From Fetuses to Newborns: A Continuity Perspective on the Development of Speech Perception*

Wang Menghuan Xu Qiuyang Liang Dandan

2026-07-14

ChinaRxiv

Abstract

Human newborns exhibit high sensitivity to their native language, their mother's voice, and phonetic details within hours after birth, and the developmental origins of these abilities are likely established in utero. From the perspective of developmental neuroscience, this article systematically reviews the low-pass-filtered acoustic environment of the womb and the maturation of the fetal auditory system, with a focus on how prenatal auditory experience contributes to shaping the neural mechanisms of neonatal speech perception. Existing studies indicate that, within the low-frequency-filtered intrauterine soundscape, fetuses can perceive and remember prosodic patterns related to their native language, and these experiences may persist into the early stages of speech processing after birth. After birth, full-spectrum speech input and rapid cortical maturation jointly contribute to the development of neonatal segmental processing abilities. The underlying neural mechanisms involve activation of bilateral temporal–frontal networks, as well as enhanced long-range temporal correlations of neural oscillations matched to native-language prosody and phoneme processing. On this basis, building on the existing perspective of perinatal continuity, this article constructs a prenatal–postnatal continuity framework for the development of speech perception, proposing that neonatal speech perception develops continuously under the joint influence of maturation and functionalization of the auditory system, prenatal low-frequency prosodic experience, and postnatal full-spectrum speech input. Future research should integrate multimodal techniques, further strengthen longitudinal follow-up of high-risk populations such as preterm infants, and expand samples from tonal-language and multilingual environments. At the applied level, it should also explore the potential value of related research in clinical screening and intervention, with the aim of providing a more robust neuroscientific basis for the early identification and precise intervention of language developmental disorders.

Full Text

From Fetuses to Newborns: A Continuity Perspective on the Development of Speech Perception*

Wang Menghuan Xu Qiuyang Liang Dandan

(School of Liberal Arts, Nanjing Normal University, Nanjing 210097)

Abstract Human newborns can, within only a few hours after birth, show high sensitivity to their native language, their mother's voice, and fine-grained speech details; the starting point of this development is very likely already laid down in utero. From the perspective of developmental neuroscience, this article systematically reviews the intrauterine low-pass-filtered acoustic environment and the maturation of the fetal auditory system, with a focus on how prenatal auditory experience participates in shaping the neural mechanisms of newborn speech perception. Existing studies show that, within the intrauterine soundscape shaped by low-frequency filtering, fetuses are able to perceive and remember rhythm patterns related to the maternal language; these experiences may extend into the early postnatal stages of speech processing. After birth, full-spectrum speech input and rapid cortical maturation jointly participate in the development of newborns' phonemic-processing abilities. The neural mechanisms involve activation of bilateral temporo-frontal networks, as well as strengthened long-range temporal correlations of neural oscillations matched to maternal-language rhythm and phonemic processing. On this basis, grounded in the existing perspective of perinatal continuity, this article constructs a prenatal–postnatal continuity framework for the development of speech perception, arguing that newborn speech perception develops continuously under the joint influence of auditory-system maturation and functionalization, prenatal low-frequency rhythmic experience, and postnatal full-spectrum speech input. Future research should integrate multimodal techniques, deepen longitudinal tracking of high-risk populations such as preterm infants, and expand samples from tonal-language and multilingual environments. At the same time, it should explore, at the application level, the potential value of related research for clinical screening and intervention, in order to provide a more solid neuroscientific basis for the early identification of language developmental disorders and for precise intervention.

Keywords prenatal auditory experience; intrauterine acoustic environment; newborn speech perception; prosodic processing; neural mechanisms

1 Introduction

Language development runs through the entire human life cycle and has long been a core issue in developmental science. Among these topics, research on language ability in the earliest stages of life is an important breakthrough point for tracing the starting point of language development and delineating early developmental trajectories. Within a very short time after birth, newborns already display rich and fine-grained capacities for speech perception (for reviews, see Song Xinyan & Meng Xiangzhi, 2012; Chen Yu & Zhang Dandan, 2020; Li Sijin et al., 2023). A large body of behavioral evidence shows that newborns can distinguish the rhythmic structures of their native language and foreign languages (Mehler et al., 1988; Moon et al., 1993), recognize and prefer their mother's voice (DeCasper & Fifer, 1980), process segmental differences in vowels or consonants (Cheour et al., 1998; Moon et al., 2015), and exhibit differentiated behavioral responses to various types of affective prosody (Mastropieri & Turkewitz, 1999). Together, these findings indicate that humans are very likely already excellent language "learners" at the very beginning of life, and that newborns are not "blank

Received: 2025-12-12

* Funded by the 2022 Key Project of the National Social Science Fund of China (22AYY013).
Corresponding author: Liang Dandan, E-mail: 03275@njnu.edu.cn

slate),” but rather come into the world with a brain state that is highly plastic and sensitive to linguistic input.

In addition to behavioral research, neuroimaging studies have provided more direct evidence for neonates’ language-processing abilities. Studies have found that when newborns listen to the rhythm of their native language, their mother’ s voice, or speech materials repeatedly presented during pregnancy, their patterns of brain activity show clear characteristics of familiarity processing (Ma et al., 2025; Ortiz Barajas & Gervain, 2021), and they can produce specific neural responses to emotional prosody (Hou et al., 2024; Zhang et al., 2017, 2019). These neural findings corroborate behavioral evidence, jointly indicating that newborns already possess, at birth, an initial foundation for language-related auditory processing. Why, then, do newborns have such rich capacities for speech perception at birth? Where exactly does this “innate capacity” come from? Is it preset by innate mechanisms, or has it already been shaped by auditory experience before birth? From the perspective of developmental neuroscience, Prechtl (1984) relatively early proposed that neural function shows continuity from the prenatal to the postnatal period, arguing that the organization of neural functions already formed during the fetal period continues and constrains behavioral development after birth. In recent years, this continuity perspective[1] has been further integrated, from a broader perspective on brain and behavioral development, into a perinatal developmental framework, which understands birth as a process in which the organization of neural functions in the fetal period is extended, adjusted, and recalibrated after a sudden change in the external environment (Weiss et al., 2024). Existing research shows that the auditory system begins functioning at around 24 weeks of gestation (Helwany et al., 2025), suggesting that the starting point of newborns’ speech-perception abilities may lie well before birth, while the relevant postnatal responses are a concrete manifestation of this continuity in the domain of speech perception.

Although research on prenatal auditory experience has made some progress, there is still a lack of systematic clarification of its mechanisms of action during the fetal–neonatal transition. Methodological limitations have also, to some extent, hindered deeper exploration of the above issues. Because the fetus is located inside the mother’ s body, it is difficult for researchers to directly and continuously capture the dynamic process of prenatal brain development. Traditional studies have mostly relied on newborns’ behavioral responses after birth, or have used preterm infants as subjects in “natural experiments”

[1] It should be noted that the “continuity” referred to in this article does not mean the continuity issue concerning the origins of linguistic ability and the nature of representation in generative linguistics or constructivist theories of language acquisition; rather, it specifically refers to the developmental continuity of the speech-perception system from the fetal period to the neonatal period.

for indirect inference; however, such approaches are inevitably affected by confounding factors such as differences in developmental stage. In recent years, with the development of fetal functional magnetic resonance imaging (fMRI), fetal magnetoencephalography (fMEG), and ultrasound techniques with high spatiotemporal resolution (Mappa et al., 2024; Calixto et al., 2024), researchers have been able to observe fetal brain neural activity and response patterns at earlier stages, providing new possibilities for directly exploring how prenatal experience shapes the nervous system. From the perspective of developmental neuroscience, this article intends to systematically review and integrate domestic and international research findings on prenatal auditory experience and the development of fetal-neonatal speech perception. It aims to summarize existing evidence, construct a possible mechanistic framework, and propose directions and methodological recommendations for future research, with the goal of promoting deeper development in this field of study.

