Received: 13 June 2022 | Revised: 17 February 2023 | Accepted: 21 February 2023

DOI: 10.1111/desc.13387

Developmental Science WILEY

RESEARCH ARTICLE

Preliminary evidence for selective cortical responses to music in one-month-old infants

Heather L. Kosakowski[1,2,3] | Samuel Norman-Haignere[4] | Anna Mynick[5] | Atsushi Takahashi[1,2] | Rebecca Saxe[1,2,3] | Nancy Kanwisher[1,2,3]

[1]Department of Brain and Cognitive Sciences, Massachusetts Institute, of Technology, Cambridge, Massachusetts, USA

[2]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[3]Center for Brains, Minds and Machines, Cambridge, Massachusetts, USA

[4]Department of Neuroscience, University of Rochester, Rochester, New York, USA

[5]Psychological and Brain Sciences, Dartmouth College, Hannover, New Hampshire, USA

Correspondence
Heather L. Kosakowski, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA.
Email: heatherlkosakowski@gmail.com

Funding information
NIH, Grant/Award Numbers: #1F99NS124175, #8K00DA058542, #NIDCD-K99-DC018051, #NIDCD-R00-DC018051, #R21-HD090346-0, #DP1HD091947, S10OD021569; NSF STC award, Grant/Award Number: CCF-1231216

Part of the Special Issue "Music in Development", edited by Heather Bortfeld and Samuel Mehr.

Abstract

Prior studies have observed selective neural responses in the adult human auditory cortex to music and speech that cannot be explained by the differing lower-level acoustic properties of these stimuli. Does infant cortex exhibit similarly selective responses to music and speech shortly after birth? To answer this question, we attempted to collect functional magnetic resonance imaging (fMRI) data from 45 sleeping infants (2.0- to 11.9-weeks-old) while they listened to monophonic instrumental lullabies and infant-directed speech produced by a mother. To match acoustic variation between music and speech sounds we (1) recorded music from instruments that had a similar spectral range as female infant-directed speech, (2) used a novel excitation-matching algorithm to match the cochleagrams of music and speech stimuli, and (3) synthesized "model-matched" stimuli that were matched in spectrotemporal modulation statistics to (yet perceptually distinct from) music or speech. Of the 36 infants we collected usable data from, 19 had significant activations to sounds overall compared to scanner noise. From these infants, we observed a set of voxels in non-primary auditory cortex (NPAC) but not in Heschl's Gyrus that responded significantly more to music than to each of the other three stimulus types (but not significantly more strongly than to the background scanner noise). In contrast, our planned analyses did not reveal voxels in NPAC that responded more to speech than to model-matched speech, although other unplanned analyses did. These preliminary findings suggest that music selectivity arises within the first month of life. A video abstract of this article can be viewed at https://youtu.be/c8IGFvzxudk.

KEYWORDS
auditory cortex, fMRI, infants, music, speech

Research Highlights

- Responses to music, speech, and control sounds matched for the spectrotemporal modulation-statistics of each sound were measured from 2- to 11-week-old sleeping infants using fMRI.
- Auditory cortex was significantly activated by these stimuli in 19 out of 36 sleeping infants.

- Selective responses to music compared to the three other stimulus classes were found in non-primary auditory cortex but not in nearby Heschl's Gyrus.
- Selective responses to speech were not observed in planned analyses but were observed in unplanned, exploratory analyses.

1 | INTRODUCTION

Music and speech sounds are found across human societies (Brown & Jordania, 2013; Mehr et al., 2019). Our sophisticated ability to perceive pitch, beat, and phonemes emerges early (Baruch & Drake, 1997; Bertoncini et al., 1988; Chang & Trehub, 1977; Demany et al., 1977; Hannon & Trehub, 2005; Mehler et al., 1988; Plantinga & Trainor, 2009) and is fine-tuned with experience (Bergelson & Swingley, 2012; Gasparini et al., 2021; Hannon & Johnson, 2005; Hannon & Trehub, 2005a, 2005b). Adults have neural populations that are specialized for music and speech perception that are distinct from each other and from cortical responses to language meaning (Angulo-Perkins et al., 2014; Boebinger et al., 2021; Fedorenko et al., 2010, 2011, 2012; Leaver & Rauschecker, 2010; Mineroff et al., 2018; Norman-Haignere & Mcdermott, 2018; Norman-Haignere et al., 2015, 2022; Overath & Paik, 2021; Overath et al., 2015; Perani et al., 1996; Peretz & Coltheart, 2003). How do these music- and speech-selective responses emerge over development? To find out, we used functional magnetic resonance imaging (fMRI) to measure human cortical responses to music or speech shortly after birth.

1.1 | Human music and speech perception

Auditory signals are composed of frequencies that vary in amplitude and change over time at different rates. Throughout the earliest stages of acoustic processing, neurons are frequency-tuned and organized tonotopically (Graven & Browne, 2008). Neurons in primary auditory cortex (PAC) exhibit tuning for spectrotemporal changes which can be modeled using linear spectrotemporal receptive fields (STRFs) (reviewed in (Mcdermott, 2013; Norman-Haignere & Mcdermott, 2018)). Naturally occurring music and speech sounds have different spectrotemporal modulation statistics (Ding et al., 2017; Fernald & Kuhl, 1987; Grieser & Kuhl, 1988) that can result in the appearance of a category-selective response when not accounted for (Landemard et al., 2021). Thus, without careful controls, even adult PAC would respond differently to music and speech, not because of high-level category selective responses but because the stimuli evoke different patterns of activity in neural populations selective for different frequencies and spectrotemporal modulations. Yet several recent studies have shown that the music- and speech-selective neural populations in adult non-primary auditory cortex (NPAC) (Norman-Haignere & Mcdermott, 2018; Norman-Haignere et al., 2015, 2022; Perani et al., 1996; Peretz & Coltheart, 2003) cannot be explained by the spectrotemporal modulation statistics of music and speech: the response is weaker for synthetic sounds with spectrotemporal modulation statistics matched to music and speech (while the original and synthetic sounds evoke similar responses in PAC) (Leaver & Rauschecker, 2010; Norman-Haignere & Mcdermott, 2018). How do these selective neural populations develop?

