2022) indicating that increased human voice selectivity is not driven strictly by any linguistic information contained within the voice signal.

These changes in neural responses to human voice throughout development could be mediated by a phenomenon called “perceptual

narrowing,” in which frequent experience with a relevant stimulus modulates sensitivity and specificity of responses (Aslin and Pisoni, 1980). Increased sensitivity to relevant stimulus differences comes at the expense of ability to discriminate between stimuli with differences that, through experience, become less relevant. Perceptual narrowing is one proposed mechanism driving the increase in discrimination of native phonemes/contrasts with simultaneous decreases in non-native phonemes (Kuhl et al., 1992; Tsushima et al., 1994). This trade-off has also been observed in investigations of cross-species voice processing in infants, who demonstrate a decreased ability to distinguish between monkey voices from 6- to 12-months (Friendly et al., 2014). Some theories have suggested that the auditory system uses networks shaped by and optimized to meet the demands of “ecologically relevant” tasks and that the creation of novel sound categories specifically alters neural processing to enhance behaviorally relevant features (Ley et al., 2012; Staeren et al., 2009). In other words, experience with biologically meaningful stimuli like human voices and native language influences the formation of neural networks that further enhance responses to those stimuli. Experience-driven increases in the efficiency of voice processing networks may extend through childhood and adolescence. A shift from relatively diffuse to increasingly focal activation in task-related regions is generally thought to represent learning and enhanced cognitive performance (Durston et al., 2006). In line with this idea, an fMRI study has shown that human voice initially evokes more diffuse activation patterns within the STG/S at 8–9 years old, which becomes both increasingly selective and focal (i.e., less widely distributed) from adolescence (14–15 years old) to early adulthood (20–30 years old) (Bonte et al., 2013). A similar pattern of increasingly focal task-selective responses was observed in the same age groups during a voice discrimination task (Bonte et al., 2016).

The idea of a perceptual narrowing to biologically relevant stimuli is particularly compelling when considering research investigating response to familiar voices, specifically maternal voice. There is an evolutionary basis pointing to the primacy of response to familiar voices. Both humans and other species demonstrate very early abilities in recognizing voices of caregivers and kin (Kreiman and Sidtis, 2011; Seyfarth and Cheney, 2003; Sidtis and Kreiman, 2012). This phenomenon could also be due to the unique multisensory transmission of maternal voice while in utero, which may promote early learning (Kumar et al., 2023). The environmental conditions in utero create a natural low-pass filter and favor transmission of prosodic level information (Ghio et al., 2021; Granier-Deferre et al., 2011; J.-P. Lecanuet et al., 1993). While signals transmitted at this stage of development are not as rich with information as those experienced ex-utero, some argue that the reduced complexity naturally available to fetuses decreases cognitive load and enhances the ability to integrate acoustic information over longer timescales (Turkewitz and Kenny, 1982; Vogelsang et al., 2023). Early experience with prosodic level information over longer integration windows may prime the developing auditory system to key into speaker-specific information, conveyed for example by the fundamental frequency contours of a voice that survive the diminishment of higher-frequency information by the low-pass filtering properties of the prenatal environment.

Beyond voices, there is some evidence that responses within the STS are driven by experience-expectant responses to more general socially relevant stimuli. Biological relevance seems to play a role in the functioning of extended portions of the STS outside of traditionally proposed TVAs, especially in posterior regions. Some have suggested that the STS functions cross-modally as a social processor (Redcay, 2008). Indeed, the human STS is responsive to species-specific signals in a much broader sense than just vocalization, demonstrating significant activation to a variety of dynamic visual stimuli that are perceived as human. Non-auditory stimuli that have been noted to induce activity in the STS include facial expressions (Redcay, 2008), sign language (Levänen et al., 2001; Sadato et al., 2004), and even simple line drawings evoking human features (Kingstone et al., 2004).

3.3. Cerebral specialization

Models of voice processing do not necessarily provide a mechanistic explanation for how anatomical or physiological characteristics of the auditory cortex subserve response to voice. Especially relevant in this discussion are the potential contributions of cerebral specialization. It has long been suggested and supported in the literature that the left hemisphere demonstrates a propensity for speech and language processing (Broca, 1861). The neural substrate of this lateralization is suggested to be structural asymmetries in the perisylvian region (Fox, 1991; Steinmetz, 1996; Witelson and Pallie, 1973), which may be further modified by genetic factors (P. M. Thompson et al., 2001). Sulcal and volumetric differences between the left and right temporal lobes are evident even in utero (Dehaene-Lambertz et al., 2006) and persist through adulthood (Bonte et al., 2013; Leroy et al., 2015). Interestingly, peak voice selectivity has been observed to be associated with sulcal depth in the STS (Bodin et al., 2018). The presence of this asymmetry early in development along with findings of right hemisphere bias in voice selectivity may indicate an early specialization specifically of the right STS for voice processing (Simon et al., 2009; Grossmann et al., 2010; Blasi et al., 2011).

A frequent explanation of observed right hemisphere biases for voice processing (Belin et al., 2002; Fecteau et al., 2004) is its apparent involvement in analyzing affect or prosody (Fox, 1991). Given that prosodic level information can be key to identifying idiosyncrasies among individual speakers, this may explain the heavy implication of the right anterior STS in processing identity-specific information in voices (Aglieri et al., 2021; Belin et al., 2004; Belin and Zatorre, 2003; Kriegstein and Giraud, 2004; Schall et al., 2015; von Kriegstein et al., 2003; Zhang et al., 2021). A more recent theory explaining this propensity suggests that the left and right hemispheres are specialized for processing fast temporal and slow spectral changes, respectively (Albouy et al., 2020; Belin et al., 1998; Flinker et al., 2019; Zatorre and Belin, 2001). Complex auditory stimuli comprise dynamic changes in both temporal and spectral domains. Temporal resolution is implicated in processing rapid changes in the frequency domain that characterize speech sounds, while spectral resolution is implicated in slower state changes in harmonic structure related to prosody and melody (Albouy et al., 2020; Baum and Pell, 1999). If the right hemisphere is predisposed to processing information naturally carried over longer timescales and, as previously discussed, this information is the most salient to the early developing auditory system, then findings of an advantage for right hemisphere voice processing, especially in fetal and infant research, may not be surprising (Blasi et al., 2011; Cheng et al., 2012; Grossmann et al., 2010; Simon et al., 2009). Taken together, early emerging anatomical differences or asymmetries and prenatal auditory experience primarily with prosodic-level acoustic information may underlie the functional specialization of the right STS for voice processing (Schönwiesner et al., 2005).

3.4. Language acquisition

Perhaps one of the most intriguing questions surrounding the development of voice processing is how it interacts with language acquisition. While speech- and talker-specific information can be independently extracted and are distinct to a degree (Belin et al., 2004; Maguinness et al., 2018; von Kriegstein et al., 2003) they are inherently intertwined due to the simple fact that they are simultaneously carried by the same signal. For this reason, it can be difficult to untangle responses to voice from those that are also related to linguistic processing (Luthra, 2021). Results of studies published to date point toward an interaction between speech and voice processing, in which speech and voice perception may be facilitative of one another. For example, familiarity with a voice increases intelligibility when presented in the context of competing stimuli (Barker and Newman, 2004; Johnsrude et al., 2013; Nygaard and Pisoni, 1998) even when voices are not

explicitly recognized as familiar (Holmes et al., 2018). Similarly, increased language experience in the form of bilingualism benefits voice perception by allowing for faster learning of novel voices (Levi, 2018), and earlier acquisition of a second language seems to augment this effect (Bregman and Creel, 2014). Infants as young as a few days old already exhibit differential neurophysiological and behavioral responses to voices speaking familiar versus unfamiliar languages, suggesting that prenatal experience primes the auditory system to tune to voice stimuli in the native language (May et al., 2011; Moon et al., 2013).

The Language Familiarity Effect (LFE) is one of the most well-established pieces of evidence highlighting the relationship between speech and voice processing. LFE refers to the behavioral observation that familiarity with a language increases a listener's ability to discriminate, recognize, and identify voices (Goggin et al., 1991; C. P. Thompson, 1987). The LFE has been demonstrated in school-age children (7-15-years-old) in talker identification, recognition, and discrimination tasks (Goggin et al., 1991; Levi, 2018; Perea et al., 2014). The point at which language familiarity benefits voice identification is not known but, crucially, one study observed the LFE in 6-year-olds but not 5-year-olds (Fecher and Johnson, 2018a, 2021). This effect appears to not be as powerful in tasks with lower cognitive demand and may not be observed in adults or children in simple discrimination tasks (Fecher et al., 2019; Levi and Schwartz, 2013).

A current debate is the mechanism by which language acquisition and familiarity support voice processing; namely, whether full semantic or linguistic knowledge is required (Bregman and Creel, 2014; Perrachione and Wong, 2007) or if phonological knowledge of a language is sufficient (Fecher and Johnson, 2018b; Fleming et al., 2014; Johnson et al., 2011). In general, studies that manipulate the level of phonological and semantic content within their stimuli have found that, while phonological information alone may be sufficient, increasing the amount of familiar linguistic information in a stimulus is correlated with increased performance on talker identification tasks (Goggin et al., 1991; Perrachione et al., 2015; Zarate et al., 2015). Like the suggestion that familiarity with a voice decreases the cognitive demand of speech perception (Heald and Nusbaum, 2014; Holmes and Johnsrude, 2020), it could be that increasing familiarity with the phonetic and semantic content of a language decreases the cognitive load of voice identity processing, perhaps by engaging additional left-hemisphere circuits in a right-hemisphere predominated process (Perrachione et al., 2009). Further research with infants and young children over the first few years of life may help in examining how the relationship between speech and voice processing develops by analyzing potential correlations or associations between phonetic and semantic knowledge with performance on voice identification tasks.

4. Effects of pathophysiology

Examining the typical developmental trajectory of voice processing is the primary goal of this review, however examining its functioning in pathophysiological states is also valuable. While voice processing may be impacted by both congenital and acquired disorders, for the purposes of this review we limit our discussion to three specific populations: premature infants, individuals with autism spectrum disorder, and individuals with developmental phonagnosia. These populations were selected because they implicate human voice processing from its nascence and represent disruption to the developmental trajectory during childhood rather than impairment of an already developed mechanism as might be the case with acquired neurological injury in adulthood. Continued study of the abnormalities evident in auditory and voice processing in these clinical populations not only provides us with information on how voice perception is impacted in pathological states but may also inform our understanding of the mechanisms that underlie voice processing and how these neural underpinnings typically develop.