2 Characteristics and Availability of Prenatal Auditory Experience

2.1 Physical Characteristics of the Intrauterine Acoustic Environment

The auditory experience of the perinatal fetus is mainly shaped by physiological conditions. On the one hand, the acoustic environment inside the uterus plays a decisive role in the effective transmission of external acoustic information; on the other hand, the auditory features that the fetus can process gradually become richer as the auditory system—from the peripheral auditory organs to the central nervous system—matures. The prevailing view is that the uterus and amniotic fluid together constitute a natural "low-pass filtering system": high-frequency sound waves (> 1000 Hz) are substantially attenuated as they pass through maternal soft tissue, the uterine wall, and amniotic fluid, whereas the attenuation of low-frequency sound waves (≤ 250 Hz) can be considered negligible (Gélat et al., 2025). Because the main energy of human speech is concentrated between 300 and 3000 Hz, this acoustic filtering property allows the fetus to stably perceive low-frequency information such as speech-stream rhythm, stress patterns, and prosodic contours. However, for speech features that rely heavily on high-frequency information—such as consonants and voicing contrasts—its perceptual capacity is relatively limited (Anikin et al., 2023).

The main pathways by which sound enters the intrauterine environment are of two kinds. One is the longitudinal transmission pathway, which primarily conducts the mother's speech: when the mother vocalizes, vocal-fold vibrations resonate through the thoracic and abdominal cavities and then pass through the uterine wall into the amniotic fluid. The other is the transverse transmission pathway, which primarily conducts sounds from the external environment: sound is transmitted into the uterus through successive media, including air, the soft tissues of the abdominal wall, and amniotic fluid (Gélat et al., 2019). Compared with transverse transmission, attenuation through the media is lower along the longitudinal pathway, meaning that the mother's speech sounds clearer and more intense inside the uterus. Studies have shown that, on average, the mother' s speech is approximately 20-24 dB stronger than the speech of an unfamiliar person outside the body (Moon & Fifer, 2000). Therefore, for the fetus, maternal speech is not only the loudest, most stable, and most predictable input source, but also the core sound source for the formation of early auditory memory.

2.2 Types of Prenatal Auditory Experience

The acoustic information of language can be roughly divided into two levels. Segments refer to the basic speech units in speech that can be linearly divided; in research on language perception, they are typically associated with correspondences at the phonemic/phoneme level. In contrast, suprasegmental or prosodic features span several segments and operate over a larger temporal window, such as stress, rhythm, intonation, and pitch contour. Under the constraints of intrauterine filtering, acoustic information at these two levels differs in accessibility and processability, thereby influencing fetal speech perception. Specifically, prosodic and rhythmic information both belong to low-frequency fluctuations (mainly in the range of 0.5–20 Hz), and therefore are the easiest to preserve into the intrauterine environment; they are the earliest speech features that the fetus can perceive (Mehler et al., 1988; Goswami, 2019). For example, emotional prosody is basically composed of the fundamental-frequency (F_0) contour, changes in intensity, and rhythmic features. These acoustic cues are located in low-frequency ranges and can penetrate the intrauterine environment relatively completely, allowing the fetus to receive them (Hou et al., 2024). By contrast, for segmental information in the speech stream, its high-frequency components —such as consonantal obstruction noise and plosive bursts—are severely affected by attenuation. Essentially, only low-frequency cues such as the first formant of vowels (F_1) can be retained in utero. This makes it possible for the fetus to form a coarse categorization of the native-language vowel system, but difficult to perceive fine-grained phonemic contrasts (Moon et al., 2013). However, the sequencing and repetition patterns among phonemes are highly dependent on rhyth—

rhythms, and fetuses in late gestation may also initially perceive and learn these structures (Ortiz Barajas & Gervain, 2021). In sum, speech-sound processing in the prenatal period does not have a complete foundation in direct experience, but low-frequency cues such as the vowel F_1 may constitute a limited interface for prenatal experience.

2.3 Developmental course of the fetal auditory system and evidence of auditory learning

The development of the fetal auditory system follows a progressive maturation sequence from the periphery to the central nervous system and from lower to higher levels, broadly proceeding through the stepwise development of peripheral auditory organs, brainstem auditory pathways, the thalamic auditory relay system, and finally auditory regions of the cerebral cortex (Leibovitz et al., 2022; Monson et al., 2018). The development of the nervous system is closely related to the prenatal acoustic environment: the gradual maturation of neural structures builds the physiological basis for auditory processing, while external acoustic stimulation provides experience-driven and regulatory mechanisms for the realization and plasticity of system functions (Lecanuet, 1996). At 24 weeks of gestation, the development of fetal peripheral auditory organs (such as the cochlea and ossicular chain) is approaching completion, and the auditory system begins to become functional (Helwany et al., 2025); the fetus can produce reflexive motor responses to vibration or relatively strong sound pressure (Birnholz & Benacerraf, 1983). At 28–34 weeks of gestation, as the brainstem auditory nuclei and the thalamo-cortical pathways gradually mature, the fetus can distinguish low-frequency sound stimuli of different frequencies and show changes in heart rate. Hepper and Shahidullah (1994) found that fetuses at 28–32 weeks produced stable heart-rate changes in response to pure-tone frequency differences (250 Hz vs. 500 Hz). Meanwhile, at this stage, fetal auditory evoked fields (fAEF) can be recorded by fMEG. Schleussner and Schneider (2004) found that fetuses at around 28 weeks of gestation already showed relatively stable fAEF responses, suggesting that their sound-encoding ability had developed to a central level. By late gestation (35–40 weeks), cortical-level networks are gradually established and begin the process of myelination. Measurable functional connections emerge between the primary auditory cortex and language-related brain regions such as the temporal lobe and inferior frontal gyrus, indicating that the fetal auditory system may begin to enter an advanced processing stage (Taymourtash et al., 2023). During this period, fetuses can not only perceive sounds, but may also neurally encode acoustic features such as speech rhythm and sequential structure; even so, the fine-grained discriminative capacity of the temporal and prefrontal cortices for speech-segment information does not mature rapidly until after birth (Bethlehem et al., 2022; Buzi et al., 2024).

Evidence from behavioral, physiological-response, and neuroimaging data indicates that fetuses can not only "hear" sounds, but also begin auditory learning before birth. The classic study by Lecanuet et al. (1986) first revealed that fetal responses to sound stimulation are affected by factors such as heart-rate change patterns, sound intensity, and repetition frequency, suggesting that fetuses can initially encode relevant acoustic attributes. Subsequently, a large number of habituation-paradigm studies further confirmed this learning ability in fetuses. Shahidullah and Hepper (1994) found that 35-week fetuses may possess frequency-discrimination ability: they showed a habituation pattern of heart-rate deceleration during repeated pure-tone stimulation, and showed a dishabituation response when the frequency changed. A study by Lecanuet et al. (2000) showed that, in fetuses at 36–39 weeks, heart rate decreased significantly when the fundamental frequency of piano notes was switched (D4 $\rightarrow$ C5), indicating that fetal discrimination of sounds was based on changes in fundamental frequency rather than on mixed high-frequency harmonics. Notably, as gestational age increased, fetal responses to complex sounds also showed a transition from arousal to attention. Kisilevsky et al. (2004) found that, before 33 weeks of gestation, fetuses listening to sound

Heart rate during music listening most often shows an acceleratory (alerting) response; after 33 weeks of gestation, however, the corresponding pattern gradually shifts toward a deceleratory (attentional processing) response. This suggests that fetal attentional resources may participate in music processing.

With the development of fetal magnetoencephalography (fMEG), researchers have been able to directly observe fetal neural responses in utero. Using an oddball paradigm, Draganova et al. (2007) found that, as early as 28 weeks of gestation, the fetal brain can generate mismatch negativity (MMN)-like responses or late discriminative negativity (LDN) to changes in the fundamental frequency of complex tones (500 Hz $\rightarrow$ 750 Hz). This indicates that fetuses in utero may already possess an automatic capacity to detect changes in sound and to engage in early "prediction–prediction violation" processing, an important neural basis for auditory learning. Researchers have further examined the fetal capacity to process more complex acoustic features. Draganova et al. (2018) found that the fetal brain responds to amplitude-modulated (AM) sounds at different rates. Among these, 4 Hz amplitude modulation corresponds to a typical syllabic rhythm, and the latency of the corresponding fetal response shortens with increasing gestational age. In addition, the fetal response rate to 27 Hz amplitude modulation is the highest (approximately 86%), markedly exceeding that to pure-tone stimulation. This suggests that complex acoustic features more readily activate the fetal auditory system. Niepel et al. (2020) subsequently used fMEG to successfully record, for the first time, auditory steady-state responses (ASSR) in perinatal fetuses. ASSR reflect stable, synchronized neural oscillations generated by the brain in response to continuous, rhythmic stimulation. Studies show that the fetal brain can produce significant phase-locked responses to 27 Hz amplitude-modulated sounds. This means that the fetal auditory system can not only detect the onset and change of sounds, but that its neural activity can also maintain continuous synchronization with the rhythmic features of external sounds—an ability crucial for the future parsing of continuous speech streams. Movalled et al. (2023) noted that music or speech stimuli perceived during the fetal period can not only elicit immediate heart-rate or neural responses, but also form stimulus-specific memory representations that persist into the neonatal period. This suggests that fetal auditory learning may not occur only during a single stimulation episode, but may also induce enduring changes in neural plasticity, thereby constituting an early foundation for newborns' general auditory abilities and language-processing capacities.