1.2 | Music and speech perception in infancy

Shortly after birth, infants discriminate changes in melody (Chang & Trehub, 1977; Hannon & Trehub, 2005; Plantinga & Trainor, 2009), rhythm (Chang & Trehub, 1977; Demany et al., 1977), and tempo (Baruch & Drake, 1997). Cultural differences in music perception, such as sensitivity to culture-specific meter and rhythm, begin to emerge at about 6 months of age (Hannon & Johnson, 2005; Hannon & Trehub, 2005a, 2005b). A similar trajectory has been observed in speech perception (Kuhl, 2004; Werker & Gervail, 2012; Werker & Hensch, 2015): neonates reliably discriminate between all phonemes (Bertoncini et al., 1988; Mehler et al., 1988) and discriminate between languages from different rhythmical classes (Gasparini et al., 2021), but by the middle of the first year, infants lose the ability to discriminate between phonemes that are not distinguished in their language, and begin to ascribe meaning to the phonemic pairings of common nouns (Bergelson & Swingley, 2012). Taken together, behavioral evidence indicates that music and speech perception emerge early and are subsequently fine-tuned to culturally relevant acoustic features.

Activations of human auditory cortex (we use the acronym "AC" to refer to activations anywhere in primary or non-primary auditory cortex) have been reported as early as the third trimester in utero (Hykin et al., 1999; Jardri et al., 2012; Moore et al., 2001) and soon after birth (Anderson et al., 2001; Benavides-Varela & Gervain, 2017; Blasi et al., 2011; Bortfeld et al., 2007; Cristia et al., 2014; Dehaene-Lambertz et al., 2002, 2010; Giordano et al., 2021; Homae et al., 2006, 2007, 2012; Kotilahti et al., 2010; Lloyd-Fox et al., 2012; Nishida et al., 2008; Peña et al., 2003; Perani et al., 2010; Sato et al., 2010, 2012; Shultz et al., 2014; Telkemeyer et al., 2011) (though see (Anderson et al., 2001; Dehaene-Lambertz et al., 2002)). Thus, infants appear to have reliable cortical responses to sounds before or soon after birth. But do those responses include, or go beyond, processing frequency and spectrotemporal modulation statistics of sounds? Answering this question requires controlling for both the spectrotemporal properties of the stimulus, and for the presence of familiar high-level acoustic structure, like music and speech.

To our knowledge, only two studies have compared infants' cortical response to music (western piano music) to another meaningful sound category such as speech (thus controlling for the presence of high-level structure). Using functional near-infrared spectroscopy (fNIRS),

Kotilahti et al. (2010) observed activations to music that were significantly greater than activations to speech in some neonates, but this effect was not reliable across subjects. Similarly, in a group analysis of fMRI data from seven infants, Dehaene-Lambertz et al. (2010) found a response to both speech and music stimuli that was greater than the response to scanner noise alone but did not observe a differential response to music compared to speech or speech compared to music. Other studies have compared neural responses to music to responses to altered music (Perani et al., 2010; Wild et al., 2017) but these contrasts cannot distinguish selectivity for music from a response to familiar auditory structure. Taken together, these results do not answer whether music-selective responses are found in infant auditory cortex.

Findings concerning infants' neural response to speech have been more variable, with some studies reporting speech-sensitive responses (i.e., speech compared to another stimulus condition) in auditory (Benavides-Varela & Gervain, 2017; Fló et al., 2019; Lloyd-Fox et al., 2012; May et al., 2011; Minagawa-Kawai et al., 2010; Sato et al., 2012; Shultz et al., 2014), frontal language (Arimitsu et al., 2011; Gervain et al., 2008; Minagawa-Kawai et al., 2010; Sato et al., 2010; Shultz et al., 2014), or temporal language (Cristia et al., 2014; Dehaene-Lambertz et al., 2002; Gervain et al., 2008; May et al., 2018; Shultz et al., 2014) regions. Meanwhile, other studies fail to find any significant response to speech compared to another auditory category (Benavides-Varela & Gervain, 2017; Blasi et al., 2011; Bortfeld et al., 2007; Cristia et al., 2014; Dehaene-Lambertz et al., 2010; Kotilahti et al., 2010; Nishida et al., 2008; Peña et al., 2003). Also, it remains unclear whether apparent speech-sensitive responses in infants could be explained by differences in spectrotemporal modulation statistics between speech and other sounds (Minagawa-Kawai et al., 2011; Reybrouck & Podlipniak, 2019; Sato et al., 2010; Sulpizio et al., 2018; Telkemeyer et al., 2009, 2011).

Why might infants have cortical responses to music and speech that go beyond selectivity to their lower-level acoustic features such as spectrotemporal modulation statistics? One possibility is that infants' pre- and early post-natal auditory environment supports slow experience-dependent development such that cortical neurons learn important sound categories such as music and speech. This bottom-up view of cortical development might predict that AC of very young infants would have frequency-tuned neurons but not spectrotemporal tuning. On this view, sounds that have matched cochlear excitation patterns would evoke similar activations across PAC and NPAC (Figure 1a). Then, extensive auditory experience would drive neurons in AC to acquire specific spectrotemporal modulation statistics (Figure 1b). Finally, in the last stage of development, subpopulations of neurons in NPAC would develop category-selective responses (Figure 1c). At the other extreme, it is possible that from very early in infancy, populations of neurons already preferentially respond to music and speech sounds beyond the spectrotemporal modulation statistics of these sounds (Figure 1c). The key difference between these hypotheses is the extent to which the neural tuning to music and speech perception is gradually constructed from auditory experience or whether instead it arises with little or no instructive role from experience.