4.1. Premature birth

Premature birth disrupts the normal course of neural development, which undergoes several critical processes during the third trimester. It is associated with a variety of health complications and poor long-term developmental outcomes (McCormick et al., 2011). As it relates specifically to auditory development, prematurity is associated with abnormalities both in the structure of primary auditory and association cortex (Therien et al., 2004; Wang et al., 2023), and in functional measures of central auditory processing (Didoné et al., 2021; Durante et al., 2018). While structural abnormalities in the premature brain may independently be associated with unfavorable functional outcomes (Dubois et al., 2008), it is equally important to consider the environmental circumstances of prematurity that may pose a detriment to the development of auditory processing. Principally, premature infants may be impoverished in their exposure to human voice and exposed to excessive environmental noise compared to their full-term peers (McMahon et al., 2012).

In the domain of voice processing, there is some evidence indicating that preterm infants may process maternal and unfamiliar voices in a different manner than their full-term peers (Adam-Darque et al., 2020; Filippa et al., 2023), and may not show typical markers of maternal voice recognition at term-equivalent age (Therien et al., 2004). Given the suspected role of maternal voice in shaping auditory and linguistic development, these findings may have important implications for infants born preterm. Some groups have shown preliminary evidence that targeted exposure to maternal voice may improve short-term medical outcomes in preterm infants (see (Filippa et al., 2017) for a review).

On the other hand, there is an argument that increased extrauterine auditory exposure afforded to preterm infants may benefit the auditory system by stimulating cortical growth and enhancing neural processing (Webb et al., 2015). For example, several studies have found more mature event-related potential responses (Adam-Darque et al., 2020) and fMRI activation patterns (Simon et al., 2009) among preterm infants compared to full-term peers. Whether or not this affords any functional advantage for preterm infants relative to their full-term peers remains unknown. While several studies have identified that premature infants are capable of syllable discrimination equivalent to full-term infants, no study has yet demonstrated superior performance of preterm infants on measures of auditory processing or perception (Mahmoudzadeh et al., 2013; Peña et al., 2012).

4.2. Autism spectrum disorder

Research on autism spectrum disorder (ASD) has identified differences in auditory processing compared to neurotypical listeners, including responses to human voice. Implicated in this observation are atypical developmental patterns in the region of the superior temporal sulcus/gyrus (Bailey et al., 1998; Boddaert et al., 2004b; Zilbovicius et al., 2006). These structural abnormalities may be accompanied by physiological anomalies like reduced blood flow during the presentation of complex sounds including speech (Boddaert et al., 2004; Lombardo et al., 2015). Structural and physiological differences may be irrelevant if they don't also constitute a functional impairment. Certain studies, however, have demonstrated that physiological measures like cerebral blood flow (Ohnishi et al., 2000) and resting state connectivity (Abrams et al., 2013) are correlated with overall symptom severity in ASD.

Evidence of the impact of ASD on voice processing has been inconsistent. Several studies examining voice-selective responses in individuals with ASD have been unable to localize any voice-selective regions (Gervais et al., 2004) or responses (Bidet-Caulet et al., 2017), though others have identified TVAs similar to those identified in neurotypical populations (Schelinski et al., 2016). Studies investigating voice processing in both adults and adolescents with ASD have shown similar performance to neurotypical controls in both voice discrimination and identification tasks (Clopper et al., 2013; Groen et al., 2008; Lin

et al., 2015). One study even found that adults with ASD demonstrated faster reaction times than neurotypical controls in classifying stimuli as human voice (Lin et al., 2016). Conversely, a study on preschoolers with ASD found behavioral preferences for non-speech stimuli over those that contained human voice (Kuhl et al., 2005), and infants 4-7-months-old at high risk for developing ASD showed similar preference patterns (Blasi et al., 2015). Given mixed results, it is difficult to determine if and to what extent voice processing is impacted in ASD.

The overall social impairment characterizing ASD and structural/functional abnormalities localized to the STS provide support for theories that consider the STS a "social processor" (Redcay, 2008; Zilbovicius et al., 2006). Even more convincing of the role of the STS in responding to biologically relevant stimuli like human voice, are studies that have identified underconnectivity between putative TVAs and reward circuitry (Abrams et al., 2013). Connectivity between these same systems has been observed in neurotypical children as correlated with and predictive of social communication abilities (Abrams et al., 2016). Given the previous discussion of maternal voice as a potentially powerful modulator of auditory processing, it is also interesting to consider that young children with ASD may not show the usual preference for maternal voice observed in their neurotypical peers (Klin, 1991).

4.3. Developmental phonagnosia

Phonagnosia, a term first proposed as an auditory correlate of visual prosopagnosia (Van Lancker and Canter, 1982), describes the inability to recognize familiar voices. This impairment has been documented as occurring both subsequent to stroke (Neuner and Schweinberger, 2000; Van Lancker et al., 1988, 1989) and developmentally (Garrido et al., 2009; Roswadowitz et al., 2014; Xu et al., 2015), the latter of which we will focus on here. Presently, literature published on developmental phonagnosia is sparse and an important limitation as it relates to the content of this review is that it has been conducted in adult participants who report lifelong difficulty with voice recognition. Therefore, in discussing their findings we operate under the same assumption as authors of these studies that participants' deficits are congenital in nature.

Current estimates suggest that developmental phonagnosia is prevalent in roughly 3–4% of the population (Shilowich and Biederman, 2016). While rare, developmental phonagnosia provides a unique opportunity to examine voice processing, specifically voice identity processing, in relative isolation. Individuals with developmental phonagnosia present with normal peripheral hearing acuity and with no evidence of structural abnormalities on imaging of the brain (Garrido et al., 2009; Roswadowitz et al., 2017; Xu et al., 2015). This impairment in voice recognition is not representative of wholesale impairment in person recognition as studies have documented intact recognition of familiar faces. Similarly, developmental phonagnosia does not seem to be the product of a central auditory processing disorder given normal performance on a variety of control tasks including speech comprehension (Garrido et al., 2009; Roswadowitz et al., 2017), music perception (Garrido et al., 2009; Roswadowitz et al., 2014), perception of emotion in voice (Garrido et al., 2009), and recognition/discrimination of other auditory categories including environmental sounds (Garrido et al., 2009). Perhaps most interestingly, much of the existing literature has documented intact discrimination of unfamiliar voices on basic "same/different" judgements and intact discrimination of voices based on basic talker characteristics such as gender (Garrido et al., 2009; Xu et al., 2015). These findings suggest that developmental phonagnosia is not a difficulty in the "online" acoustic or perceptual analysis of voices, but rather points toward an impairment in the long-term encoding of talker-specific acoustic information and the association of acoustic with semantic information (e.g., name, face) to successfully form an identity representation.

There is some debate surrounding the functional basis of developmental phonagnosia. One study has documented difficulty with

discrimination-based tasks (i.e., judging whether two sentences were spoken by the same or two different talkers) alongside impairment in recognizing familiar voices (Roswadowitz et al., 2014). This has led to the proposal of two distinct phenotypes of phonagnosia: aperceptual phonagnosia in which a deficit in lower-level perceptual analysis disrupts processing and recognition of voices, and associative phonagnosia in which a higher-level deficit disrupts encoding of acoustic information into a percept of identity (Gainotti et al., 2023; Roswadowitz et al., 2017). Roswadowitz et al. (2014) documented this dissociation in two participants who both performed significantly worse on a voice identification task compared to control subjects but differed in their ability to perform a voice discrimination task. These behavioral observations were later backed up by imaging data demonstrating distinct deviations from controls during talker recognition tasks, with one participant exhibiting reduced activation to familiar voices in voice-selective regions and the other with reduced connectivity between voice-selective regions and extra-temporal regions implicated in identity processing including the amygdala (Roswadowitz et al., 2017). Compared to other studies of developmental phonagnosia, this research suggests that the observed inability to recognize familiar voices may not be due to the same core deficit across cases.

Though explicit examination of developmental phonagnosia in children has not yet been published, early identification of phonagnosia and recruitment of child participants would provide an interesting opportunity to compare performance between participants with immature, but typically developing voice identity processing and those with phonagnosia. The fact that the typical developmental trajectory indicates that voice identity processing is not fully mature until later in childhood introduces some potential difficulty into the investigation of phonagnosia in children; however, we have already discussed that identification of personally familiar voices is likely evident by roughly 3- to 4-years. Studies investigating phonagnosia across a wide range of ages in childhood (i.e., from 3- to 12-years-old) may provide evidence of when and how the development of voice processing diverges from the typical trajectory, increasing our knowledge of the true basis of developmental phonagnosia. Additionally, studies of a longitudinal nature would allow for investigation of the possibility that individuals with phonagnosia develop compensatory strategies for voice identity processing (Roswadowitz et al., 2017).

5. Future research consideration and conclusions

Perception and processing of the human voice subserves a multitude of important social functions including forming interpersonal bonds, recognizing familiar individuals from strangers, and perceiving the thoughts and emotions of others. The current state of the developmental literature on voice perception, however, leaves us with an incomplete picture of how response to human voice evolves in childhood and how external factors may help or hinder achieving adult-like competence. In this review, we synthesized the available literature on early development of human voice perception providing evidence that there is indeed developmental change in voice selectivity and voice identity processing. Crucially, we have demonstrated that this topic is under-researched and suggest that continued investigation is necessary not only in understanding how voice perception develops but in furthering our understanding of its substrates and neural underpinning.

Of primary importance, continued research in this area will assist in fully describing the developmental trajectory of human voice perception. Currently, much of the research on voice processing in children is concentrated on infants and school-aged children with very few studies targeting the age range between these two groups. This creates a substantial gap in our knowledge during the critical period of overall development that occurs during the first few years of life (1- to 3-years-old). For example, we know that infants recognize their mother's voices and that children 4- to 5-years-old recognize the voices of their classmates and teachers at levels above chance but have not precisely defined

when children begin to demonstrate the ability to recognize familiar voices beyond that of their mother (e.g., other family members), a skill that may very well emerge within this under-investigated age range. Considering the potential impacts of cognitive, auditory, and language development discussed in this review, and their rapid maturation during the first few years of life, future research that methodically controls for these factors could be particularly useful in unraveling how development in these areas interact with one another. An accurate understanding of how voice processing develops will need to consider this interplay and carefully examine how variables such as stimulus type (e.g., human voice, animal vocalizations, familiarity of voices), stimulus content (e.g., duration, linguistic content, combining auditory and visual modalities), and response modality (e.g., visual, motor, verbal) impact results and, especially, their interpretation and generalization.