Taken together, the above studies indicate that development of the fetal auditory system presents a continuous process that gradually transitions from structural/functional maturation to experience-dependent learning. In this process, the hierarchical maturation of the peripheral–brainstem–thalamic–cortical pathway and the emergence of fAEF primarily reflect neural maturation unfolding according to a developmental timetable, whereas learning phenomena such as habituation, frequency discrimination, and prenatal prosodic memory reflect experience-dependent shaping. Together, these two processes underpin the language sensitivity that newborns already possess at birth, such as a strong preference for human speech (Ma et al., 2025), learning of sequential structures and repetitive patterns (Gervain et al., 2012), and processing of different types of emotional prosody (Zhang et al., 2019; Zhang Dandan et al., 2023; Hou et al., 2024). It can therefore be seen that fetal speech perception is first subject to the dual constraints of the maturity of the auditory system and the availability of acoustic input in utero; these constraints make low-frequency prosodic cues the most stable and learnable speech information before birth.

It should be noted that the strength of conclusions in studies of fetal and neonatal speech perception depends to a large extent on the methods employed. Measures of fetal movement, heart rate, and behavioral preference have relatively high ecological validity and can reveal fetal or neonatal discrimination of, and preference for, auditory stimuli, but they usually allow only indirect inferences about auditory processing; whereas fMCG, fMEG, and electroencephalography/event-related potentials

(electroencephalography/event-related potential, EEG/ERP), functional near-infrared spectroscopy (fNIRS), fMRI, and other neurophysiological measurement methods each have their own advantages and limitations in temporal resolution, spatial localization, ecological validity, and applicable populations. Therefore, when integrating relevant evidence, it is necessary to distinguish the boundaries of the conclusions that different methods can support (Table 1).

Table 1. Comparison of major methods in studies of fetal and neonatal speech perception

Research method/pathway	Common participant population	Main measures	Advantages	Limitations	Questions suited for examination (and inferential limitations)
Fetal movement and heart-rate monitoring	Fetuses	Fetal-movement frequency, changes in fetal heart rate, heart-rate deceleration/acceleration, habituation responses	Usable in utero; high ecological validity; moderate temporal resolution	Reflects only autonomic/behavioral levels; affected by sleep–wake state	Whether fetuses show differentiated physiological responses to differences in frequency, rhythm, sound familiarity, or language type
Fetal magnetocardiography (fMCG)	Fetuses	Fetal cardiac magnetic signals, heart rate, and changes in heart-rate variability	Usable in utero; high temporal resolution; relatively high ecological validity	An autonomic nervous-system index; does not directly reflect cortical processing; requires magnetically shielded equipment	Whether fetuses show differentiated responses at the heart-rate/autonomic level to language or rhythm changes

Research method/pathway	Common participant population	Main measures	Advantages	Limitations	Questions suited for examination (and inferential limitations)
Fetal magnetoencephalography (fMEG)	Fetuses	Auditory evoked magnetic fields, mismatch responses, and other central nervous-system indices	Usable in utero; high temporal resolution; can reflect central auditory responses	Weak signals; susceptible to interference from fetal movement/fetal position; low spatial localization; equipment is scarce	Whether fetal central (cortical) auditory responses exist and whether they are differentiated
Electroencephalography / event-related potentials (EEG/ERP)	Neonates	MMR/MMN, evoked potentials, neural oscillations	High temporal resolution; portable; suitable for examining components such as MMR	Low spatial resolution; affected by sleep–wake state	The time course and neural differentiation of postnatal prosodic/segmental processing
Functional near-infrared spectroscopy imaging (fNIRS)	Neonates	Changes in HbO_2 and HbR; cortical activation and lateralization in regions such as the frontal and temporal lobes	Relatively high spatial resolution; high ecological validity; tolerant of movement; portable	Relatively low temporal resolution; cannot detect subcortical structures	Cortical localization and lateralization of neonatal speech processing

Research method/pathway	Common participant population	Main measures	Advantages	Limitations	Questions suited for examination (and inferential limitations)
Structural/functional magnetic resonance imaging (MRI/fMRI)	Fetuses, neonates	Brain structural maturation, functional connectivity, task-related activation, or resting-state networks	High spatial resolution; can examine whole-brain structure and connectivity	Low temporal resolution; sensitive to noise and motion; limited ecological validity	Structural maturation and development of functional connectivity in the auditory network
Behavioral preference / high-amplitude sucking paradigm	Neonates	Sucking frequency, looking/orienting responses, or sound-choice responses	Directly reflects differences in preference, familiarity, or motivation; high ecological validity	Behavioral level only; affected by arousal, fatigue, and motor ability	Neonatal preference for and discrimination of the mother tongue and familiar speech/prosody

Research method/pathway	Common participant population	Main measures	Advantages	Limitations	Questions suited for examination (and inferential limitations)
Preterm-infant studies / natural-control designs	Preterm infants	Behavioral responses, EEG/ERP, fNIRS, fMRI, and other multimodal indices	Treats the absence/premature interruption of in utero experience as a natural control, allowing examination of the role of prenatal experience	The absence of prenatal experience and restricted neurodevelopment are mutually confounded pathways and are difficult to fully separate	The relative roles of prenatal auditory experience and neural maturation in development (limited by confounding; causal attribution requires caution)

Against the above physiological and methodological background, whether—and how—low-frequency prosodic processing formed during the fetal period continues after birth constitutes the starting point for understanding the continuity of speech perception from the prenatal to the postnatal period.

3 Continuity of Prenatal-Postnatal Prosodic Processing and Its Neural Mechanisms

3.1 Prosodic Input in the Fetal Period and Prenatal Prosodic Learning

Among all acoustic information that can enter the uterus, prosodic and rhythmic features are the most complete, stable, and accessible speech cues. These prosodic components include the rhythmic type of speech streams—stress-timed, syllable-timed, or mora-timed languages—as well as foot structure and stress patterns, such as strong-weak alternations and differences in syllable duration, and intonational contours and pitch changes, such as rises and falls in fundamental frequency and patterns of prominence. Together, they constitute the characteristic signals of the rhythm of the native language, provide the earliest statistical cues for language classification and recognition, and represent the primary speech input available to the fetus before birth (Ortiz Barajas & Gervain, 2021).

A large body of evidence has revealed the influence of prenatal prosodic experience on fetal auditory processing. Behavioral and physiological studies show that, by late gestation, fetuses can distinguish languages with different rhythmic types and can form stable behavioral response patterns based on the prosodic characteristics of the maternal language. Kisilevsky et al. (2009) found that fetuses in an English-language environment during late gestation showed heart-rate change patterns more similar to those elicited by the mother' s voice when listening to unfamiliar Mandarin Chinese, whereas their responses were weaker when listening to English speech produced by an unfamiliar speaker. The researchers argued that fetal heart-rate acceleration more strongly reflected "novelty processing" of the distinctive pitch, tone, and stress patterns of Chinese, and that this processing mainly arose from the pronounced differences in prosodic features between the two languages. Similar results have also been verified in more precise fetal magnetocardiography (fMCG) studies. Using fMCG, Minai et al. (2017) presented fetuses with English and Japanese speech materials read by the same bilingual speaker, and likewise found that when the stimulus switched from familiar English to mora-timed Japanese, fetuses still showed stable heart-rate acceleration, indicating a general physiological sensitivity to differences in linguistic prosody and rhythm.