1.3 | Current experiment

To identify selective responses to music and speech in infants, we created stimuli that are matched on spectrotemporal modulation statistics. A key contribution of the current project is the creation and dissemination of these stimuli. Our choice of speech and music stimuli for this study arose from the following rationale. Speech sounds are emitted from a single source (i.e., vocal cords) while piano music (used in prior studies that compared music and speech in infants (Dehaene-Lambertz et al., 2010; Kotilahti et al., 2010)) emits sound from multiple sources (i.e., many keys strike strings simultaneously). Although the evolution of music remains a mystery (Akkermann et al., 2021; James, 1890; Lieberman & Billingsley, 2021; Mcdermott, 2008; Mehr et al., 2021; Pinker, 1997; Savage et al., 2021; Trevor & Frühholz, 2021), vocal music likely emerged earlier in human history than instrumental music[1] and perhaps even earlier than speech (Montagu, 2017). Flutes, like untrained vocalists, emit a steady tone from a single source and are used in modern, remote cultures (Jacoby et al., 2019). Thus, we decided to use simple monophonic instrumental melodies as music stimuli, which likely reflect evolutionarily relevant properties of music without the confound of speech-like sounds entailed in singing. To create these music stimuli, we recorded lullabies, a universal category of infant-directed human song (Mehr et al., 2018, 2019). Selected lullabies (Feierabend, 2000) (Figure 2a) were played on instruments that emitted sound from a single source (see Section 2) in a similar frequency range as female infant-directed speech (IDS) (Fernald & Kuhl, 1987; Grieser & Kuhl, 1988). For speech stimuli, we recorded novel instances of female IDS (Figure 2b). All music and speech stimuli were 18 s long, with a clear onset and offset. In order to prevent neural habituation (Dehaene-Lambertz et al., 2002), we used 60 unique music recordings and 60 unique speech recordings.

Music and speech recordings differed in their overall frequency and amplitude (Figure S1), so we used a novel algorithm that matched the time-averaged cochleograms of music and speech stimuli (see Section 2) while keeping the high-level structure of each stimulus. This procedure matches the average spectral content of the cochleagram but does not control for differences in the rate of spectrotemporal modulations that occur within the cochleagram. To control for spectrotemporal differences, we synthesized a "model-matched" stimulus for each music and speech stimulus that was designed to have the same spectrotemporal modulation statistics as the original stimulus (Chi et al., 2005) (Figure 1a,b; Norman-Haignere & Mcdermott, 2018) and was thus expected to produce the same response in PAC but not NPAC in adults (Norman-Haignere & Mcdermott, 2018). In practice, the perception of these spectrotemporally matched synthetic sounds differs markedly from their natural counterparts (Norman-Haignere & Mcdermott, 2018), indicating that modulation statistics fail to capture perceptually important features of natural speech and music, despite producing similar responses in adult PAC. This experimental design allowed us to test three different theoretical predictions. First, infant AC neurons might have frequency tuning but may not have acquired STRF tuning, which would predict that PAC and NPAC have similar responses to all four stimulus categories (Figure 2a). Second, the

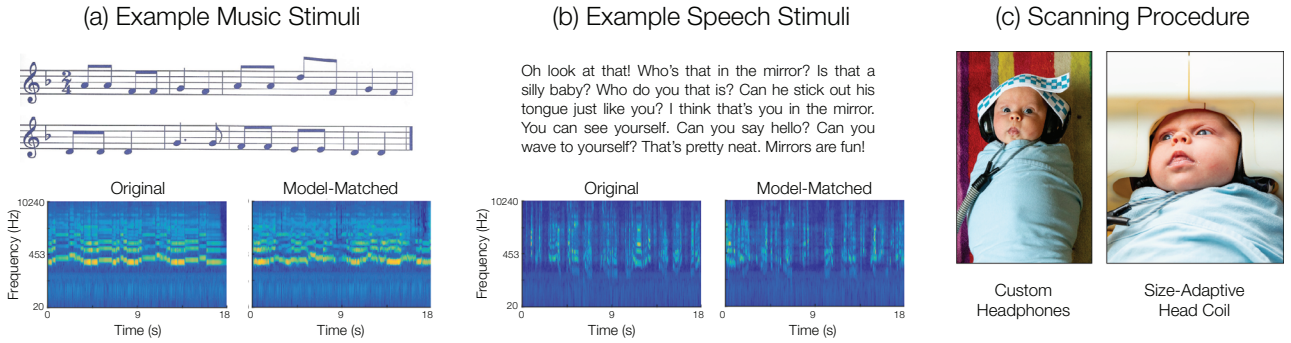

FIGURE 1 Infant fMRI protocol. (a) An example musical score from (Feierabend, 2000) (top) along with a cochleogram for the original (bottom left) and model-matched music stimulus (bottom right). (b) An example speech stimulus (top) and an example cochleogram for the original (bottom left) and corresponding model-matched speech stimulus (bottom right). (c) Infants were swaddled and custom MR-safe headphones that provided auditory stimulation were applied (left). Then, infants were placed in a custom, size-adaptive infant head coil (Ghotra et al., 2021) (right).