We have also made the argument that research with infants and children has the potential to resolve current debates surrounding models of voice selectivity and identity processing in the adult literature. Applying the stimuli and analysis methods utilized in adult studies of voice selectivity (Agus et al., 2017; Norman-Haignere and McDermott, 2018; Staib and Frühholz, 2021) in studies with children will not only provide additional data supporting the investigation of how human voice is encoded in the auditory cortex (i.e., feature or category level), but will also allow us to investigate how encoding mechanisms may change with development. Specifically, we suggest further examination of the transition from a broadly-tuned sensitivity to human voice that may initially include the vocalizations of other species, to true conspecific voice selectivity observed in adolescents and adults. This research may also consider how this transition is supported by auditory experience, perceptual narrowing, increased efficiency in auditory category learning/encoding, or a combination of these factors. Related to voice identity processing, we have pointed out a potential discrepancy between prototype-based models from adult studies and observations of pre- and neonatal recognition of maternal voice. Replication of methods used in the adult literature (Andics et al., 2013; Latinus et al., 2013; Latinus and Belin, 2011) could identify whether children appear to process voices in reference to a prototype and delineate at what point in development children evidence existence of a prototypical voice representation. Further conflicting evidence may cause us to question, as others have (Lavan and McGettigan, 2023), whether prototype-based voice encoding is essential to identity perception and how identity perception might alternatively be achieved in the absence of a prototypical representation. Studies of this nature could be revelatory not only in examining voice identity perception in infants and children, but also in describing the development of general identity processing and how abilities may overlap or diverge across perceptual modalities.

Further, we encourage the investigation of voice processing in both neurotypical and neurodivergent populations including infants born preterm and children with autism spectrum disorder, which both implicate the development of voice processing. Early identification and examination of children with developmental prosopagnosia additionally provides a unique platform for the investigation of voice identity processing in isolation. Exploration of how the development voice perception deviates from the typical trajectory in these populations can inform our knowledge of the role of early auditory experience and environment, the extended voice processing network, and the mechanistic foundations of voice processing.

Existing research suggests that the response to and analysis of human voice goes through a protracted period of development. While newborns certainly evidence emerging preferences for voice over other sounds, voice-selective responses and processing identity through voice do not reach adult-like levels until late childhood or early adolescence. The prolonged nature of this trajectory is likely due to a combination of predetermined development, anatomical maturation, functional cerebral specialization, and auditory experience, though we have yet to determine how all these factors affect development of voice perception and how they may interact with one another. Continued investigation of

voice processing abilities at discrete intervals from prenatal through early childhood stages will map the ontogeny of voice perception and contribute to discussions surrounding the variables that drive its development. A more detailed description of the developmental timeline may additionally provide us with a better understanding of the underpinnings of voice processing and how deviations from the normal developmental trajectory can be used as an early indicator of pathology. Considering the significance of human voice in social, emotional, and linguistic development, further research on voice perception in infants, children, and adolescents is not only interesting but a necessary contribution to the fields of auditory and developmental neuroscience.

Funding

Research was supported by the National Institutes of Health-National Institute on Deafness and Other Communication Disorders (R21DC019217).

CRediT authorship contribution statement

Emily E. Harford: Conceptualization, Writing – original draft, Writing – review & editing. **Lori L. Holt:** Conceptualization, Writing – review & editing, Supervision. **Taylor J. Abel:** Conceptualization, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Taylor J. Abel reports funding was provided by University of Pittsburgh School of Medicine.

Data availability

No data was used for the research described in the article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.crneur.2024.100127>.

References

- Abrams, D.A., Chen, T., Odriozola, P., Cheng, K.M., Baker, A.E., Padmanabhan, A., Ryali, S., Kochalka, J., Feinstein, C., Menon, V., 2016. Neural circuits underlying mother's voice perception predict social communication abilities in children. *Proc. Natl. Acad. Sci. USA* 113 (22), 6295–6300. <https://doi.org/10.1073/pnas.1602948113>.
- Abrams, D.A., Lynch, C.J., Cheng, K.M., Phillips, J., Supekar, K., Ryali, S., Uddin, L.Q., Menon, V., 2013. Underconnectivity between voice-selective cortex and reward circuitry in children with autism. *Proc. Natl. Acad. Sci. USA* 110 (29), 12060–12065. <https://doi.org/10.1073/pnas.1302982110>.
- Abrams, D.A., Mistry, P.K., Baker, A.E., Padmanabhan, A., Menon, V., 2022. A neurodevelopmental shift in reward circuitry from mother's to nonfamilial voices in adolescence. *J. Neurosci.* 42 (20), 4164–4173. <https://doi.org/10.1523/JNEUROSCI.2018-21.2022>.
- Abrams, D.A., Padmanabhan, A., Chen, T., Odriozola, P., Baker, A.E., Kochalka, J., Phillips, J.M., Menon, V., 2019. Impaired voice processing in reward and salience circuits predicts social communication in children with autism. *Elife* 8, e39906. <https://doi.org/10.7554/eLife.39906>.
- Adam-Darque, A., Pittet, M.P., Grouiller, F., Rihs, T.A., Leuchter, R.H.-V., Lazeyras, F., Michel, C.M., Hüppi, P.S., 2020. Neural correlates of voice perception in newborns and the influence of preterm birth. *Cerebr. Cortex* 30 (11), 5717–5730. <https://doi.org/10.1093/cercor/bhaa144>.
- Aglieri, V., Cagna, B., Velly, L., Takerkart, S., Belin, P., 2021. fMRI-based identity classification accuracy in left temporal and frontal regions predicts speaker recognition performance. *Sci. Rep.* 11 (1) <https://doi.org/10.1038/s41598-020-79922-7>. Article 1.
- Agus, T.R., Paquette, S., Suied, C., Pressnitzer, D., Belin, P., 2017. Voice selectivity in the temporal voice area despite matched low-level acoustic cues. *Sci. Rep.* 7 (1) <https://doi.org/10.1038/s41598-017-11684-1>. Article 1.
- Albouy, P., Benjamin, L., Morillon, B., Zatorre, R.J., 2020. Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science* 367 (6481), 1043–1047. <https://doi.org/10.1126/science.aaz3468>.
- Allen-Meares, P., Wightman, F., 1992. Spectral pattern discrimination by children. *J. Speech Hear. Res.* 35 (1), 222. <https://doi.org/10.1044/jshr.3501.222>.
- Andics, A., McQueen, J.M., Petersson, K.M., 2013. Mean-based neural coding of voices. *Neuroimage* 79, 351–360. <https://doi.org/10.1016/j.neuroimage.2013.05.002>.
- Andrianopoulos, M.V., Darrow, K., Chen, J., 2001. Multimodal standardization of voice among four multicultural populations formant structures. *J. Voice* 15 (1), 61–77. [https://doi.org/10.1016/S0892-1997\(01\)00007-8](https://doi.org/10.1016/S0892-1997(01)00007-8).
- Aslin, R.N., Pisoni, D.B., 1980. Some developmental processes in speech perception. In: *Child Phonology*, vol. 2. Brill Academic Publishers, pp. 67–96.
- Bailey, A., Luthert, P., Dean, A., Harding, B., Janota, I., Montgomery, M., Rutter, M., Lantos, P., 1998. A clinicopathological study of autism. *Brain: J. Neurol.* 121 (5), 889–905.
- Barbero, F.M., Calce, R.P., Talwar, S., Rössion, B., Collignon, O., 2021. Fast periodic auditory stimulation reveals a robust categorical response to voices in the human brain. *eNeuro* 8 (3). <https://doi.org/10.1523/ENEURO.0471-20.2021>. ENEURO.0471-20.2021.
- Barker, B.A., Newman, R.S., 2004. Listen to your mother! The role of talker familiarity in infant streaming. *Cognition* 94 (2), B45–B53. <https://doi.org/10.1016/j.cognition.2004.06.001>.
- Bartholomew, B., 1973. Voice identification by nursery school children. *Canadian Journal of Psychology/Revue Canadienne de Psychologie* 27 (4), 464–472. <https://doi.org/10.1037/h0082498>.
- Baum, S.R., Pell, M.D., 1999. The neural bases of prosody: insights from lesion studies and neuroimaging. *Aphasiology* 13 (8), 581–608. <https://doi.org/10.1080/026870399401957>.
- Beauchemin, M., González-Frankenberger, B., Tremblay, J., Vannasing, P., Martínez-Montes, E., Belin, P., Béland, R., Francoeur, D., Carceller, A.-M., Wallois, F., Lassonde, M., 2011. Mother and stranger: an electrophysiological study of voice processing in newborns. *Cerebr. Cortex* 21 (8), 1705–1711. <https://doi.org/10.1093/cercor/bhq242>.
- Belin, P., Boehme, B., McAleer, P., 2017. The sound of trustworthiness: acoustic-based modulation of perceived voice personality. *PLoS One* 12 (10), e0185651. <https://doi.org/10.1371/journal.pone.0185651>.
- Belin, P., Fecteau, S., Bédard, C., 2004. Thinking the voice: neural correlates of voice perception. *Trends Cognit. Sci.* 8 (3), 129–135. <https://doi.org/10.1016/j.tics.2004.01.008>.
- Belin, P., Grosbras, M.-H., 2010. Before speech: cerebral voice processing in infants. *Neuron* 65 (6), 733–735. <https://doi.org/10.1016/j.neuron.2010.03.018>.
- Belin, P., Zatorre, R.J., 2003. Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport* 14 (16), 2105.
- Belin, P., Zatorre, R.J., Ahad, P., 2002. Human temporal-lobe response to vocal sounds. *Cognit. Brain Res.* 13 (1), 17–26. [https://doi.org/10.1016/S0926-6410\(01\)00084-2](https://doi.org/10.1016/S0926-6410(01)00084-2).
- Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. *Nature* 403 (6767), 309–312. <https://doi.org/10.1038/35002078>.
- Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M.-C., Samson, Y., 1998. Lateralization of speech and auditory temporal processing. *J. Cognit. Neurosci.* 10 (4), 536–540. <https://doi.org/10.1162/089892998562834>.
- Bidet-Caulet, A., Latinus, M., Roux, S., Malvy, J., Bonnet-Brilhault, F., Bruneau, N., 2017. Atypical sound discrimination in children with ASD as indicated by cortical ERPs. *J. Neurodev. Disord.* 9 (1), 13. <https://doi.org/10.1186/s11689-017-9194-9>.
- Blasi, A., Lloyd-Fox, S., Sethna, V., Brammer, M.J., Mercure, E., Murray, L., Williams, S.C.R., Simmons, A., Murphy, D.G.M., Johnson, M.H., 2015. Atypical processing of voice sounds in infants at risk for autism spectrum disorder. *Cortex* 71, 122–133. <https://doi.org/10.1016/j.cortex.2015.06.015>.
- Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., Deeley, Q., Barker, G.J., Renvall, V., Deoni, S., Gasston, D., Williams, S.C.R., Johnson, M.H., Simmons, A., Murphy, D.G.M., 2011. Early specialization for voice and emotion processing in the infant brain. *Curr. Biol.* 21 (14), 1220–1224. <https://doi.org/10.1016/j.cub.2011.06.009>.
- Boddaert, N., Chabane, N., Belin, P., Bourgeois, M., Royer, V., Barthelemy, C., Mouren-Simeoni, M.-C., Philippe, A., Brunelle, F., Samson, Y., Zilbovicius, M., 2004a. Perception of complex sounds in autism: abnormal auditory cortical processing in children. *Am. J. Psychiatr.* 161 (11), 2117–2120. <https://doi.org/10.1176/appi.ajp.161.11.2117>.
- Boddaert, N., Chabane, N., Gervais, H., Good, C.D., Bourgeois, M., Plumet, M.-H., Barthélémy, C., Mouren, M.-C., Artiges, E., Samson, Y., Brunelle, F., Frackowiak, R.S.J., Zilbovicius, M., 2004b. Superior temporal sulcus anatomical abnormalities in childhood autism: a voxel-based morphometry MRI study. *Neuroimage* 23 (1), 364–369. <https://doi.org/10.1016/j.neuroimage.2004.06.016>.
- Bodin, C., Belin, P., 2019. Exploring the cerebral substrate of voice perception in primate brains. *Phil. Trans. Biol. Sci.* 375 (1789), 20180386 <https://doi.org/10.1098/rstb.2018.0386>.
- Bodin, C., Takerkart, S., Belin, P., Coulon, O., 2018. Anatomic-functional correspondence in the superior temporal sulcus. *Brain Struct. Funct.* 223 (1), 221–232. <https://doi.org/10.1007/s00429-017-1483-2>.
- Bodin, C., Trapeau, R., Nazarian, B., Sein, J., Degiovanni, X., Baurberg, J., Rapha, E., Renaud, L., Giordano, B.L., Belin, P., 2021. Functionally homologous representation of vocalizations in the auditory cortex of humans and macaques. *Curr. Biol.* 31 (21), 4839–4844.e4. <https://doi.org/10.1016/j.cub.2021.08.043>.
- Bomba, P.C., Siqueland, E.R., 1983. The nature and structure of infant form categories. *J. Exp. Child Psychol.* 35 (2), 294–328. [https://doi.org/10.1016/0022-0965\(83\)90085-1](https://doi.org/10.1016/0022-0965(83)90085-1).