Fetuses can also form lasting memories of complex prosodic structures before birth. DeCasper and Spence (1986) asked pregnant women to repeatedly read aloud a story with a highly stable rhythmic structure during the final six weeks of pregnancy, and found that newborns a few hours after birth preferred the story with the familiar rhythm. This indicates that fetuses can discriminate prosody and can already form long-term memories of prosodic structure in utero. Partanen et al. (2013) further confirmed with EEG that Finnish fetuses who had been exposed prenatally to vowel changes in syllable sequences ("tatata" → "tatota") or pitch changes ("tatata" → "taTAta") produced enhanced mismatch responses (MMR) after birth. Notably, this learning effect could also generalize to speech features that had not been directly trained, such as changes in intensity and duration, and the strength of the neural response was positively correlated with the total amount of speech stimulation received during the fetal period. These studies indicate that fetuses do not merely passively perceive prosodic input; rather, they can actively extract the internal structural regularities of prosody and form abstract representations.

3.2 Neural Mechanisms of Prosodic Processing in Newborns

After birth, newborns show prosodic sensitivity at both the behavioral and neural levels that is consistent with prenatal sensitivity. Behavioral studies indicate that newborns can distinguish the maternal language from foreign languages solely on the basis of rhythmic class, and show a sucking preference for the rhythm of the maternal language; even when processed through low-pass

After filtering removed segmental information, newborns were still able to perceive the rhythmic framework of the speech stream (Moon et al., 1993). Nazzi et al. (1998) pointed out that suprasegmental features are important cues for newborns to discriminate different languages. Researchers found that newborns with prenatal exposure to French could distinguish low-pass-filtered English from Japanese, but could not distinguish rhythmically similar Spanish from Catalan. Studies of preterm infants have also provided a critical perspective for understanding the role of prenatal prosodic experience. Peña et al. (2010) showed that preterm infants born approximately 3 months early, although exposed to substantial native-language input after birth, still could not discriminate other languages rhythmically similar to their mother tongue at a chronological age of 6 months; only at a chronological age of 9 months (equivalent to a corrected age of 6 months) did they exhibit abilities similar to those of full-term infants. This finding indicates that postnatal linguistic experience alone is unlikely to fully compensate for the in-utero prosodic input lost because of preterm birth, suggesting that prenatal prosodic experience and neural maturation both play irreplaceable roles in language acquisition. In addition, prosodic information also appears to influence vocalization patterns at birth. Mampe et al. (2009) found that the cries of German and French newborns showed prosodic patterns in their pitch/fundamental-frequency contours that resembled their native languages. This seemed to suggest that fetal prosodic exposure had already influenced the pitch trajectory of crying, and the authors therefore argued that prenatal prosodic experience not only affects newborns' auditory perception but also directly shapes their vocal behavior. It should be noted, however, that the statistical methods of this study were subsequently questioned. Gustafson et al. (2017) argued that multiple cry recordings from the same infant should not be treated as independent samples. If future studies can verify this finding through more rigorous statistical analyses, they will provide more comprehensive evidence for mechanisms of early prenatal learning of prosody.

At the neural level, newborns' processing of prosodic structure likewise shows a "native-language" signature of prenatal experience. Using fNIRS, Abboub et al. (2016) examined how newborns use duration, pitch, and intensity to achieve prosodic grouping. They found that only when the acoustic grouping pattern violated the prosodic rules of the language encountered prenatally did the temporo-parietal junction show a significantly stronger hemodynamic response. This suggests that newborns' bias in prosodic grouping may already have been tuned by prenatal language experience, manifesting as selective sensitivity to native-language-specific prosodic prominence cues, such as duration or pitch. Benavides-Varela and Gervain (2017) further found that when multiword sequences were presented as isolated lists, newborns could track changes in word order in the temporal lobe and the left inferior frontal gyrus; however, once the sequences were "packaged" as whole natural sentences by intonation, the brain instead responded to the overall sentential intonation and no longer additionally parsed the internal word order. This indicates that newborns with limited processing resources are more inclined to prioritize the use of prosodic boundaries to segment the speech stream. Overall, existing evidence suggests that newborns' sensitivity to mother-tongue-specific prominence cues may be related to prenatal prosodic experience, and that the temporal–inferior frontal network is involved relatively early in language-related processing such as prosodic grouping and sequence tracking.

Beyond the lexical level, Martinez-Alvarez et al. (2023) further extended research on newborns' prosodic processing to the discourse level. The researchers had French newborns listen to the "natural prosodic contour" of their native language and to a temporally reversed "reversed prosodic contour," while using fNIRS to record the corresponding hemodynamic changes. The results showed that both types of speech elicited hemodynamic activation in the bilateral temporal lobes, but the reversed condition produced a significantly stronger neural response in the right temporal region than did the natural condition, and this difference showed a right-lateralized advantage. Because the two types of stimuli were matched in speech content and segmental material, with the main difference lying in the natural temporal organization of the prosodic contour, this result suggests that newborns have early

neural sensitivity, and its right-lateralized hemispheric pattern suggests that auditory experience with the native language during the fetal period has already led the right temporal lobe to begin forming specialized processing of prosody. In addition, using EEG and long-range temporal correlation analysis, Mariani et al. (2023) found that neonates showed enhanced temporal organization of neural oscillations in the θ band (4–8 Hz) after listening to a prenatally familiar language. The θ band corresponds to the time scale of syllable processing and may participate in tracking and integrating the speech envelope and syllabic rhythmic structure (Draganova et al., 2018). This finding suggests that prenatal experience with a familiar language may influence the neural-dynamical organization of early postnatal speech-rhythm processing. However, that study did not directly examine prosodic processing at the phrase, stress, or discourse levels corresponding to the δ band; therefore, its results are more appropriately regarded as evidence that prenatal familiar-language experience affects the organization of speech over slow temporal scales. Overall, sustained fetal exposure to low-frequency acoustic information such as the native language' s fundamental frequency, rhythm, and syllabic contours may make the postnatal neonatal brain more likely to form sustained temporal organization of low-frequency neural oscillations after native-language input, thereby providing an early temporal framework for subsequent speech processing.

Overall, existing evidence rather consistently suggests that there is a certain continuity between prenatal and early postnatal prosodic processing. Because the intrauterine acoustic environment is more conducive to preserving low-frequency information about prosody, rhythm, and fundamental-frequency contours, fetal speech experience is mainly manifested as perception, memory, and neural tuning of prosodic structure; the low-frequency rhythmic system tuned by prenatal prosodic experience may also provide a temporal-organizational framework for postnatal segmental processing. Thus, this prosody-dominant processing mode may constitute an important starting point for early postnatal speech perception.

4 Development of Postnatal Segmental Processing and Its Neural Mechanisms

4.1 Rapid development of neonatal segmental-processing ability

Segmental processing mainly refers to neonates' perception, encoding, and discrimination of vowels, consonants, and related phonetic features, including formant structure, place and manner of articulation, voicing, voice onset time, and so forth; the segmental level refers to phonemic/segmental-granularity representations that are distinct from the prosodic level. Unlike prosodic cues, segments have very limited availability prenatally. Constrained by intrauterine low-pass filtering, the fetus can generally obtain only the low-frequency components of vowels, whereas fine-grained features such as consonants and voicing, which depend on high-frequency information, are almost unavailable before birth; at present, there is also no evidence that the fetus can learn them. Therefore, the refinement of segmental processing mainly occurs after birth. In contrast to the prosody-dominated processing of the prenatal stage, after birth the full-band speech signal transmitted through air can directly stimulate the neonatal auditory system, allowing high-frequency segmental features to enter the processing system for the first time in a relatively complete form. This abrupt change in input quality constitutes an important physical condition for the processing of fine-grained speech features. At the same time, the superior temporal gyrus, the primary auditory cortex A1, and its connections with the left inferior frontal gyrus—cortical structures responsible for processing acoustic spectral features—also mature rapidly within several weeks after birth (Taymourtash et al., 2023), bringing infants into a "rapid segmental tuning" window.

Behavioral and neuroscientific evidence consistently indicates that neonates' segmental-processing ability enters a stage of rapid development after birth. EEG and MEG studies show that neonates only a few hours to a few days old can already distinguish native-language from non-native-language vowels, and for vowels

Categorical changes elicit stable MMR/MMN responses (Dehaene-Lambertz & Peña, 2001; Cheour et al., 2002), which is related to the fact that vowel energy is concentrated mainly in low frequencies and that, during the fetal period, there is already a coarse-grained perceptual basis. By contrast, fine-grained segmental features that depend on high-frequency components—such as consonantal contrasts, voicing differences, burst duration, and frication noise—must be gradually established under full-spectrum, air-conducted speech input; therefore, stable neural sensitivity to these features emerges only at 1–3 months after birth or even later (Eimas et al., 1971; Werker & Tees, 1984). Thus, the opening of the rapid tuning window for speech segments may be determined primarily by rapid cortical maturation, whereas the specific shaping of segmental representations within that window depends on postnatal full-spectrum speech experience.