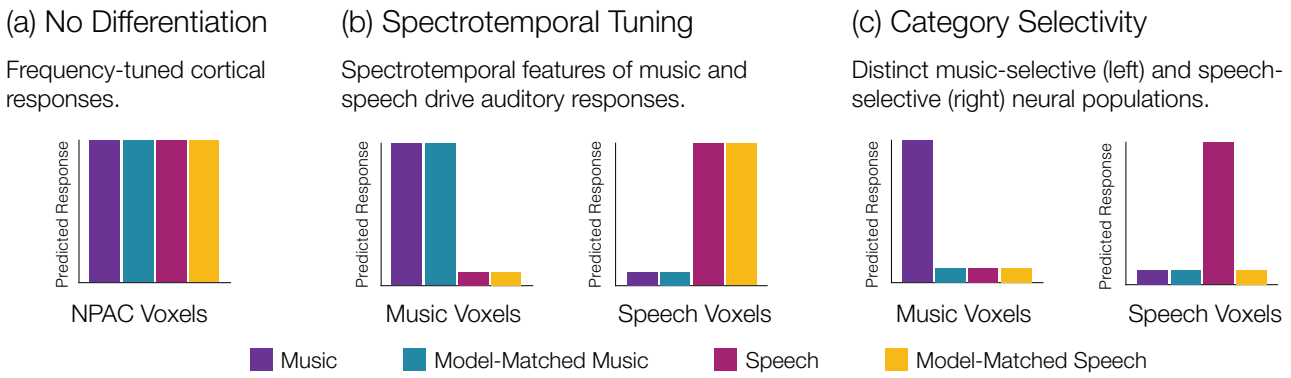

FIGURE 2 Alternative hypotheses of functional organization of infant auditory cortex. (a) NPAC organization might be dominated by selectivity to sound energy at different frequencies and thus show similar responses to all four conditions. (b) Different regions within NPAC might be tuned to spectrotemporal statistics found in music and speech, and thus show distinct responses to music and speech but similar responses to synthetic sounds with matched spectrotemporal statistics. (c) NPAC may have music-selective responses and/or speech-selective responses that reflect higher-order properties (e.g., notes, syllables) not explained by modulation selectivity.

response to music and speech in infant NPAC could be driven by the overall spectrotemporal modulation statistics of the stimuli, resulting in a similar response to original and model-matched pairs, but different responses to music and speech (Figure 2b). A third possibility is that infant NPAC has populations of neurons with higher-order category selectivity for music or speech that is not explained by spectrotemporal modulation statistics (Figure 2c).

To test these alternative hypotheses, we first sought to determine if our stimuli would elicit music- and speech-selective responses in adult NPAC (see Section 2). This is an important first step as music-selective neural populations are less spatially clustered in NPAC (Boebinger et al., 2021; Norman-Haignere et al., 2015, 2022; Tierney et al., 2013) than speech-selective neural populations (Boebinger et al., 2021; Leaver & Rauschecker, 2010; Norman-Haignere et al., 2015, 2022; Overath & Paik, 2021; Overath et al., 2015; Tierney et al., 2013) and are therefore harder to detect with standard voxel-wise contrasts (Angulo-Perkins et al., 2014; Fedorenko et al., 2012; Leaver & Rauschecker, 2010). Once we validated our stimuli in adults, we collected and analyzed fMRI data (Figure 2c) from 45 sleeping infants (2.0–11.9 weeks) while we played them music, model-matched music, speech, and model-matched speech sounds. Previous research has primarily investigated infants' auditory responses while they are asleep, with mixed success. Thus, a first question is the degree to which our stimuli activate AC. We predicted that each stimulus category would elicit a significant, and similar, response (compared to scanner noise alone). Second, we predicted that if spectrotemporal modulation statistics are sufficient to explain the response to music, voxels selected to have a greater response to music than speech would have a similar response to model-matched music (Figure 2b). Alternatively, if infants have music-selective responses, the response to music will be greater than the response to both speech and model-matched music (Figure 2c). A similar prediction for speech follows—if spectrotemporal modulation statistics are sufficient to explain auditory responses to speech, then the response to speech and model-matched speech should be similar (Figure 2b), but if speech responses are selective for higher-order speech features (e.g., phonemes, syllables) than the response to speech will be greater than the response to music and model-matched speech (Figure 2c). Finally, previous research with

infants has reported speech preferences in cortical areas that are language-selective in adults (i.e., respond selectively to written and auditory language) (Fedorenko et al., 2011, 2012). Thus, we also ask if adults' language-selective regions exhibit a speech preference in sleeping infants.

2 | METHODS

2.1 | Subject information

Participants were recruited through Boston Children's Hospital and Massachusetts Institute of Technology and received a gift card. We recruited 50 infants (2.0 to 13.1 weeks old; 21 female, 9 Asian, 3 African American, 31 White, 7 unknown; 4 Hispanic/Latino, 45 not Hispanic/Latino, and 1 unknown) and collected data from 45 infants (2.0 to 11.9 weeks-old; 17 female), 36 of whom had at least 144 volumes of data (see "subrun creation" in Methods for criteria) and 19 of whom had significant auditory activations.

2.2 | Paradigm information

Music recordings. Lullabies are traditionally sung by a caregiver (i.e., a single sound source, non-professional musician) and are soothing for infants (Bainbridge et al., 2021). We selected instruments that emit sound from a single source and have a frequency range that overlapped the mean fundamental frequency of American female IDS (267 Hz) (Fernald & Kuhl, 1987; Grieser & Kuhl, 1988). Professional musicians were recruited from the Boston area and paid $60 per hour to play selected melodies from The Book of Lullabies (Feierabend, 2000). To ensure melodies would start and end within 18 s, the scores of selected melodies were shorted or lengthened by excluding or adding bars and adjusting the tempo. As none of the hired musicians were parents, we provided instructions to enhance naturalistic lullaby performance by (1) placing pictures of babies at strategic locations in the recording booth, (2) instructing musicians to play without vibrato, and (3) instructing musicians to play as if trying to lull a baby to sleep. All music stimuli were recorded in a double-walled sound booth. Recordings were normalized and exported using Audacity (https://www.audacityteam.org/).

Speech recordings. Speech stimuli were designed to be as naturalistic as possible. To ensure speech content that was consistent with utterances typical parents from the northeastern United States use with their infants, we recorded mothers as they spoke to their babies. These recordings did not make ideal stimuli in their original form because they contained unrelated sounds and non-uniform silent periods. However, from the recordings we created 37 written vignettes that approximately matched the content and repetition in the original recordings. As onsets and offsets are important acoustic features for AC (Norman-Haignere et al., 2015), we additionally ensured each speech stimulus had a clear onset and offset. We recruited three mothers to say each vignette aloud. To ensure each stimulus was as naturalistic as possible, had a clear onset and offset, and was 18 total seconds, words or phrases within individual vignettes were added or removed to match the specific speech cadence of each mother. While recording speech stimuli, mothers were instructed to look at a photo of their baby and recite each speech vignette as if talking to their baby. All speech stimuli were recorded in the same double-walled sound booth as the music recordings. Recordings were normalized and exported using Audacity (https://www.audacityteam.org/). We randomly selected a subset of 60 speech recordings (20 from each of three speakers) for the final speech stimulus set.