- Bonte, M., Frost, M.A., Rutten, S., Ley, A., Formisano, E., Goebel, R., 2013. Development from childhood to adulthood increases morphological and functional inter-individual variability in the right superior temporal cortex. *Neuroimage* 83, 739–750. <https://doi.org/10.1016/j.neuroimage.2013.07.017>.
- Bonte, M., Ley, A., Scharke, W., Formisano, E., 2016. Developmental refinement of cortical systems for speech and voice processing. *Neuroimage* 128, 373–384. <https://doi.org/10.1016/j.neuroimage.2016.01.015>.
- Bregman, M.R., Creel, S.C., 2014. Gradient language dominance affects talker learning. *Cognition* 130 (1), 85–95. <https://doi.org/10.1016/j.cognition.2013.09.010>.
- Broca, P., 1861. Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). *Bulletin de La Société Anatomique* 6, 330–357.
- Buss, E., Hall, J.W., Grose, J.H., 2012. Development of auditory coding as reflected in psychophysical performance. In: Werner, L., Fay, R.R., Popper, A.N. (Eds.), *Human Auditory Development*. Springer, pp. 107–136. https://doi.org/10.1007/978-1-4614-1421-6_4.
- Calce, R.P., Rekow, D., Barbero, F.M., Kiseleva, A., Talwar, S., Leleu, A., Collignon, O., 2023. Voice categorization in the four-month-old human brain. *Curr. Biol.* <https://doi.org/10.1016/j.cub.2023.11.042>.
- Cheng, Y., Lee, S.-Y., Chen, H.-Y., Wang, P.-Y., Decety, J., 2012. Voice and emotion processing in the human neonatal brain. *J. Cognit. Neurosci.* 24 (6), 1411–1419. https://doi.org/10.1162/jocn_a.00214.
- Clopper, C.G., Rohrbeck, K.L., Wagner, L., 2013. Perception of talker age by young adults with high-functioning autism. *J. Autism Dev. Disord.* 43 (1), 134–146. <https://doi.org/10.1007/s10803-012-1553-5>.
- Creel, S.C., Jimenez, S.R., 2012. Differences in talker recognition by preschoolers and adults. *J. Exp. Child Psychol.* 113 (4), 487–509. <https://doi.org/10.1016/j.jecp.2012.07.007>.
- Darwin, C.J., Brungart, D.S., Simpson, B.D., 2003. Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *J. Acoust. Soc. Am.* 114 (5), 2913–2922. <https://doi.org/10.1121/1.1616924>.
- de Haan, M., Johnson, M.H., Maurer, D., Perrett, D.I., 2001. Recognition of individual faces and average face prototypes by 1- and 3-month-old infants. *Cognit. Dev.* 16 (2), 659–678. [https://doi.org/10.1016/S0885-2014\(01\)00051-X](https://doi.org/10.1016/S0885-2014(01)00051-X).
- de Haan, M., Pascalis, O., Johnson, M.H., 2002. Specialization of neural mechanisms underlying face recognition in human infants. *J. Cognit. Neurosci.* 14 (2), 199–209. <https://doi.org/10.1162/089992902317236849>.
- DeCasper, A.J., Fifer, W.P., 1980. Of human bonding: newborns prefer their mothers' voices. *Science* 208 (4448), 1174–1176. <https://doi.org/10.1126/science.7375928>.
- DeCasper, A.J., Lecanuet, J.-P., Busnel, M.C., Granier-Deferre, C., Maugeais, R., 1994. Fetal reactions to recurrent maternal speech. *Infant Behav. Dev.* 17 (2), 159–164. [https://doi.org/10.1016/0163-6383\(94\)90051-5](https://doi.org/10.1016/0163-6383(94)90051-5).
- Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., 2006. Nature and nurture in language acquisition: anatomical and functional brain-imaging studies in infants. *Trends Neurosci.* 29 (7), 367–373. <https://doi.org/10.1016/j.tins.2006.05.011>.
- Dehaene-Lambertz, G., Montavont, A., Jobert, A., Alliro, L., Dubois, J., Hertz-Pannier, L., Dehaene, S., 2010. Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain Lang.* 114 (2), 53–65. <https://doi.org/10.1016/j.bandl.2009.09.003>.
- Didoné, D.D., Oliveira, L.S., Durante, A.S., Almeida, K. de, Garcia, M.V., Riesgo, R. dos S., Sleifer, P., 2021. Cortical auditory-evoked potential as a biomarker of central auditory maturation in term and preterm infants during the first 3 months. *Clinics* 76, e2944. <https://doi.org/10.6061/clinics/2021/e2944>.
- Dubois, J., Benders, M., Borradori-Tolsa, C., Cachia, A., Lazeyras, F., Ha-Vinh Leuchter, R., Sizonenko, S.V., Warfield, S.K., Mangin, J.F., Hüppi, P.S., 2008. Primary cortical folding in the human newborn: an early marker of later functional development. *Brain* 131 (8), 2028–2041. <https://doi.org/10.1093/brain/awn137>.
- Durante, A.S., Mariano, S., Pachi, P.R., 2018. Auditory processing abilities in prematurely born children. *Early Hum. Dev.* 120, 26–30. <https://doi.org/10.1016/j.earlhumdev.2018.03.011>.
- Durston, S., Davidson, M.C., Tottenham, N., Galvan, A., Spicer, J., Fossella, J.A., Casey, B.J., 2006. A shift from diffuse to focal cortical activity with development. *Dev. Sci.* 9 (1), 1–8. <https://doi.org/10.1111/j.1467-7687.2005.00454.x>.
- Eggermont, J.J., Moore, J.K., 2012. Morphological and functional development of the auditory nervous system. In: *Human Auditory Development*. Springer, pp. 61–105.
- Fecher, N., Johnson, E.K., 2018a. Effects of language experience and task demands on talker recognition by children and adults. *J. Acoust. Soc. Am.* 143 (4), 2409–2418. <https://doi.org/10.1121/1.5032199>.
- Fecher, N., Johnson, E.K., 2018b. The native-language benefit for talker identification is robust in 7.5-month-old infants. *J. Exp. Psychol. Learn. Mem. Cognit.* 44 (12), 1911–1920. <https://doi.org/10.1037/xlm0000555>.
- Fecher, N., Johnson, E.K., 2021. Developmental improvements in talker recognition are specific to the native language. *J. Exp. Child Psychol.* 202, 104991. <https://doi.org/10.1016/j.jecp.2020.104991>.
- Fecher, N., Paquette-Smith, M., Johnson, E.K., 2019. Resolving the (apparent) talker recognition paradox in developmental speech perception. *Infancy* 24 (4), 570–588. <https://doi.org/10.1111/infa.12290>.
- Fecteau, S., Armony, J.L., Joanette, Y., Belin, P., 2004. Is voice processing species-specific in human auditory cortex? An fMRI study. *Neuroimage* 23 (3), 840–848. <https://doi.org/10.1016/j.neuroimage.2004.09.019>.
- Fifer, W., Moon, C., 1994. The role of mother's voice in the organization of brain function in the newborn. *Acta Paediatr.* 83 (s397), 86–93. <https://doi.org/10.1111/j.1651-2227.1994.tb13270.x>.
- Filippa, M., Benis, D., Adam-Darqac, A., Grandjean, D., Hüppi, P.S., 2023. Preterm Infants Show an Atypical Processing of the Mother's Voice. <https://doi.org/10.1101/2022.04.26.489394> (p. 2022.04.26.489394). bioRxiv.