Studies of preterm infants have provided important clues about the relationship between segmental processing and postnatal speech input, but the interpretation of this evidence requires caution. Preterm infants are exposed earlier after birth to extrauterine full-spectrum speech input, which can be used to examine the influence of postnatal experience on segmental processing. However, they also experience premature termination of the low-pass-filtered auditory experience of the late-gestational intrauterine environment, interruption of the development of brain structures and auditory pathways, and possible accompanying medical complications. Therefore, differences in speech processing among preterm infants must be considered in terms of both an "experiential pathway" and a "maturational pathway." From the perspective of the experiential pathway, Gonzalez-Gomez and Nazzi (2012) found that preterm infants' performance in a consonant-sequence perception task corresponded more closely to "hearing age" than to actual chronological age; that is, their segmental perception was associated to some extent with the duration of postnatal auditory exposure. This result suggests that postnatal speech input and the time course of auditory exposure may participate in the development of segmental processing. However, because "hearing age" itself is intertwined with the duration of auditory-system use, overall maturational status, and differences in the early environment, this study is better understood as indicating the importance of postnatal auditory exposure. From the perspective of the maturational pathway, the high-density EEG study by Novitskiy et al. (2023) showed that preterm infants still exhibit certain deficits in the neural encoding of speech detail. Using the frequency-following response (FFR) to examine Cantonese-learning preterm infants' ability to encode lexical tone, the study found that they were weaker than full-term infants on neural synchrony indices such as pitch tracking and inter-trial phase coherence. Because tone processing depends on stable pitch tracking and precise neural synchrony, this finding suggests that preterm birth may affect auditory neural encoding on a rapid timescale, thereby weakening the analysis of tones and related speech details. More importantly, this deficit appeared under both native- and non-native-tone conditions; on this basis, the researchers argued that its source is more likely related to premature interruption of prenatal neural development than simply to insufficient language experience. In sum, evidence from preterm infants both reflects the shaping effect of postnatal experience on segmental processing and suggests constraints imposed by the time course of neural development itself. Existing designs still make it difficult to draw a complete causal distinction between the two. Future studies need to control for corrected gestational age, duration of postnatal auditory exposure, medical risk factors, and the quality of language input, so as to distinguish more reliably between the two explanatory pathways of "prenatal experience" and "neural maturation."

4.2 Neural mechanisms of neonatal segmental processing

Beyond developmental trajectories, existing studies have further focused on the plasticity of postnatal segmental processing. Wu et al. (2022) used fNIRS to record brain activation responses in neonates only a few hours after birth following exposure to natural and time-reversed vowels. The researchers found that only 5 hours of speech input was sufficient to induce a significant increase in oxygenated hemoglobin in the left temporal lobe and the left inferior frontal gyrus. This change indicates that neonates can form new phonemic representations within a relatively short period of time. In addition, neonates were able to distinguish natural and time-

vowel sequences after temporal reversal, indicating that they are not only sensitive to the segments themselves, but can also rapidly learn the temporal organizational structure of segments. This powerful immediate-learning effect suggests that, although systematic segmental experience cannot be acquired during the fetal period, the auditory cortex has already entered a state of readiness after birth for rapidly tuning to high-frequency speech details; postnatal input of full-band speech signals drives rapid adaptation and reorganization of the left auditory cortex to the segmental features of the native language.

Cabrera and Gervain (2020) used a multimodal approach combining EEG and fNIRS to investigate the neural coding patterns of segmental processing in neonates. Using an EEG oddball paradigm with degraded speech stimuli, the researchers observed, in the left prefrontal region of neonates, a mismatch response (MMR) resembling the adult MMN, which was most sensitive to plosive sounds and vowel transitions, indicating that neonates can use features on a rapid time scale to distinguish speech segments. The fNIRS data showed that rapid speech cues mainly elicited activation in the left inferior frontal gyrus and the left temporal lobe, whereas slow temporal structures (such as the prosodic envelope) more strongly activated the right superior temporal gyrus and middle temporal gyrus. This functional differentiation between the left and right hemispheres at different temporal scales suggests that the neonatal brain may already possess the rudiments of a dual-channel processing mechanism based on a division of labor between fast and slow temporal scales, implying that neonatal segmental processing may be built upon an early capacity to decompose and integrate the temporal structure of speech. This finding can be linked to developmental research on speech-envelope tracking. Ortiz Barajas et al. (2021) used EEG to examine neural tracking responses to natural sentences in French, Spanish, and English in neonates and 6-month-old infants. They found that neonates could perform both amplitude tracking and phase tracking of the speech envelopes of all three languages, and that this tracking was not modulated by language familiarity, indicating that the brain at birth already possesses the basic auditory capacity to track the slow temporal structure of speech. By 6 months, phase tracking remained widespread, but amplitude tracking weakened for the native language while being retained for unfamiliar languages. This suggests that native-language processing may undergo, within the first few months after birth, a reorganization from acoustic-envelope/syllabic-scale processing toward language-specific phonemic processing. At birth, suprasegmental units such as syllables and prosodic contours occupy a relatively large weight in speech representation; with the initiation of native-phonology tuning, early word-form learning, and grammatical learning, infants gradually need to process finer-grained phonemic information.

In light of the preceding discussion of prenatal prosodic experience, it can be inferred that postnatal segmental processing may unfold gradually within an early low-frequency temporal framework for speech. Mariani et al. (2023) found that, after neonates listened to a prenatally familiar language, long-duration temporal organization in the θ band (4–8 Hz) was enhanced, whereas the γ band (>35 Hz) did not show a corresponding enhancement. The θ band is closer to the temporal scale of syllable rate and speech-envelope integration, whereas the γ band is associated with faster temporal scales of segmental/phonemic processing. This result suggests that experience with a prenatally familiar language may first affect the association between low-frequency prosodic information and θ-band syllabic temporal organization; after birth, as full-band speech experience increases, finer segmental cues such as vowels and consonants only then gradually enter the processing system and become coordinated with the existing slow-temporal-scale framework. However, because existing studies have not directly tested coupling between θ–γ oscillations, nor have they longitudinally tracked the development of neonatal segmental processing, it is currently more appropriate to regard these findings as indirect evidence for the prosody–segment bridging hypothesis.

In sum, the existing evidence more strongly supports a limited continuity in segmental processing. During the fetal period, only low-frequency cues such as vowel F_1 can be accessed, while high-frequency fine-grained features such as consonants and voicing depend primarily on postnatal full-band input and cortical maturation. Prenatal low-frequency prosodic experience, meanwhile, may provide slow-temporal-scale organizational conditions for the extraction, grouping, and integration of postnatal segmental cues. Behavioral and

Neural evidence also indicates that although newborns can form new phonemic representations within hours after birth (Wu et al., 2022), more fine-grained consonant correspondences often require several weeks or even months to mature gradually (Eimas et al., 1971; Werker & Tees, 1984). Therefore, it should not be assumed that newborns have already completed fine-grained segment learning prenatally. Rather, the rapid development of segmental processing is the result of the joint effects of the limited prenatal interface, the slow-timescale organizational conditions provided by prenatal low-frequency prosodic experience, and, after birth, full-band input together with rapid cortical maturation.

5 Summary and Outlook

To integrate the evidence reviewed above more intuitively, this article places the maturation of auditory-system structure, the acoustic input available at each stage, behavioral evidence, and neural evidence within a single developmental framework in the form of a timeline (Figure 1). As can be seen, the development of fetal–neonatal speech perception is a continuous developmental process spanning the prenatal period and early postnatal life. In the prenatal stage, it is conditioned primarily by the functionalization of the auditory system and by low-frequency prosodic input in utero; in the postnatal stage, full-band speech input and rapid cortical maturation jointly drive the rapid tuning of segmental/phonemic processing.