Music and speech stimuli. Music sounds are more spectrally heterogenous than speech sounds (Figure S1a). To increase the spectral homogeneity of the music stimuli, we first computed the mean excitation pattern for each music stimulus by averaging its cochleagram across time (Mcdermott & Simoncelli, 2011) (Figure S1a). We then computed the standard error between the excitation pattern for each music stimulus and the average excitation across all music stimuli (i.e., the standard deviation across cochlear frequencies for the difference vector between the mean excitation and the excitation of a given stimulus, divided by the square root of the number of frequencies). The stimulus for the excitation pattern with the greatest standard error was removed from the set. We iterated through this process, greedily discarding the stimulus with highest standard error, until we had 60 music stimuli (out of an original 275), which included recordings from cello, flute, clarinet, and violin.

After increasing the spectral homogeneity of the music stimuli there were still differences in the average excitation pattern between music and speech stimuli (Figure S1b). To reduce these differences, we matched the average excitation patterns between the two stimulus sets. Specifically, we computed cochleagrams for each stimulus and then separately scaled the magnitude of each frequency channel in the cochleagram such that the average magnitude across time and stimuli for each category was equal to the grand mean across both categories (for that frequency). We then reconstructed waveforms from these modified cochleagrams using standard procedures described previously (Mcdermott & Simoncelli, 2011; Norman-Haignere & Mcdermott, 2018). Because the cochlear filters overlap, the reconstructed waveform was not fully consistent with the desired cochleagram (i.e., the cochleagram that has been rescaled to match the mean excitation pattern across categories). To increase the match between the measured and intended cochleagram, we iteratively measured cochleagrams and resynthesized waveforms 10 times. This procedure was successful in generating stimuli with closely matched average excitation patterns (Figure S1c) that were perceptually very similar to the original music and speech recordings.

Model-matched music and speech. Our excitation-matched music and speech stimuli still differ along many acoustic properties, such as spectrotemporal modulation statistics, which have been shown to drive fMRI responses in adults' PAC. A higher neural response to music than speech or vice versa could thus reflect differences in modulation statistics between speech and music. We therefore created two new control conditions with synthetic stimuli that were matched to the music and speech in both cochlear and spectrotemporal modulation statistics. We used a model and synthesis procedure that has been shown to yield stimuli with closely matched fMRI responses in

PAC (Norman-Haignere & Mcdermott, 2018). Following the procedure of Norman-Haignere and Mcdermott (2018), the synthesis algorithm starts with an unstructured noise stimulus, and iteratively modifies the noise stimulus to match the modulation statistics of a natural sound. The synthesis procedure alters the noise stimulus to match the histogram of response magnitudes across time for each filter in the model, which has the effect of matching all time-averaged statistics (such as mean and variance) of the filter responses. Although the synthetic stimuli have closely matched spectrotemporal modulation statistics, they lack higher-order properties that are poorly captured by modulation statistics. The resulting model-matched stimuli perceptually sound very different from natural speech or music to adults (Norman-Haignere & Mcdermott, 2018). Example stimuli are available online (heatherkosakowski.com/stimuli/) and the full stimulus set is available on OSF (https://osf.io/8ty34/).

Experimental paradigm. Infants listened to the 18 s clips of music, speech, model-matched music, and model-matched speech, which were presented consecutively in sets of four such clips (one per stimulus category), in a random order within each set, followed by a block of silence. For adults and the first 10 subjects, silent blocks were 12–17 s. Each run had fifteen 18-s clips of each of the four conditions. Adults were instructed to press a button at the end of each clip. Three out of four adults heard every stimulus; the remaining adult (who was highly familiar with the stimulus set) did not finish all experimental runs. After our new headphones were designed and we resumed data collection, we increased the silent blocks to 24 s to increase the measurement quality of the silent blocks.

2.3 | Data collection and processing

Hearing protection and sound presentation. The first 10 infants were scanned with headphones from MR Confon (http://www.mr-confon.de/). Once we discovered that the headphones did not produce the manufacturer-reported attenuation at 1 kHz, we discontinued use of the headphones. Subsequent subjects were scanned with headphones that were designed in collaboration with Sensimetrics (www.sens.com). Newly designed infant headphones featured a speaker that sat in infants' concha and was adhered to infants' ears using neonatal noise guards (shorturl.at/lrsV9). Infant-sized earmuffs (Ems-for-Kids Baby Earmuffs; earmuffsforkids.com) were placed over noise guards and held in place with a cloth and Velcro. If muffs had metal screws, they were replaced with plastic screws. Additional custom pillows made with foam and material commonly used for diapers (e.g., PUL or TPU) were used when necessary.

Scanning session. Infants were swaddled and rocked to sleep by a parent or researcher. After the infant fell asleep, headphones were applied, and the infant was placed in the head coil. All sounds were presented at 75 dB, a comfortable listening level for infants. A researcher stood outside the scanner for the duration of the scanner session. Caregivers were offered the option to go into the scanner with the infant. The session ended when infants woke up or the parent requested to end the session.

Acquisition parameters. Adult fMRI data were collected with a Siemens 3T Prisma scanner using a 32-channel head coil and an EPI with standard trajectory with 46 near-axial slices (repetition time, TR = 2 s, echo time, TE = 30 s, flip angle = 90º, field of view, FOV = 208 mm, matrix = 104 × 104, slice thickness = 2 mm, slice gap = 0.2 mm). Infant fMRI data were collected on the same Siemens 3T Prisma scanner using an EPI with standard trajectory with 52 near-axial slices (repetition time, TR = 2 s, echo time, TE = 30 ms, flip angle = 90º, field of view, FOV = 208 mm, matrix = 104 × 104, slice thickness = 2 mm, slice gap = 0 mm). Data were collected from the first 21 subjects using an adult 32-channel head coil and from the remaining subjects using a custom infant-specific head coil (Ghotra et al., 2021).