- Filippa, M., Panza, C., Ferrari, F., Frassoldati, R., Kuhn, P., Balduzzi, S., D'Amico, R., 2017. Systematic review of maternal voice interventions demonstrates increased stability in preterm infants. *Acta Paediatr.* 106 (8), 1220–1229. <https://doi.org/10.1111/apa.13832>.
- Fleming, D., Giordano, B.L., Caldara, R., Belin, P., 2014. A language-familiarity effect for speaker discrimination without comprehension. *Proc. Natl. Acad. Sci. USA* 111 (38), 13795–13798. <https://doi.org/10.1073/pnas.1401383111>.
- Flinker, A., Doyle, W.K., Mehta, A.D., Devinsky, O., Poeppel, D., 2019. Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. *Nat. Human Behav.* 3 (4), 393–405. <https://doi.org/10.1038/s41562-019-0548-z>.
- Floccia, C., Nazzi, T., Bertoncini, J., 2000. Unfamiliar voice discrimination for short stimuli in newborns. *Dev. Sci.* 3 (3), 333–343. <https://doi.org/10.1111/1467-7687.00128>.
- Fontaine, M., Love, S.A., Latinus, M., 2017. Familiarity and voice representation: from acoustic-based representation to voice averages. *Front. Psychol.* 8, 1180. <https://doi.org/10.3389/fpsyg.2017.01180>.
- Fox, N.A., 1991. If it's not left, it's right: electroencephalograph asymmetry and the development of emotion. *Am. Psychol.* 46, 863–872. <https://doi.org/10.1037/0003-066X.46.8.863>.
- Friendly, R.H., Rendall, D., Trainor, L.J., 2014. Learning to differentiate individuals by their voices: infants' individuation of native- and foreign-species voices. *Dev. Psychobiol.* 56 (2), 228–237. <https://doi.org/10.1002/dev.21164>.
- Gainotti, G., Quaranta, D., Luzzi, S., 2023. Apperceptive and associative forms of phonagnosia. *Curr. Neurol. Neurosci. Rep.* 23 (6), 327–333. <https://doi.org/10.1007/s11910-023-01271-5>.
- Galván, A., 2010. Neural plasticity of development and learning. *Hum. Brain Mapp.* 31 (6), 879–890. <https://doi.org/10.1002/hbm.21029>.
- Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J.R., Schweinberger, S.R., Warren, J.D., Duchaine, B., 2009. Developmental phonagnosia: a selective deficit of vocal identity recognition. *Neuropsychologia* 47 (1). <https://doi.org/10.1016/j.neuropsychologia.2008.08.003>. Article 1.
- Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P., Gore, J.C., 1999. Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nat. Neurosci.* 2 (6). <https://doi.org/10.1038/9224>. Article 6.
- Gervais, H., Belin, P., Boddaert, N., Leboyer, M., Coez, A., Sfaello, I., Barthélémy, C., Brunelle, F., Samson, Y., Zilbovicius, M., 2004. Abnormal cortical voice processing in autism. *Nat. Neurosci.* 7 (8). <https://doi.org/10.1038/nn1291>. Article 8.
- Ghio, M., Cara, C., Tettamanti, M., 2021. The prenatal brain readiness for speech processing: a review on foetal development of auditory and primordial language networks. *Neurosci. Biobehav. Rev.* 128, 709–719. <https://doi.org/10.1016/j.neubiorev.2021.07.009>.
- Gil-da-Costa, R., Braun, A., Lopes, M., Hauser, M.D., Carson, R.E., Herscovitch, P., Martin, A., 2004. Toward an evolutionary perspective on conceptual representation: species-specific calls activate visual and affective processing systems in the macaque. *Proc. Natl. Acad. Sci. USA* 101 (50), 17516–17521. <https://doi.org/10.1073/pnas.0408077101>.
- Goggin, J.P., Thompson, C.P., Strube, G., Simental, L.R., 1991. The role of language familiarity in voice identification. *Mem. Cognit.* 19 (5), 448–458. <https://doi.org/10.3758/BF03199567>.
- Gogtay, N., Giedd, J.N., Lusk, L., Hayashi, K.M., Greenstein, D., Vaituzis, A.C., Nugent, T. F., Herman, D.H., Clasen, L.S., Toga, A.W., Rapoport, J.L., Thompson, P.M., 2004. Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl. Acad. Sci. USA* 101 (21), 8174–8179. <https://doi.org/10.1073/pnas.0402680101>.
- Goldberg, E., McKenzie, C.A., de Vrijer, B., Eagleson, R., de Ribaupierre, S., 2020. Fetal response to a maternal internal auditory stimulus. *J. Magn. Reson. Imag.* 52 (1), 139–145. <https://doi.org/10.1002/jmri.27033>.
- Granier-Deferre, C., Ribeiro, A., Jacquet, A.-Y., Bassereau, S., 2011. Near-term fetuses process temporal features of speech. *Dev. Sci.* 14 (2), 336–352. <https://doi.org/10.1111/j.1467-7687.2010.00978.x>.
- Greenough, W.T., Black, J.E., Wallace, C.S., 1987. Experience and brain development. *Child Dev.* 58 (3), 539–559.
- Groen, W.B., van Orsouw, L., Zwiers, M., Swinkels, S., van der Gaag, R.J., Buitelaar, J.K., 2008. Gender in voice perception in autism. *J. Autism Dev. Disord.* 38 (10), 1819–1826. <https://doi.org/10.1007/s10803-008-0572-8>.
- Gros-Louis, J., West, M.J., King, A.P., 2014. Maternal responsiveness and the development of directed vocalizing in social interactions. *Infancy* 19 (4), 385–408. <https://doi.org/10.1111/infa.12054>.
- Grossmann, T., Oberecker, R., Koch, S.P., Friederici, A.D., 2010. The developmental origins of voice processing in the human brain. *Neuron* 65 (6), 852–858. <https://doi.org/10.1016/j.neuron.2010.03.001>.
- Heald, S., Nusbaum, H., 2014. Speech perception as an active cognitive process. *Front. Syst. Neurosci.* 8. <https://www.frontiersin.org/articles/10.3389/fnsys.2014.00035>.
- Holmes, E., Domingo, Y., Johnsrude, I.S., 2018. Familiar voices are more intelligible, even if they are not recognized as familiar. *Psychol. Sci.* 29 (10), 1575–1583. <https://doi.org/10.1177/0956797618779083>.
- Holmes, E., Johnsrude, I.S., 2020. Speech spoken by familiar people is more resistant to interference by linguistically similar speech. *J. Exp. Psychol. Learn. Mem. Cognit.* 46 (8), 1465–1476. <https://doi.org/10.1037/xlm0000823>.
- Huttenlocher, P.R., Dabholkar, A.S., 1997. Regional differences in synaptogenesis in human cerebral cortex. *J. Comp. Neurol.* 387 (2), 167–178. [https://doi.org/10.1002/\(SICI\)1096-9861\(19971020\)387:2<167::AID-CNEI>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1096-9861(19971020)387:2<167::AID-CNEI>3.0.CO;2-Z).
- Hykin, J., Moore, R., Duncan, K., Clare, S., Baker, P., Johnson, I., Bowtell, R., Mansfield, P., Gowland, P., 1999. Fetal brain activity demonstrated by functional magnetic resonance imaging. *Lancet* 354 (9179), 645–646. [https://doi.org/10.1016/S0140-6736\(99\)02901-3](https://doi.org/10.1016/S0140-6736(99)02901-3).