Text in Figure 1

- Timeline: gestational week 24; gestational weeks 28–34; gestational weeks 35–40; birth; hours → weeks; months.
- Structure/functional maturation:
  - Peripheral auditory organs approach completeness; the auditory system begins to become functional.
  - Brainstem–thalamic pathways gradually mature; fAEP emerges.
  - Cortical auditory networks are gradually established; frontal–temporal functional connectivity appears.
  - The cortex gradually matures and functions become refined; neural remodeling continues.
- Acoustic input:
  - Low-frequency prosody can be transmitted, whereas high-frequency speech signals are attenuated.
  - Full-band speech is transmitted, with input of high-frequency segmental details.
- Behavioral evidence:
  - Reflexive motor responses appear.
  - Frequency discrimination; heart-rate habituation and dishabituation responses.
  - Discrimination of speech rhythm; metric beat and musical memory.
  - Preference for native-language prosody; discrimination between native- and foreign-language rhythms.
  - Phonemic contrasts can be perceived.
  - Discrimination between native- and foreign-language vowels.
  - Lateralization of prosodic-processing networks; activation of the right superior temporal gyrus and middle frontal gyrus.
  - Neonatal phonemic processing; activation in the left temporal region and responses in association areas.
  - Formation of phonemic categories; statistical learning and attenuation of non-native contrasts.
- Neural evidence:
  - fAEP: MMN/LDN-like responses can already be recorded reliably.

  - ASSR can be recorded.
- Legend:
  - Early general auditory responses (not specific to prosody or segments).
  - Prosody-related evidence.
  - Segment/phoneme-related evidence.
  - Structural maturation and acoustic input.
- Overall trajectory: prosody-dominated → rapid segmental tuning.

Figure 1. Continuity timeline of speech-perception development from the fetal period to early postnatal life

Note: The horizontal axis indicates key developmental nodes from gestational week 24 to 1–3 months after birth and beyond. The four tracks summarize, respectively, structural/functional maturation of the auditory system, acoustic input available at different stages, behavioral evidence, and neural evidence. Box colors indicate the processing type of the evidence: purple denotes general early auditory responses, green denotes prosody-related evidence, orange denotes segment/phoneme-related evidence, and gray denotes structural maturation and acoustic input.

Based on the evidence above, this article builds, on the existing perspective of perinatal continuity, a prenatal–postnatal continuity framework for speech-perception development. For ease of exposition, this article summarizes this prenatal–postnatal continuity framework of speech-perception development into the following five core mechanisms and hypotheses: Boundary conditions of the framework: the low-pass-filtered soundscape formed by the uterus and amniotic fluid means that the speech information available to the fetus consists mainly of low-frequency prosody, rhythm, and fundamental-frequency contours, whereas the availability of high-frequency segmental details is clearly limited (Gélat et al., 2025; Moon et al., 2013). This constitutes the physical premise of the continuity framework. The prosody-dominant mechanism in the prenatal stage: in utero

In a low-pass-filtered auditory scene, after the fetal auditory system becomes functional, what it primarily encounters and learns are slow-timescale information in the native language, such as low-frequency prosody, rhythm, and fundamental-frequency contours; repeated exposure may form familiarization or memory traces (Kisilevsky et al., 2009; Partanen et al., 2013). After birth, this is manifested as selective sensitivity to native-language prosody (Abboub et al., 2016) and enhanced long-range correlations in the θ band (Mariani et al., 2023).

Hypothesis of a prosody–segment transition mechanism: the low-frequency prosodic temporal structure formed prenatally and the related syllabic temporal organization in the θ band may continue after birth as a slow-timescale framework, providing temporal-organizational conditions for the extraction, grouping, and integration of segmental cues, while low-frequency cues such as vowel F_1 may constitute a limited prenatal interface for segmental processing. This bridging mechanism remains to be further tested experimentally. Mechanism of rapid postnatal tuning of segments: after birth, full-spectrum speech input and rapid cortical maturation jointly promote, within the above temporal framework, the rapid refinement of segmental/phonemic processing (Wu et al., 2022; Cabrera & Gervain, 2020), enabling speech perception to shift gradually from prenatal prosody dominance to rapid postnatal segmental tuning. Mechanistic attribution within the framework: prenatal speech experience, postnatal environmental input, and neural maturation both divide labor and interact with one another. Neural maturation provides the structural basis and plasticity window for processing, and its course runs through all prenatal and postnatal stages, including the functionalization of peripheral and central auditory pathways, the emergence of fetal auditory evoked responses, and the rapid postnatal development of the superior temporal gyrus/primary auditory cortex and its connections with frontal language-related regions. Prenatal speech experience is mainly reflected in familiarization with, and experience-dependent tuning to, the mother' s voice, native-language prosody, familiar speech materials, and low-frequency melodic contours. Postnatal environmental input, through air-conducted full-spectrum speech, allows fine segmental features such as vowels and consonants to enter the processing system in a relatively complete form. The continuity of prosodic processing is mainly reflected in the coupling between prenatal low-frequency input and neural maturation, whereas the rapid refinement of segmental processing depends on the joint action of the prenatal prosodic framework, postnatal input, and cortical maturation. Together, these three factors determine when early speech perception is plastic, what kinds of input it accepts, and toward what type of representation it is ultimately tuned.

Because this framework is primarily an integrative inference from existing evidence, many of its components remain hypotheses to be tested. Around the core claims of this framework, future research can proceed in the following directions.

First, future research needs to distinguish more directly the relative contributions of prenatal experience, postnatal input, and neural maturation. The framework proposed in this paper holds that neural maturation provides the structural basis and plasticity window for processing, prenatal low-frequency prosodic experience forms early familiarization and a temporal-organizational framework, and postnatal full-spectrum speech input drives the rapid refinement of segmental/phonemic processing. However, existing studies often have difficulty fully separating these three factors. Future work could test the following questions through longitudinal designs, matching of corrected gestational age and auditory age, comparisons between preterm and full-term infants, records of prenatal exposure dose, and control of postnatal input amount: whether the degree of prenatal prosodic exposure can predict the strength with which postnatal low-frequency neural oscillations track native-language prosody; whether full-spectrum speech input during the first few hours to days after birth is sufficient to induce rapid changes in segmental/phonemic processing; and whether the level of native-language prosodic tuning at birth can predict the speed of subsequent segmental tuning. Answers to these questions will help clarify how speech experience and neural maturation contribute to prenatal–postnatal continuity.

Second, future research needs to further test the “prosody–segment transition mechanism.” This paper proposes that the prenatally formed low-frequency

prosody–θ-oscillatory temporal structure may provide a slow-timescale scaffold for postnatal segmental processing, but this bridging mechanism is still inferred mainly from cross-study integration and lacks direct experimental validation. Future research could adopt standardized acoustic-manipulation paradigms to separately manipulate rhythm, fundamental-frequency contours, temporal envelopes, frequency-band ranges, and the strength of segmental contrasts, comparing newborns' behavioral preferences and neural responses under conditions such as preserved prosodic cues, weakened prosodic cues, and enhanced segmental cues. In particular, it is necessary to avoid mistaking novelty effects for responses induced by prosodic or segmental differences. Therefore, designs using forward/reversed homologous materials, low-pass/full-band materials, and envelope/fine-structure separation materials should all serve the same core question: whether the low-frequency temporal structure available prenatally can truly facilitate the extraction and tuning of fine-grained segmental cues after birth.

Third, future research needs to examine the dose–response relationship of prenatal sound exposure under stricter experimental control and safety frameworks. Existing studies of prenatal auditory stimulation differ substantially in stimulus type, sound intensity/frequency band, exposure duration, gestational week of onset, control conditions, and outcome measures; moreover, some studies have problems such as insufficient randomization, lack of allocation concealment and blinding, and incomplete reporting, which limits the reproducibility and generalizability of their conclusions (Movalled et al., 2023). The background soundscape in clinical environments may also alter newborns' behavioral state and autonomic nervous activity, thereby affecting observation of the effects of target stimuli (Philbin, 2000). Therefore, within ethical and safety boundaries, future research should specify what kinds of sounds, at what intensity, for what duration, and during which gestational windows are most likely to produce stable and beneficial neurobehavioral effects. Accordingly, parallel recording of multiple indicators is of great value: during the fetal period, measures such as heart rate, respiration, movement, fMCG/fMEG/fMRI can be combined; after birth, EEG/fNIRS, behavioral preferences, and follow-up of language development can be combined, so as to establish a continuous evidence chain from acoustic input and neural responses to behavioral outcomes. On this basis, computational tools such as deep learning and large language models can serve as auxiliary modeling approaches for constructing interpretable neurocomputational models, further revealing how auditory experience affects neural circuits and early language functions across different timescales.