Data selection. To be included in the analysis, data had to meet criteria for low head motion. Data were cleaved between consecutive timepoints having more than 2° or 2 mm of frame-to-frame displacement, creating *subruns*, each of which contained at least 48 consecutive low-motion (less than 2°/mm of motion) volumes. These motion exclusion criteria are similar to previously reported thresholds (Deen et al., 2017; Kosakowski et al., 2022). All volumes included in a subrun were extracted from the original run data and combined to create a new Nifti file for each subrun. Paradigm files were similarly updated for each subrun. Within each subrun, volumes with greater than 0.5°/mm of motion between volumes were scrubbed. As in our previous research, if more than three consecutive images were scrubbed, there had to be at least seven consecutive low-motion volumes following the scrubbed volume in order for those volumes to be included in the analysis (Kosakowski et al., 2022). Each subrun had to have at least 48 volumes after accounting for motion. To be included in the group random effects and overlap analyses, subjects had to have at least 144 volumes, which ensures inclusion of some data from all four conditions and rest blocks.

Preprocessing and Data image registration. We followed standard protocols for pre-processing and data image registration as has been reported in (Deen et al., 2017; Kosakowski et al., 2022). For details, see Supplemental Materials.

Auditory parcels. Precise tonotopic mapping of primary auditory cortex in infants has never been conducted. Thus, we used several strategies (fully described in Table S1) before settling on these parcel definitions. We used three anatomical parcels to constrain our selection of functionally defined voxels. The first parcel was constrained to Heschl's Gyrus (HG) using a common adult atlas computed in volume space. Note that PAC in adults extends well beyond HG (e.g., planum temporale), as measured based on tonotopic criteria (Da Costa et al., 2011; Humphries et al., 2010; Norman-Haignere et al., 2015), but HG is nonetheless a standard anatomical landmark of PAC. The second parcel corresponded to broader AC, including HG and NPAC, and was defined using Glasser parcels A1, LBelt, Pbelt, and Mbelt transformed from surface space to volume space (Figure 3). A third parcel was created by removing voxels that corresponded to HG from the AC parcel, which we refer to as the NPAC parcel. We further note that our ability to localize functional responses in infants was limited because our registrations heavily relied on low-resolution functional data and because registration between infants and adults is necessarily imperfect.