- Innsley, S.J., 2000. Long-term vocal recognition in the northern Fur seal. *Nature* 406 (6794). <https://doi.org/10.1038/35019064>. Article 6794.
- Jardri, R., Houfflin-Debarge, V., Delion, P., Pruvo, J.-P., Thomas, P., Pins, D., 2012. Assessing fetal response to maternal speech using a noninvasive functional brain imaging technique. *Int. J. Dev. Neurosci.* 30 (2), 159–161. <https://doi.org/10.1016/j.ijdevneu.2011.11.002>.
- Jardri, R., Pins, D., Houfflin-Debarge, V., Chaffiotte, C., Rocourt, N., Pruvo, J.-P., Steinling, M., Delion, P., Thomas, P., 2008. Fetal cortical activation to sound at 33 weeks of gestation: a functional MRI study. *Neuroimage* 42 (1), 10–18. <https://doi.org/10.1016/j.neuroimage.2008.04.247>.
- Jeffries, E., 2015. Pre-school children's identification of familiar speakers and the role of accent features. *York Papers in Linguistics Series 2* (14), 15–40.
- Johnson, E.K., Westrek, E., Nazzi, T., Cutler, A., 2011. Infant ability to tell voices apart rests on language experience. *Dev. Sci.* 14 (5), 1002–1011. <https://doi.org/10.1111/j.1467-7687.2011.01052.x>.
- Johnsrude, I.S., Mackey, A., Hakymez, H., Alexander, E., Trang, H.P., Carlyon, R.P., 2013. Swinging at a cocktail party: voice familiarity aids speech perception in the presence of a competing voice. *Psychol. Sci.* 24 (10), 1995–2004. <https://doi.org/10.1177/0956797613482467>.
- Joly, O., Pallier, C., Ramus, F., Pressnitzer, D., Vanduffel, W., Orban, G.A., 2012a. Processing of vocalizations in humans and monkeys: a comparative fMRI study. *Neuroimage* 62 (3), 1376–1389. <https://doi.org/10.1016/j.neuroimage.2012.05.070>.
- Joly, O., Ramus, F., Pressnitzer, D., Vanduffel, W., Orban, G.A., 2012b. Interhemispheric differences in auditory processing revealed by fMRI in awake rhesus monkeys. *Cerebr. Cortex* 22 (4), 838–853. <https://doi.org/10.1093/cercor/bhr150>.
- Kanber, E., Lavan, N., McGettigan, C., 2022. Highly accurate and robust identity perception from personally familiar voices. *J. Exp. Psychol. Gen.* 151 (4), 897–911. <https://doi.org/10.1037/xge0001112>.
- Kingstone, A., Tipper, C., Ristic, J., Ngan, E., 2004. The eyes have it!: an fMRI investigation. *Brain Cognit.* 55 (2), 269–271. <https://doi.org/10.1016/j.bandc.2004.02.037>.
- Kisilevsky, B.S., Hains, S.M.J., Brown, C.A., Lee, C.T., Cowperthwaite, B., Stutzman, S.S., Swansburg, M.L., Lee, K., Xie, X., Huang, H., Ye, H.-H., Zhang, K., Wang, Z., 2009. Fetal sensitivity to properties of maternal speech and language. *Infant Behav. Dev.* 32 (1), 59–71. <https://doi.org/10.1016/j.infbeh.2008.10.002>.
- Kisilevsky, B.S., Hains, S.M.J., Lee, K., Xie, X., Huang, H., Ye, H.H., Zhang, K., Wang, Z., 2003. Effects of experience on fetal voice recognition. *Psychol. Sci.* 14 (3), 220–224. <https://doi.org/10.1111/1467-9280.02435>.
- Klin, A., 1991. Young autistic children's listening preferences in regard to speech: a possible characterization of the symptom of social withdrawal. *J. Autism Dev. Disord.* 21 (1), 29–42. <https://doi.org/10.1007/BF02206995>.
- Kojima, S., Izumi, A., Ceugniet, M., 2003. Identification of vocalizers by pant hoots, pant grunts and screams in a chimpanzee. *Primates* 44 (3), 225–230. <https://doi.org/10.1007/s10329-002-0014-8>.
- Kreiman, J., Park, S.J., Keating, P., Alwan, A., 2015. The relationship between acoustic and perceived intraspeaker variability in voice quality. In: *Sixteenth Annual Conference of the International Speech Communication Association*.
- Kreiman, J., Sidtis, D., 2011. Voices and listeners: toward a model of voice perception. *Acoust. Today* 7 (4), 7–15.
- Kriegstein, K.V., Giraud, A.-L., 2004. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage* 22 (2), 948–955. <https://doi.org/10.1016/j.neuroimage.2004.02.020>.
- Kuhl, P.K., Coffey-Corina, S., Padden, D., Dawson, G., 2005. Links between social and linguistic processing of speech in preschool children with autism: behavioral and electrophysiological measures. *Dev. Sci.* 8 (1), F1–F12. <https://doi.org/10.1111/j.1467-7687.2004.00384.x>.
- Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., Lindblom, B., 1992. Linguistic experience alters phonetic perception in infants by 6 Months of age. *Science* 255 (5044), 606–608.
- Kumar, N., Kamath, S., Kumar, G., Vaishali, K., Sinha, M.K., Amin, R., Chamallamudi, M. R., 2023. Prenatal learning and memory: review on the impact of exposure. *Curr. Pediatr. Rev.* 19 (2), 108–120. <https://doi.org/10.2174/1573396318666220601160537>.
- Latinus, M., Belin, P., 2011. Anti-voice adaptation suggests prototype-based coding of voice identity. *Front. Psychol.* 2. <https://www.frontiersin.org/articles/10.3389/fpsyg.2011.00175>.
- Latinus, M., McAleer, P., Bestelmeyer, P.E.G., Belin, P., 2013. Norm-based coding of voice identity in human auditory cortex. *Curr. Biol.* 23 (12), 1075–1080. <https://doi.org/10.1016/j.cub.2013.04.055>.
- Lavan, N., Knight, S., Hazan, V., McGettigan, C., 2019a. The effects of high variability training on voice identity learning. *Cognition* 193, 104026. <https://doi.org/10.1016/j.cognition.2019.104026>.
- Lavan, N., Knight, S., McGettigan, C., 2019b. Listeners form average-based representations of individual voice identities. *Nat. Commun.* 10 (1) <https://doi.org/10.1038/s41467-019-10295-w>. Article 1.
- Lavan, N., McGettigan, C., 2023. A model for person perception from familiar and unfamiliar voices. *Communications Psychology* 1 (1). <https://doi.org/10.1038/s44271-023-00001-4>. Article 1.
- Lavner, Y., Rosenhouse, J., Gath, I., 2001. The prototype model in speaker identification by human listeners. *Int. J. Speech Technol.* 4 (1), 63–74. <https://doi.org/10.1023/A:1009656816383>.
- Leaver, A.M., Rauschecker, J.P., 2010. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J. Neurosci.* 30 (22), 7604–7612. <https://doi.org/10.1523/JNEUROSCI.0296-10.2010>.
- Lecanuet, J.P., Granier-Deferre, C., DeCasper, A.J., Maugeais, R., Andrieu, A.J., Busnel, M.C., 1987. Fetal perception and discrimination of speech stimuli; demonstration by cardiac reactivity; preliminary results. *Comptes rendus de l'Académie des sciences. Serie III, Sciences de la vie* 305 (5), 161–164.
- Lecanuet, J.-P., Granier-Deferre, C., Jacquet, A.-Y., Capponi, I., Ledru, L., 1993. Prenatal discrimination of a male and a female voice uttering the same sentence. *Early Dev. Parent.* 2 (4), 217–228. <https://doi.org/10.1002/edp.2430020405>.
- Lee, G.Y., Kisilevsky, B.S., 2014. Fetuses respond to father's voice but prefer mother's voice after birth. *Dev. Psychobiol.* 56 (1), 1–11. <https://doi.org/10.1002/dev.21084>.
- Leech, R., Holt, L.L., Devlin, J.T., Dick, F., 2009. Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. *J. Neurosci.* 29 (16), 5234–5239. <https://doi.org/10.1523/JNEUROSCI.5758-08.2009>.
- Leopold, D.A., O'Toole, A.J., Vetter, T., Blanz, V., 2001. Prototype-referenced shape encoding revealed by high-level aftereffects. *Nat. Neurosci.* 4 (1) <https://doi.org/10.1038/82947>. Article 1.
- Leroy, F., Cai, Q., Bogart, S.L., Dubois, J., Coulon, O., Monzalvo, K., Fischer, C., Glasel, H., Van der Haegen, L., Bénézit, A., Lin, C.-P., Kennedy, D.N., Ihara, A.S., Hertz-Pannier, L., Moutard, M.-L., Poupon, C., Brysbaert, M., Roberts, N., Hopkins, W.D., et al., 2015. New human-specific brain landmark: the depth asymmetry of superior temporal sulcus. *Proc. Natl. Acad. Sci. USA* 112 (4), 1208–1213. <https://doi.org/10.1073/pnas.1412389112>.
- Levänen, S., Uutela, K., Salenius, S., Hari, R., 2001. Cortical representation of sign language: comparison of deaf signers and hearing non-signers. *Cerebr. Cortex* 11 (6), 506–512. <https://doi.org/10.1093/cercor/11.6.506>.
- Levi, S.V., 2015. Individual differences in learning talker categories: the role of working memory. *Phonetica* 71 (3), 201–226. <https://doi.org/10.1159/000370160>.
- Levi, S.V., 2018. Another bilingual advantage? Perception of talker-voice information. *Biling. Lang. Cognit.* 21 (3), 523–536. <https://doi.org/10.1017/S1366728917000153>.
- Levi, S.V., Schwartz, R.G., 2013. The development of language-specific and language-independent talker processing. *J. Speech Lang. Hear. Res.* 56 (3), 913–925. [https://doi.org/10.1044/1092-4388\(2012\)12-0095](https://doi.org/10.1044/1092-4388(2012)12-0095).
- Ley, A., Vroomen, J., Hausfeld, L., Valente, G., Weerd, P.D., Formisano, E., 2012. Learning of new sound categories shapes neural response patterns in human auditory cortex. *J. Neurosci.* 32 (38), 13273–13280. <https://doi.org/10.1523/JNEUROSCI.0584-12.2012>.
- Liebenthal, E., Desai, R., Ellingson, M.M., Ramachandran, B., Desai, A., Binder, J.R., 2010. Specialization along the left superior temporal sulcus for auditory categorization. *Cerebr. Cortex* 20 (12), 2958–2970. <https://doi.org/10.1093/cercor/bhq045>.
- Lin, I.-F., Agus, T.R., Sued, C., Pressnitzer, D., Yamada, T., Komine, Y., Kato, N., Kashino, M., 2016. Fast response to human voices in autism. *Sci. Rep.* 6 (1), 1–7.
- Lin, I.-F., Yamada, T., Komine, Y., Kato, N., Kato, M., Kashino, M., 2015. Vocal identity recognition in autism spectrum disorder. *PLoS One* 10 (6), e0129451. <https://doi.org/10.1371/journal.pone.0129451>.
- Liu, P., Cole, P.M., Gilmore, R.O., Pérez-Edgar, K.E., Vigeant, M.C., Moriarty, P., Scherf, K.S., 2019. Young children's neural processing of their mother's voice: an fMRI study. *Neuropsychologia* 122, 11–19. <https://doi.org/10.1016/j.neuropsychologia.2018.12.003>.
- Lloyd-Fox, S., Blasi, A., Mercure, E., Elwell, C.E., Johnson, M.H., 2012. The emergence of cerebral specialization for the human voice over the first months of life. *Soc. Neurosci.* 7 (3), 317–330. <https://doi.org/10.1080/17470919.2011.614696>.
- Lombardo, M.V., Pierce, K., Eyster, L.T., Carter Barnes, C., Ahrens-Barbeau, C., Solso, S., Campbell, K., Courchesne, E., 2015. Different functional neural substrates for good and poor language outcome in autism. *Neuron* 86 (2), 567–577. <https://doi.org/10.1016/j.neuron.2015.03.023>.
- Luthra, S., 2021. The role of the right hemisphere in processing phonetic variability between talkers. *Neurobiology of Language* 2 (1), 138–151. https://doi.org/10.1162/nol_a_00028.
- Maguinness, C., Roswandowitz, C., von Kriegstein, K., 2018. Understanding the mechanisms of familiar voice-identity recognition in the human brain. *Neuropsychologia* 116, 179–193. <https://doi.org/10.1016/j.neuropsychologia.2018.03.039>.
- Mahmoudzadeh, M., Dehaene-Lambertz, G., Fournier, M., Kongolo, G., Goudjil, S., Dubois, J., Grebe, R., Wallois, F., 2013. Syllabic discrimination in premature human infants prior to complete formation of cortical layers. *Proc. Natl. Acad. Sci. USA* 110 (12), 4846–4851. <https://doi.org/10.1073/pnas.1212220110>.
- Mann, V.A., Diamond, R., Carey, S., 1979. Development of voice recognition: parallels with face recognition. *J. Exp. Child Psychol.* 27 (1), 153–165. [https://doi.org/10.1016/0022-0965\(79\)90067-5](https://doi.org/10.1016/0022-0965(79)90067-5).
- Mathias, S.R., von Kriegstein, K., 2019. Voice processing and voice-identity recognition. In: *Siedenburg, K., Saitis, C., McAdams, S., Popper, A.N., Fay, R.R. (Eds.), Timbre: Acoustics, Perception, and Cognition*. Springer International Publishing, pp. 175–209. https://doi.org/10.1007/978-3-030-14832-4_7.
- May, L., Byers-Heinlein, K., Gervain, J., Werker, J., 2011. Language and the newborn brain: does prenatal language experience shape the neonate neural response to speech? *Front. Psychol.* 2. <https://www.frontiersin.org/articles/10.3389/fpsyg.2011.00222>.
- McAleer, P., Todorov, A., Belin, P., 2014. How do you say 'hello'? Personality impressions from brief novel voices. *PLoS One* 9 (3), e90779. <https://doi.org/10.1371/journal.pone.0090779>.
- McCormick, M.C., Litt, J.S., Smith, V.C., Zupancic, J.A.F., 2011. Prematurity: an overview and public health implications. *Annu. Rev. Publ. Health* 32 (1), 367–379. <https://doi.org/10.1146/annurev-publhealth-090810-182459>.
- McDonald, N.M., Perdue, K.L., Eilbott, J., Loyal, J., Shic, F., Pelphrey, K.A., 2019. Infant brain responses to social sounds: a longitudinal functional near-infrared