Fourth, future research needs to expand the types of languages, populations, and developmental stages studied in order to test the universality and boundary conditions of the framework. Current evidence comes mainly from European and American language environments, while research on tone languages, bilingual/multilingual environments, and high-risk populations remains clearly insufficient. Languages differ markedly in rhythmic type, stress structure, tone systems, and phonemic distributions; these differences may alter the prosodic cues available during the fetal period and may also affect the pathway of postnatal segmental tuning. For example, in tone languages, fundamental-frequency contours are both prosodic information and lexical-contrast cues, providing a relatively special window for testing the "prosody–segment transition mechanism." At the same time, high-risk groups such as preterm infants can be used to examine the relationships among altered intrauterine experience, postnatal input compensation, and constrained neural maturation, but medical risks, corrected gestational age, and duration of auditory exposure must be carefully controlled. In the future, cross-stage longitudinal cohorts should be established to connect prenatal acoustic exposure, fetal neural responses, postnatal speech processing, and later language-development outcomes, thereby identifying key sensitive periods, cross-linguistic differences, and early risk indicators.

Overall, research on prenatal auditory experience helps deepen our understanding of the starting point of language development. The fetus already possesses rudimentary auditory-processing capacities in utero; prenatal low-frequency prosodic experience may provide a temporal organizational framework for postnatal speech processing, while full-band input after birth and maturation of the nervous system jointly promote rapid tuning of segmental/phonemic processing. If future research can focus on the above testable

hypotheses. Further advances in longitudinal designs, acoustic control, multimodal measurements, and cross-linguistic samples will further deepen scientific understanding of the origins of language ability and provide a more solid foundation for the early screening and intervention of speech and language disorders.

References

Chen Yu, Zhang Dandan. (2020). Brain mechanisms of neonatal speech perception. *Journal of Psychological Science*, *43*(4), 844-849.

Li Sijin, Wang Tingdong, Peng Zhilin, Zhang Dandan. (2023). Newborns' perception, discrimination, and learning of speech. *Advances in Psychological Science*, *31*(12), 2295-2305.

Song Xinyan, Meng Xiangzhi. (2012). Development of infant speech perception and its mechanisms. *Advances in Psychological Science*, *20*(6), 843-852.

Zhang Dandan, Li Yiwei, Yu Wenwen, Mo Licheng, Peng Cheng, Liu Lili. (2023). Development of emotional bias in infants aged 0–1 year: A near-infrared imaging study. *Acta Psychologica Sinica*, *55*(6), 920-929.

Abboub, N., Nazzi, T., & Gervain, J. (2016). Prosodic grouping at birth. *Brain and Language*, *162*, 46-59.

Anikin, A., Canessa-Pollard, V., Pisanski, K., Massenet, M., & Reby, D. (2023). Beyond speech: Exploring diversity in the human voice. *iScience*, *26*(11), 108204.

Benavides-Varela, S., & Gervain, J. (2017). Learning word order at birth: A NIRS study. *Developmental Cognitive Neuroscience*, *25*, 198–208.

Bethlehem, R. A. I., Seidlitz, J., White, S. R., Vogel, J. W., Anderson, K. M., Adamson, C., Adler, S., Alexopoulos, G. S., Anagnostou, E., Areces-Gonzalez, A., Astle, D. E., Auyeung, B., Ayub, M., Bae, J., Ball, G., Baron-Cohen, S., Beare, R., Bedford, S. A., Benegal, V., ···Alexander-Bloch, A. F. (2022). Brain charts for the human lifespan. *Nature*, *604*(7906), 525–533.

Birnholz, J. C., & Benacerraf, B. R. (1983). The development of human fetal hearing. *Science*, *222*(4623), 516–518.

Buzi, G., Eustache, F., Droit-Volet, S., Desaunay, P., & Hinault, T. (2024). Towards a neurodevelopmental cognitive perspective of temporal processing. *Communications Biology*, *7*(1), 987.

Cabrera, L., & Gervain, J. (2020). Speech perception at birth: The brain encodes fast and slow temporal information. *Science Advances*, *6*(30), eaba7830.

Calixto, C., Taymourtash, A., Karimi, D., Snoussi, H., Velasco-Annis, C., Jaimes, C., & Gholipour, A. (2024). Advances in fetal brain imaging. *Magnetic Resonance Imaging Clinics of North America*, *32*(3), 459–478.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, *1*(5), 351–353.

Cheour, M., Martynova, O., Näätänen, R., Erkkola, R., Sillanpää, M., Kero, P., Raz, A., Kaipio, M.-L., Hiltunen, J., Aaltonen, O., Savela, J., & Hämäläinen, H. (2002). Speech sounds learned by sleeping newborns. *Nature*,

415(6872), 599–600.

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

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, 9(2), 133–150.

Dehaene-Lambertz, G., & Peña, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport*, 12(14), 3155–3158.

Draganova, R., Eswaran, H., Murphy, P., Lowery, C., & Preissl, H. (2007). Serial magnetoencephalographic study of fetal and newborn auditory discriminative evoked responses. *Early Human Development*, 83(3), 199–207.

Draganova, R., Schollbach, A., Schleger, F., Braendle, J., Brucker, S., Abele, H., Kagan, K. O., Wallwiener, D., Fritsche, A., Eswaran, H., & Preissl, H. (2018).

Fetal auditory evoked responses to onset of amplitude modulated sounds. A fetal magnetoencephalography (fMEG) study. *Hearing Research*, 363, 70-77.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968), 303-306.

Gélat, P., David, A. L., Haqhenas, S. R., Henriques, J., Thibaut de Maisieres, A., White, T., & Jauniaux, E. (2019). Evaluation of fetal exposure to external loud noise using a sheep model: Quantification of in utero acoustic transmission across the human audio range. *American Journal of Obstetrics and Gynecology*, 221(4), 343.e1-343.e11.

Gélat, P., van' t Wout, E., Haqshenas, R., Melbourne, A., David, A. L., Mufti, N., Henriques, J., Thibaut de Maisières, A., & Jauniaux, E. (2025). Evaluation of fetal exposure to environmental noise using a computer-generated model. *Nature Communications*, 16(1), 3916.

Gervain, J., Berent, I., & Werker, J. F. (2012). Binding at birth: The newborn brain detects identity relations and sequential position in speech. *Journal of Cognitive Neuroscience*, 24(3), 564-574.

Gonzalez-Gomez, N., & Nazzi, T. (2012). Phonotactic acquisition in healthy preterm infants. *Developmental Science*, 15(6), 885-894.

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

Gustafson, G. E., Sanborn, S. M., Lin, H.-C., & Green, J. A. (2017). Newborns' cries are unique to individuals (but not to language environment). *Infancy*, 22(6), 736-747.

Helwany, M., Arbor, T. C., & Tadi, P. (2023). *Embryology, ear.* StatPearls Publishing.

https://www.ncbi.nlm.nih.gov/books/NBK557588/

Hepper, P. G., & Shahidullah, B. S. (1994). Development of fetal hearing. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, *71*(2), F81-87.

Hou, X., Zhang, P., Mo, L., Peng, C., & Zhang, D. (2024). Sensitivity to vocal emotions emerges in newborns at 37 weeks gestational age. *eLife*, *13*, RP95393.

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 Behavior & Development*, *32*(1), 59-71.

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

Lecanuet, J.-P., Granier-Deferre, C., Cohen, H., Le Houezec, R., & Busnel, M.-C. (1986). Fetal responses to acoustic stimulation depend on heart rate variability pattern, stimulus intensity and repetition. *Early Human Development*, *13*(3), 269–283.

Lecanuet, J.-P. (1996). Prenatal auditory experience. In I. Deliège & J. Sloboda (Eds.), *Musical beginnings: Origins and development of musical competence* (pp. 2–34). Oxford University Press.

Lecanuet, J. P., Granier-Deferre, C., Jacquet, A. Y., & DeCasper, A. J. (2000). Fetal discrimination of low-pitched musical notes. *Developmental Psychobiology*, *36*(1), 29–39.