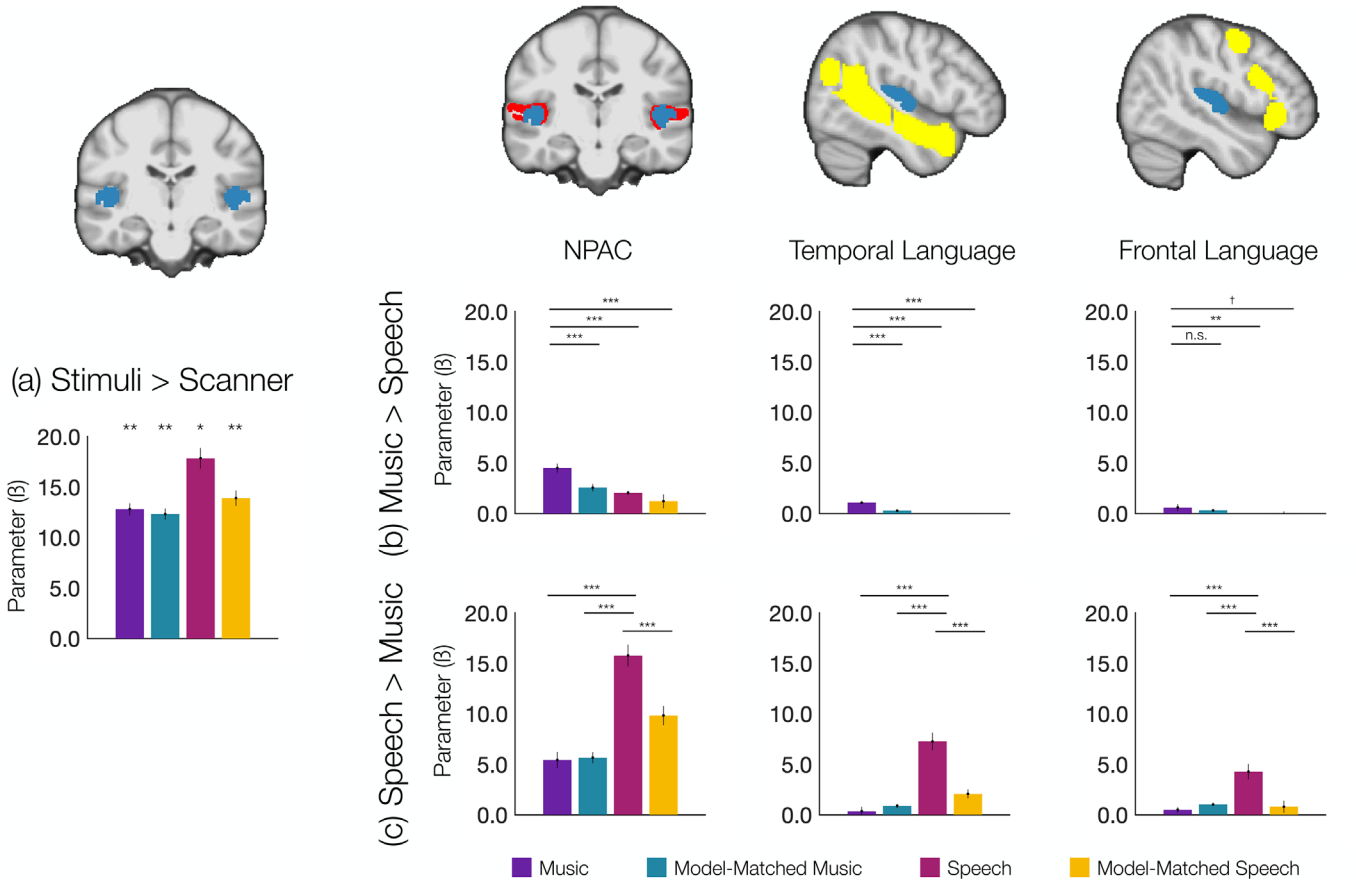

FIGURE 3 Responses to all stimuli in adult auditory and language cortices. (a) Response in independent, left-out data for the top 5% of voxels that respond more to all stimuli than scanner noise in auditory cortex (HG, highlighted in blue). Response in independent, left-out data for the top 5% of music-selective (b) and speech-selective (c) voxels in non-primary auditory cortex (NPAC) (left, highlighted in red), temporal language (middle, highlighted in yellow), and frontal language (right, highlighted in yellow). Bar graphs depict average response to music (purple), model-matched music (teal), speech (pink), and model-matched speech (yellow). Error bars indicate within-subject SE (Cousineau, 2005). Symbols used to report one-tailed statistics from linear mixed effects models: $^{n.s.}p > 0.1$; $^{\dagger}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Language parcels. Six language-selective anatomical constraint parcels were obtained from previous studies with adults (Fedorenko et al., 2010). The three posterior regions were combined into a single parcel (temporal language, Figure 3) and the three frontal regions were combined into a single parcel (frontal language, Figure 3). We removed any voxel from the lateral language parcel that overlapped with auditory parcels.

Infant parcels. Adult auditory and language parcels were transformed to infant template space. First, a representative infant anatomical image was registered to the adult MNI template using an affine transformation followed by hand-tuning. Then parcels were inverse transformed to the template infant functional image.

2.4 | Quantification and statistical analysis

Subject-level beta and contrast maps. Functional data were analyzed with a whole-brain voxel-wise general linear model (GLM) using custom MATLAB scripts. The GLM included four condition regressors (speech, music, model-matched speech, model-matched music), six motion regressors, a linear trend regressor, and five PCA noise regressors using standard methods reported in (Kosakowski et al., 2022) and (Deen et al., 2017). Condition regressors were defined as a boxcar function for the duration of the stimulus presentation. Boxcar condition regressors were convolved with an infant hemodynamic response function (HRF) that is characterized by a longer time to peak and deeper undershoot compared to the standard adult HRF (Arichi et al., 2012). Timepoints that exceeded the specified motion threshold (see above) were removed prior to model fitting. Data and regressors were demeaned for each subrun and concatenated across subruns. Beta values for each condition were computed in a whole-brain voxel-wise GLM.

Group random effects analysis. Freesurfer's mri_concat, mri_glmfit, and mri_volcluster were used for these analyses (Nichols et al., 2005). Briefly, contrast maps from each subject were concatenated and fit with a voxel-wise GLM. To correct for multiple comparisons, we used Monte Carlo simulation and randomly changed the sign of the contrast value for 5000 iterations to create a null distribution (i.e., we randomly flipped the sign for every contrast image for each subject and then averaged across subjects). From this null distribution, we recorded the

maximum cluster size for each sample and then tested if the measured cluster size exceeds this maximum for 95% of samples.

Functional region of interest (fROI) analyses. To measure cortical responses, we used a functional region of interest (fROI) approach. The strength of an fROI approach is that it allows us to quantify neural responses in functionally distinctive regions that are not perfectly aligned across participants in anatomical coordinates (Saxe et al., 2006). Further, fROI analyses do not rely on the high-quality image acquisitions necessary for inter-subject registration, which is essential given twe were often unable to obtain high-resolution anatomical images from individual infants. Due to the variable amount of data in each subrun for each subject and the impact this could have on reliable parameter estimates from the GLM, we first combined or split subruns to approximately equate the amount of data across subruns within subjects. For example, if a subject had three subruns and the first subrun was 111 volumes, the second was 57 volumes and third was 350 volumes, the first two subruns would be concatenated to create one subrun and the third subrun would be split in two resulting in a total of three subruns with approximately 175 volumes per subrun. After splitting/concatenating subruns, subjects had to have at least 144 volumes per subrun and at least two subruns (one for data selection, one for data extraction) to be included in fROI analyses.

Voxel selection and beta extraction. To constrain search areas for voxel selection, we used anatomically defined parcels (see parcels/search spaces) transformed from infant template space to subject native space. We used an iterative, leave-one-subrun-out cross-validation procedure to estimate fROI responses. Data were concatenated across all subruns except one prior to computing whole-brain voxel-wise GLMs and contrasts were computed (described above). We then selected the 5% of voxels with the greatest difference in beta weights for the contrast of interest (i.e., stimuli > scanner, music > speech, or speech > music) within an anatomical constraint parcel in subject-specific functional space. Beta values in these voxels were extracted for all four conditions from the left-out subrun. We iterated through this process for all data partitions and subjects, averaging across the betas derived from each fold in each subject. Results for all fROI analyses were generated using this method.

Auditory cortex responses. The 5% of voxels within AC that had the greatest numerical contrast value for stimuli > scanner noise were selected for analysis. To determine if each stimulus category significantly activated AC, we compared the response to each condition to zero using a *T*-test.

Subject selection for music and speech analyses. We next sought to determine if individual subjects could hear the stimuli. There are two reasons to be concerned subjects were unable to hear the stimuli: (1) the headphone set-up and (2) effects of sleep on auditory responses. With respect to the headphone set-up, the speaker is placed in the concha but, it is possible it could shift and thus not adequately stimulate AC. A second possibility is variable activation of AC during different sleep stages. As we cannot distinguish between these possibilities in our dataset, we measured infants' AC response to all sounds compared to scanner noise in NPAC, an analysis that is orthogonal to our primary analysis. To determine if the response in AC was reliable in each subject, and not due to noise, we conducted an fROI analysis. Specifically, we selected the top 5% of voxels in AC for the stimuli > scanner contrast and extracted the beta values from independent data. For each subject we determined if the average response to all four conditions was significantly greater than the response to scanner noise (i.e., zero) using a *T*-test with an alpha of 0.05.