- spectroscopy study. *Developmental Cognitive Neuroscience* 36, 100638. <https://doi.org/10.1016/j.dcn.2019.100638>.
- McMahon, E., Wintermark, P., Lahav, A., 2012. Auditory brain development in premature infants: the importance of early experience. *Ann. N. Y. Acad. Sci.* 1252 (1), 17–24. <https://doi.org/10.1111/j.1749-6632.2012.06445.x>.
- Minagawa-Kawai, Y., van der Lely, H., Ramus, F., Sato, Y., Mazuka, R., Dupoux, E., 2011. Optical brain imaging reveals general auditory and language-specific processing in early infant development. *Cerebr. Cortex* 21 (2), 254–261. <https://doi.org/10.1093/cercor/bhq082>.
- Moon, C., Lagercrantz, H., Kuhl, P.K., 2013. Language experienced in utero affects vowel perception after birth: a two-country study. *Acta Paediatr.* 102 (2), 156–160. <https://doi.org/10.1111/apa.12098>.
- Moore, J.K., Guan, Y.L., 2001. Cytoarchitectural and axonal maturation in human auditory cortex. *Journal of the Association for Research in Otolaryngology* 2 (4), 297–311.
- Moore, J.K., Linthicum, F.H., 2007. The human auditory system: a timeline of development. *Int. J. Audiol.* 46 (9), 460–478. <https://doi.org/10.1080/14992020701383019>.
- Nagels, C., Gaudrain, E., Vickers, D., Hendriks, P., Başkent, D., 2020. Development of voice perception is dissociated across gender cues in school-age children. *Sci. Rep.* 10 (1) <https://doi.org/10.1038/s41598-020-61732-6>. Article 1.
- Neuner, F., Schweinberger, S.R., 2000. Neuropsychological impairments in the recognition of faces, voices, and personal names. *Brain Cognit.* 44 (3) <https://doi.org/10.1006/brcg.1999.1196>. Article 3.
- Nittrouer, S., Manning, C., Meyer, G., 1993. The perceptual weighting of acoustic cues changes with linguistic experience. *J. Acoust. Soc. Am.* 94 (3 Suppl. ment), 1865. <https://doi.org/10.1121/1.407649>.
- Norman-Haignere, S.V., McDermott, J.H., 2018. Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex. *PLoS Biol.* 16 (12), e2005127 <https://doi.org/10.1371/journal.pbio.2005127>.
- Nygaard, L.C., Pisoni, D.B., 1998. Talker-specific learning in speech perception. *Percept. Psychophys.* 60 (3), 355–376. <https://doi.org/10.3758/BF03206860>.
- Ohnishi, T., Matsuda, H., Hashimoto, T., Kunihiro, T., Nishikawa, M., Uema, T., Sasaki, M., 2000. Abnormal regional cerebral blood flow in childhood autism. *Brain* 123 (9), 1838–1844. <https://doi.org/10.1093/brain/123.9.1838>.
- Ortiz-Rios, M., Kuśmierk, P., DeWitt, L., Archakov, D., Azevedo, F.A.C., Sams, M., Jääskeläinen, I.P., Keliris, G.A., Rauschecker, J.P., 2015. Functional MRI of the vocalization-processing network in the macaque brain. *Front. Neurosci.* 9. <https://www.frontiersin.org/articles/10.3389/fnins.2015.00113>.
- Pascalis, O., de Haan, M., Nelson, C.A., 2002. Is face processing species-specific during the first year of life? *Science* 296 (5571), 1321–1323. <https://doi.org/10.1126/science.1070223>.
- Peña, M., Werker, J.F., Dehaene-Lambertz, G., 2012. Earlier speech exposure does not accelerate speech acquisition. *J. Neurosci.* 32 (33), 11159–11163. <https://doi.org/10.1523/JNEUROSCI.6516-11.2012>.
- Pépiot, E., 2015. Voice, speech and gender: male-female acoustic differences and cross-language variation in English and French speakers. CORELA - COGNITION, REPRÉSENTATION, LANGAGE. <https://doi.org/10.4000/corela.3783>. HS-16.
- Perea, M., Jiménez, M., Suárez-Coalla, P., Fernández, N., Viña, C., Cuetos, F., 2014. Ability for voice recognition is a marker for dyslexia in children. *Exp. Psychol.* 61 (6), 480–487. <https://doi.org/10.1027/1618-3169/a000265>.
- Pernet, C.R., McAleer, P., Latinus, M., Gorgolewski, K.J., Charest, I., Bestelmeyer, P.E.G., Watson, R.H., Fleming, D., Crabbe, F., Valdes-Sosa, M., Belin, P., 2015. The human voice areas: spatial organization and inter-individual variability in temporal and extra-temporal cortices. *Neuroimage* 119, 164–174. <https://doi.org/10.1016/j.neuroimage.2015.06.050>.
- Perrachione, T.K., Dougherty, S.C., McLaughlin, D.E., Lember, R.A., 2015. The Effects of Speech Perception and Speech Comprehension on Talker Identification. *International Congress of Phonetic Sciences*.
- Perrachione, T.K., Pierrehumbert, J.B., Wong, P.C.M., 2009. Differential neural contributions to native- and foreign-language talker identification. *J. Exp. Psychol. Hum. Percept. Perform.* 35 (6), 1950–1960. <https://doi.org/10.1037/a0015869>.
- Perrachione, T.K., Wong, P.C.M., 2007. Learning to recognize speakers of a non-native language: implications for the functional organization of human auditory cortex. *Neuropsychologia* 45 (8), 1899–1910. <https://doi.org/10.1016/j.neuropsychologia.2006.11.015>.
- Petkov, C.I., Kayser, C., Studel, T., Whittingstall, K., Augath, M., Logothetis, N.K., 2008. A voice region in the monkey brain. *Nat. Neurosci.* 11 (3) <https://doi.org/10.1038/nn2043>. Article 3.
- Petkov, C.I., Logothetis, N.K., Obleser, J., 2009. Where are the human speech and voice regions, and do other animals have anything like them? *Neuroscientist* 15 (5), 419–429.
- Petrini, K., Tagliapietra, S., 2008. Cognitive maturation and the use of pitch and rate information in making similarity judgments of a single talker. *J. Speech Lang. Hear. Res.* 51 (2), 485–501. [https://doi.org/10.1044/1092-4388\(2008\)035](https://doi.org/10.1044/1092-4388(2008)035).
- Raschle, N.M., Smith, S.A., Zuk, J., Dauvermann, M.R., Figuccio, M.J., Gaab, N., 2014. Investigating the neural correlates of voice versus speech-sound directed information in pre-school children. *PLoS One* 9 (12), e115549. <https://doi.org/10.1371/journal.pone.0115549>.
- Redcay, E., 2008. The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. *Neurosci. Biobehav. Rev.* 32 (1), 123–142. <https://doi.org/10.1016/j.neubiorev.2007.06.004>.
- Rendall, D., Rodman, P.S., Emond, R.E., 1996. Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Anim. Behav.* 51 (5), 1007–1015. <https://doi.org/10.1006/anbe.1996.0103>.
- Rhodes, G., Jeffery, L., 2006. Adaptive norm-based coding of facial identity. *Vis. Res.* 46 (18), 2977–2987. <https://doi.org/10.1016/j.visres.2006.03.002>.
- Rose, J., Flaherty, M., Browning, J., Leibold, L.J., Buss, E., 2018. Pure-tone frequency discrimination in preschoolers, young school-age children, and adults. *J. Speech Lang. Hear. Res.* 61 (9), 2440–2445. https://doi.org/10.1044/2018_JSLHR-H-17-0445.
- Roswandowitz, C., Mathias, S.R., Hintz, F., Kreitewolf, J., Schelinski, S., von Kriegstein, K., 2014. Two cases of selective developmental voice-recognition impairments. *Curr. Biol.* 24 (19), 2348–2353. <https://doi.org/10.1016/j.cub.2014.08.048>.
- Roswandowitz, C., Schelinski, S., von Kriegstein, K., 2017. Developmental phonagnosia: linking neural mechanisms with the behavioural phenotype. *Neuroimage* 155, 97–112. <https://doi.org/10.1016/j.neuroimage.2017.02.064>.
- Rubenstein, A.J., Kalakanis, L., Langlois, J.H., 1999. Infant preferences for attractive faces: a cognitive explanation. *Dev. Psychol.* 35 (3), 848–855. <https://doi.org/10.1037/0012-1649.35.3.848>.
- Rupp, K., Hect, J.L., Remick, M., Ghuman, A., Chandrasekaran, B., Holt, L.L., Abel, T.J., 2022. Neural responses in human superior temporal cortex support coding of voice representations. *PLoS Biol.* 20 (7), e3001675 <https://doi.org/10.1371/journal.pbio.3001675>.
- Sadato, N., Yamada, H., Okada, T., Yoshida, M., Hasegawa, T., Matsuki, K.-I., Yonekura, Y., Itoh, H., 2004. Age-dependent plasticity in the superior temporal sulcus in deaf humans: a functional MRI study. *BMC Neurosci.* 5 (1), 56. <https://doi.org/10.1186/1471-2202-5-56>.
- Schall, S., Kiebel, S.J., Maess, B., von Kriegstein, K., 2015. Voice identity recognition: functional division of the right STS and its behavioral relevance. *J. Cognit. Neurosci.* 27 (2), 280–291. https://doi.org/10.1162/jocn_a_00707.
- Schelinski, S., Borowiak, K., von Kriegstein, K., 2016. Temporal voice areas exist in autism spectrum disorder but are dysfunctional for voice identity recognition. *Soc. Cognit. Affect Neurosci.* 11 (11), 1812–1822. <https://doi.org/10.1093/scan/nsw089>.
- Schönwiesner, M., Rübsamen, R., Von Cramon, D.Y., 2005. Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *Eur. J. Neurosci.* 22 (6), 1521–1528. <https://doi.org/10.1111/j.1460-9568.2005.04315.x>.
- Schweinberger, S.R., Kawahara, H., Simpson, A.P., Skuk, V.G., Zäske, R., 2014. Speaker perception. *WIREs Cognitive Science* 5 (1), 15–25. <https://doi.org/10.1002/wcs.1261>.
- Seyfarth, R.M., Cheney, D.L., 2003. Signalers and receivers in animal communication. *Annu. Rev. Psychol.* 54 (1), 145–173. <https://doi.org/10.1146/annurev.psych.54.101601.145121>.
- Shilowich, B.E., Biederman, I., 2016. An estimate of the prevalence of developmental phonagnosia. *Brain Lang.* 159, 84–91. <https://doi.org/10.1016/j.bandl.2016.05.004>.
- Shultz, S., Vouloumanos, A., Bennett, R.H., Pelphrey, K., 2014. Neural specialization for speech in the first months of life. *Dev. Sci.* 17 (5), 766–774. <https://doi.org/10.1111/desc.12151>.
- Sidtis, D., Kreiman, J., 2012. In the beginning was the familiar voice: personally familiar voices in the evolutionary and contemporary biology of communication. *Integr. Psychol. Behav. Sci.* 46 (2), 146–159. <https://doi.org/10.1007/s12124-011-9177-4>.
- Simon, S., Lazeyras, F., Sigrist, A.-D., Ecoffey, M., Guatieri, S., Van de Ville, D., Borradori-Tolsa, C., Pelizzone, M., Hppi, P., 2009. Nature vs nurture in newborn voice perception. An fMRI comparison of auditory processing between premature infants at term age and term born neonates. In: *International Society for Magnetic Resonance in Medicine. International Society for Magnetic Resonance in Medicine, Honolulu, HI, USA*.
- Spence, M.J., Rollins, P.R., Jerger, S., 2002. Children's recognition of cartoon voices. *J. Speech Lang. Hear. Res.* 45 (1), 214–222. [https://doi.org/10.1044/1092-4388\(2002\)016](https://doi.org/10.1044/1092-4388(2002)016).
- Staeren, N., Renvall, H., De Martino, F., Goebel, R., Formisano, E., 2009. Sound categories are represented as distributed patterns in the human auditory cortex. *Curr. Biol.* 19 (6), 498–502. <https://doi.org/10.1016/j.cub.2009.01.066>.
- Staib, M., Frühholz, S., 2021. Cortical voice processing is grounded in elementary sound analyses for vocalization relevant sound patterns. *Prog. Neurobiol.* 200, 101982. <https://doi.org/10.1016/j.pneurobio.2020.101982>.
- Staib, M., Frühholz, S., 2022. Distinct functional levels of human voice processing in the auditory cortex. *Cerebr. Cortex.* <https://doi.org/10.1093/cercor/bhac128> bhac128.
- Steinmetz, H., 1996. Structure, function and cerebral asymmetry: in vivo morphometry of the planum temporale. *Neurosci. Biobehav. Rev.* 20 (4), 587–591. [https://doi.org/10.1016/0149-7634\(95\)00071-2](https://doi.org/10.1016/0149-7634(95)00071-2).
- Stevenson, S.V., 2018. Drawing a distinction between familiar and unfamiliar voice processing: a review of neuropsychological, clinical and empirical findings. *Neuropsychologia* 116, 162–178. <https://doi.org/10.1016/j.neuropsychologia.2017.07.005>.
- Strömbergsson, S., 2013. Children's recognition of their own recorded voice: influence of age and phonological impairment. *Clin. Linguist. Phon.* 27 (1), 33–45. <https://doi.org/10.3109/02699206.2012.735744>.
- Therien, J.M., Worwa, C.T., Mattia, F.R., deRegnier, R.-A.O., 2004. Altered pathways for auditory discrimination and recognition memory in preterm infants. *Dev. Med. Child Neurol.* 46 (12), 816–824. <https://doi.org/10.1017/S0012162204001434>.
- Thompson, C.P., 1987. A language effect in voice identification. *Appl. Cognit. Psychol.* 1 (2), 121–131. <https://doi.org/10.1002/acp.2350010205>.
- Thompson, P.M., Cannon, T.D., Narr, K.L., van Erp, T., Poutanen, V.-P., Huttunen, M., Lonnqvist, J., Standertskjold-Nordenstam, C.-G., Kaprio, J., Khaledy, M., Dail, R., Zoumalan, C.I., Toga, A.W., 2001. Genetic Influences on Brain Structure | *Nature Neuroscience*. <https://www.nature.com/articles/nn758>.