Leibovitz, Z., Lerman-Sagie, T., & Haddad, L. (2022). Fetal brain development: regulating processes and related malformations. *Life*, *12*(6), 809.

Ma, Y., Zhang, J., Dang, R., Wang, N., Wang, Y., Yu, M., Chen, M., Shen, P., Wang, Q., & Huang, J. (2025). Perception of native vs. non-native language and non-speech sounds in one-week-old neonates: An fNIRS study. *Brain Research Bulletin*, *226*, 111370.

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology*, *19*(23), 1994–1997.

Mappa, I., Di Mascio, D., Carbone, L., Lu, J. L. A., Sorrenti, S., Patelli, C., D' Amico, A., Matarrelli, B., Giuliani, G. A., Neola, D., Di Girolamo, R., Sarno, L., Khalil, A., Rizzo, G., Maruotti, G. M., & D' Antonio, F. (2024). Sonographic cortical development and anomalies in the fetus: A systematic review and meta-analysis. *Biomedicines*, *12*(7), 1397.

Mariani, B., Nicoletti, G., Barzon, G., Ortiz Barajas, M. C., Shukla, M., Guevara, R., Suweis, S. S., & Gervain, J. (2023). Prenatal experience with language shapes the brain. *Science Advances*, *9*(47), eadj3524.

Martinez-Alvarez, A., Benavides-Varela, S., Lapillonne, A., & Gervain, J. (2023). Newborns discriminate utterance-level prosodic contours. *Developmental Science, 26*(2), e13304.

Mastropieri, D., & Turkewitz, G. (1999). Prenatal experience and neonatal responsiveness to vocal expressions of emotion. *Developmental Psychobiology, 35*(3), 204–214.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*(2), 143–178.

Minai, U., Gustafson, K., Fiorentino, R., Jongman, A., & Sereno, J. (2017). Fetal rhythm-based language discrimination: A biomagnetometry study. *NeuroReport, 28*(10), 561–564.

Monson, B. B., Eaton-Rosen, Z., Kapur, K., Liebenthal, E., Brownell, A., Smyser, C. D., Rogers, C. E., Inder, T. E., Warfield, S. K., & Neil, J. J. (2018). Differential rates of perinatal maturation of human primary and nonprimary auditory cortex. *eNeuro, 5*(1), ENEURO.0380-17.2017.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development, 16*(4), 495–500.

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatrica, 102*(2), 156–160.

Moon, C. M., & Fifer, W. P. (2000). Evidence of transnatal auditory learning. *Journal of Perinatology: Official Journal of the California Perinatal Association, 20*(8 Pt 2), S37–44.

Moon, C., Zernzach, R. C., & Kuhl, P. K. (2015). Mothers say "baby" and their newborns do not choose to listen: A behavioral preference study to compare with ERP results. *Frontiers in Human Neuroscience, 9*, 153.

Movalled, K., Sani, A., Nikniaz, L., & Ghojazadeh, M. (2023). The impact of sound stimulations during pregnancy on fetal learning: A systematic review. *BMC Pediatrics, 23*(1), 183.

Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development, 21*(4), 779–784.

Niepel, D., Krishna, B., Siegel, E. R., Draganova, R., Preissl, H., Govindan, R. B., & Eswaran, H. (2020). A pilot study: Auditory steady-state responses (ASSR) can be measured in human fetuses using fetal magnetoencephalography (fMEG). *PLoS ONE, 15*(7), e0235310.

Novitskiy, N., Chan, P. H. Y., Chan, M., Lai, C. M., Leung, T. Y., Leung, T. F., Bornstein, M. H., Lam, H. S., & Wong, P. C. M. (2023). Deficits in neural encoding of speech in preterm infants. *Developmental Cognitive Neuroscience, 61*, 101259.

Ortiz Barajas, M. C., & Gervain, J. (2021). The role of prenatal experience and basic auditory mechanisms in the

development of language. In M. D. Sera & M. Koenig (Eds.), *Minnesota symposia on child psychology: Human communication: Origins, mechanisms, and functions* (Vol. 40, pp. 88–112). Wiley.

Ortiz Barajas, M. C., Guevara, R., & Gervain, J. (2021). The origins and development of speech envelope tracking during the first months of life. *Developmental Cognitive Neuroscience, 48*, 100915.

Partanen, E., Kujala, T., Näätänen, R., Liitola, A., Sambeth, A., & Huotilainen, M. (2013). Learning-induced neural plasticity of speech processing before birth. *Proceedings of the National Academy of Sciences of the United States of America, 110*(37), 15145–15150.

Peña, M., Pittaluga, E., & Mehler, J. (2010). Language acquisition in premature and full-term infants. *Proceedings of the National Academy of Sciences of the United States of America, 107*(8), 3823-3828.

Philbin, M. K. (2000). The influence of auditory experience on the behavior of preterm newborns. *Journal of Perinatology, 20*(S1), S77-S87.

Prechtl, H. F. R. (1984). *Continuity of neural functions from prenatal to postnatal life.* Blackwell Scientific Publications.

Schleussner, E., & Schneider, U. (2004). Developmental changes of auditory-evoked fields in fetuses. *Experimental Neurology, 190*, 59-64.

Shahidullah, S., & Hepper, P. G. (1994). Frequency discrimination by the fetus. *Early Human Development, 36*(1), 13-26.

Taymourtash, A., Schwartz, E., Nenning, K.-H., Sobotka, D., Licandro, R., Glatter, S., Diogo, M. C., Golland, P., Grant, E., Prayer, D., Kasprian, G., & Langs, G. (2023). Fetal development of functional thalamocortical and cortico-cortical connectivity. *Cerebral Cortex, 33*(9), 5613-5624.

Weiss, S. M., Aydin, E., Lloyd-Fox, S., & Johnson, M. H. (2024). Trajectories of brain and behaviour development in the womb, at birth and through infancy. *Nature Human Behaviour, 8*(7), 1251-1262.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49-63.

Wu, Y. J., Hou, X., Peng, C., Yu, W., Oppenheim, G. M., Thierry, G., & Zhang, D. (2022). Rapid learning of a phonemic discrimination in the first hours of life. *Nature Human Behaviour, 6*(8), 1169-1179.

Zhang, D., Chen, Y., Hou, X., & Wu, Y. J. (2019). Near-infrared spectroscopy reveals neural perception of vocal emotions in human neonates. *Human Brain Mapping, 40*(8), 2434-2448.

Zhang, D., Zhou, Y., Hou, X., Cui, Y., & Zhou, C. (2017). Discrimination of emotional prosodies in human neonates: A pilot fNIRS study. *Neuroscience Letters, 658*, 62-66.

From fetus to newborn: A continuity perspective on the development of speech perception

WANG Menghuan, XU Qiuyang, LIANG Dandan

(School of Chinese Language and Culture, Nanjing Normal University, Nanjing 210097, China)

Abstract: Human newborns exhibit remarkable sensitivity to their native language, maternal voice, and fine-grained speech details within hours after birth, and these capacities likely originate during the prenatal period. From a developmental neuroscience perspective, this article systematically reviews the low-pass-filtered acoustic environment of the womb and the maturation of the fetal auditory system, with a focus on how prenatal auditory experience shapes the neural mechanisms of speech perception in newborns. Existing evidence indicates that, within the low-frequency-dominated intrauterine soundscape, fetuses can perceive and remember native-language-related prosodic patterns, and that such experience may carry over to early postnatal speech processing. After birth, full-spectrum speech input and rapid cortical maturation jointly contribute to the development of segmental processing in newborns, with the underlying neural mechanisms involving activation of bilateral temporo-frontal networks and enhanced long-range temporal correlations in neural oscillations aligned with native-language prosody and phonemes. Building on existing perinatal continuity perspectives, this paper develops a speech-perception-specific prenatal-postnatal continuity framework, arguing that newborn speech perception develops continuously under the joint action of the maturation and functional emergence of the auditory system, prenatal low-frequency prosodic experience, and postnatal full-spectrum speech input. Future research should integrate multimodal neuroimaging approaches, strengthen longitudinal designs involving preterm and other high-risk populations, expand sampling in tone-language and multilingual contexts, and further explore translational applications in early screening and intervention for developmental language disorders.

Keywords: prenatal auditory experience, intrauterine acoustic environment, newborn speech perception, prosodic processing, neural mechanisms

Note: Figure translations are in progress. See original paper for figures.