- Trehub, S.E., 2017. The maternal voice as a special signal for infants. In: Filippa, M., Kuhn, P., Westrup, B. (Eds.), *Early Vocal Contact and Preterm Infant Brain Development: Bridging the Gaps between Research and Practice*. Springer International Publishing, pp. 39–54. https://doi.org/10.1007/978-3-319-65077-7_3.
- Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Best, C., 1994. Developmental changes in perceptual discrimination of non-native speech contrasts by Japanese infants: on discrimination of American English /r, l/ and /w/. In: *The 3rd International Conference on Spoken Language Processing, ICSLP*.
- Turkewitz, G., Kenny, P.A., 1982. Limitations on input as a basis for neural organization and perceptual development: a preliminary theoretical statement. *Dev. Psychobiol.* 15 (4), 357–368. <https://doi.org/10.1002/dev.420150408>.
- Uchida-Ota, M., Arimitsu, T., Tsuzuki, D., Dan, I., Ikeda, K., Takahashi, T., Minagawa, Y., 2019. Maternal speech shapes the cerebral frontotemporal network in neonates: a hemodynamic functional connectivity study. *Developmental Cognitive Neuroscience* 39, 100701. <https://doi.org/10.1016/j.dcn.2019.100701>.
- Van Lancker, D.R., Canter, G.J., 1982. Impairment of voice and face recognition in patients with hemispheric damage. *Brain Cognit.* 1 (2), 185–195. [https://doi.org/10.1016/0278-2626\(82\)90016-1](https://doi.org/10.1016/0278-2626(82)90016-1).
- Van Lancker, D.R., Cummings, J.L., Kreiman, J., Dobkin, B.H., 1988. Phonagnosia: a dissociation between familiar and unfamiliar voices. *Cortex* 24 (2). [https://doi.org/10.1016/S0010-9452\(88\)80029-7](https://doi.org/10.1016/S0010-9452(88)80029-7). Article 2.
- Van Lancker, D.R., Kreiman, J., Cummings, J., 1989. Voice perception deficits: neuroanatomical correlates of phonagnosia. *J. Clin. Exp. Neuropsychol.* 11 (5), 665–674. <https://doi.org/10.1080/01688638908400923>.
- Vogelsang, M., Vogelsang, L., Diamond, S., Sinha, P., 2023. Prenatal auditory experience and its sequelae. *Dev. Sci.* 26 (1), e13278 <https://doi.org/10.1111/desc.13278>.
- von Kriegstein, K., Eger, E., Kleinschmidt, A., Giraud, A.L., 2003. Modulation of neural responses to speech by directing attention to voices or verbal content. *Cognit. Brain Res.* 17 (1), 48–55. [https://doi.org/10.1016/S0926-6410\(03\)00079-X](https://doi.org/10.1016/S0926-6410(03)00079-X).
- Vouloumanos, A., Hauser, M.D., Werker, J.F., Martin, A., 2010. The tuning of human neonates' preference for speech. *Child Dev.* 81 (2), 517–527. <https://doi.org/10.1111/j.1467-8624.2009.01412.x>.
- Vouloumanos, A., Werker, J.F., 2004. Tuned to the signal: the privileged status of speech for young infants. *Dev. Sci.* 7 (3), 270–276.
- Wang, W., Yu, Q., Liang, W., Xu, F., Li, Z., Tang, Y., Liu, S., 2023. Altered cortical microstructure in preterm infants at term-equivalent age relative to term-born neonates. *Cerebr. Cortex* 33 (3), 651–662. <https://doi.org/10.1093/cercor/bhac091>.
- Webb, A.R., Heller, H.T., Benson, C.B., Lahav, A., 2015. Mother's voice and heartbeat sounds elicit auditory plasticity in the human brain before full gestation. *Proc. Natl. Acad. Sci. USA* 112 (10), 3152–3157. <https://doi.org/10.1073/pnas.1414924112>.
- Witelson, S.F., Pallie, W., 1973. Left hemisphere specialization for language in the newborn: neuroanatomical evidence of asymmetry. *Brain* 96, 641–646.
- Xu, X., Biederman, I., Shilowich, B.E., Herald, S.B., Amir, O., Allen, N.E., 2015. Developmental phonagnosia: neural correlates and a behavioral marker. *Brain Lang.* 149, 106–117. <https://doi.org/10.1016/j.bandl.2015.06.007>.
- Yamauchi, A., Imagawa, H., Yokonishi, H., Sakakibara, K.-I., Tayama, N., 2022. Gender- and age- stratified normative voice data in Japanese-speaking subjects: analysis of sustained habitual phonations. *J. Voice.* <https://doi.org/10.1016/j.jvoice.2021.12.002>.
- Zaltz, Y., 2023. The effect of stimulus type and testing method on talker discrimination of school-age children. *J. Acoust. Soc. Am.* 153 (5), 2611. <https://doi.org/10.1121/10.0017999>.
- Zaltz, Y., Goldsworthy, R.L., Eisenberg, L.S., Kishon-Rabin, L., 2020. Children with normal hearing are efficient users of fundamental frequency and vocal tract length cues for voice discrimination. *Ear Hear.* 41 (1), 182–193. <https://doi.org/10.1097/AUD.0000000000000743>.
- Zarate, J.M., Tian, X., Woods, K.J.P., Poeppel, D., 2015. Multiple levels of linguistic and paralinguistic features contribute to voice recognition. *Sci. Rep.* 5 (1) <https://doi.org/10.1038/srep11475>. Article 1.
- Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. *Cerebr. Cortex* 11 (10), 946–953. <https://doi.org/10.1093/cercor/11.10.946>.
- Zhang, Y., Ding, Y., Huang, J., Zhou, W., Ling, Z., Hong, B., Wang, X., 2021. Hierarchical cortical networks of “voice patches” for processing voices in human brain. *Proc. Natl. Acad. Sci. USA* 118 (52), e2113887118. <https://doi.org/10.1073/pnas.2113887118>.
- Zilbovicius, M., Meresse, I., Chabane, N., Brunelle, F., Samson, Y., Boddaert, N., 2006. Autism, the superior temporal sulcus and social perception. *Trends Neurosci.* 29 (7), 359–366. <https://doi.org/10.1016/j.tins.2006.06.004>.