Music and speech selectivity. For the music fROI we selected the top 5% of voxels for music > speech within a parcel. For the speech fROI we selected the top 5% of voxels for speech > music within a parcel. Importantly, the response to model-matched sounds was not included in the voxel selection criteria. Betas from all four conditions were extracted from independent data in the selected voxels. To determine if a response was selective, we used a linear mixed effects (LME) model, which enabled us to account for between subject variability due to motion. Betas extracted from independent data for each individual subject were the dependent variable. We elected to have the condition of interest (e.g., music in the music > speech fROI anlaysis) be the un-modelled beta in the regression, which enabled us to test whether the response to each control condition (e.g., model-matched music, speech, and model-matched speech in the music > speech fROI analysis) was significantly lower than the response to condition of interest. Thus, we had one predictor for each control condition (e.g., in the case of the music > speech fROI analysis, there were three vectors—one for speech, one for model-matched music, and one for model-matched speech). In the infant LMEs, predictors of no interest were age, motion, and sex. Motion was computed as the number of scrubbed volumes divided by the total number of volumes. Subject was coded as a random effect with random intercepts (results were not different when random slopes were added (Barr et al., 2013)) for each control condition. We fit a model with the MATLAB expression:

```
fitlme(wß~c1+c2+c3+age+sex+motion+(1+c1|subject)+(1+c2
|subject)+(1+c3|subject))
```

where ß indicates the betas for all four conditions from each subject. The control condition vectors (c1, c2, c3) have 1 s in the location of the control condition that is represented and 0 s at all other locations. For a response to be "selective" we expected a significantly positive intercept (indicating the response to the condition of interest was greater than the response to scanner noise) and the response to each control condition to be significantly negative (indicating the response to the control condition was significantly less than the response to the condition of interest). As predictions are unidirectional, reported *p*-values are one-tailed.

Repeated measures ANOVAs were used to test for interactions between regions. All ANOVAs were computed using open-source JASP software (https://jasp-stats.org) (Goss-Sampson, 2022).

3 | RESULTS

3.1 | Auditory activations in awake adults

We first asked if adults (n = 4) would exhibit music- and speech-selective responses in NPAC. Age, sex, and motion were excluded from adult models. In non-primary auditory cortex (NPAC), we selected

TABLE 1 Responses in adult auditory and language cortices.

Parcel	Contrast	Intercept[a]	Music[a]	mmMusic[a]	Speech[a]	mmSpeech[a]
NPAC	Music > Speech	**4.51**	n/a	**−1.95**	**−2.46**	**−3.29**
		(2.39; 6.62)		(−2.79; −1.10)	(−3.32; −1.60)	(−4.78; −1.80)
	Speech > Music	**15.79**	**−10.36**	**−10.13**	n/a	**−5.95**
		(1.29; 30.29)	(−13.30–7.42)	(−12.45; −7.81)		(−8.66; −3.24)
T. Lang.	Speech > Music	**7.27**	**−6.92**	**−6.39**	n/a	**−5.20**
		(5.94; 8.61)	(−8.54; −5.30)	(−7.82; −4.96)		(−6.82; −3.58)
F. Lang.	Speech > Music	**4.27**	**−3.78**	**−3.24**	n/a	**−3.47**
		(3.02; 5.53)	(−5.08; −2.47)	(−4.47; −2.02)		(−5.21; −1.73)

[a]Parameter estimates from a linear mixed-effects model for the intercept and three control condition regressors. In the music > speech fROI analysis, the intercept indicates the effect size of the music response relative to scanner noise, the mmMusic value indicates the effect size of the model-matched music response relative to music, the speech value indicates the effect size of the speech response relative to music, and the mmSpeech value indicates the effect size of the model-matched speech response relative to music. In the speech > music fROI analysis, the intercept indicates the effect size of the speech response relative to scanner noise, the music value indicates the effect size of the music response relative to speech, the mmMusic value indicates the effect size of the model-matched music response relative to speech, and the mmSpeech value indicates the effect size of the model-matched speech response relative to speech. Confidence intervals are reported in parentheses and effect sizes with $p < 0.05$ are indicated in bold.

the top 5% of voxels that responded more to music than speech. In independent, left-out data the response to music was significantly greater than the response to each of the other conditions (Figure 3; Table 1; all *ps* < 0.0003). For the top 5% of voxels in NPAC that responded more to speech than music, the response in independent data (Figure 3; Table 1) was significantly greater to speech than each of the other conditions (all *ps* < 0.0003). Thus, these stimuli produce robust music- and speech-selective responses in adult NPAC.

Next, we sought to confirm that adult language parcels would respond selectively to speech but not music. Using the same procedure described previously, we identified the top 5% of voxels that responded more to speech than music and then, from these voxels, extracted the response to all four conditions from independent, left-out data. In both the temporal and frontal language parcels, we observed a response to speech that was significantly greater than the response to each of the other conditions (all *ps* < 0.0005; Table 1). Thus, adults have music-selective responses in NPAC and speech-selective responses in NPAC and language areas.

3.2 | Responses to sound in the auditory cortex of sleeping infants

Results from previous literature are mixed as to whether infants exhibit responses to auditory stimuli while they are asleep (Anderson et al., 2001; Benavides-Varela & Gervain, 2017; Blasi et al., 2011; Bortfeld et al., 2007; Cristia et al., 2014; Dehaene-Lambertz et al., 2002, 2010; Giordano et al., 2021; Homae et al., 2006, 2007, 2012; Kotilahti et al., 2010; Lloyd-Fox et al., 2012; Nishida et al., 2008; Peña et al., 2003; Perani et al., 2010; Sato et al., 2010, 2012; Shultz et al., 2014; Telkemeyer et al., 2011). So, we first conducted a group random effects analysis of stimuli > scanner across all infants that had sufficient data (*n* = 36; 1.9–11.7 weeks). This revealed auditory activations at the posterior-medial portion of HG in the left hemisphere that did not survive correction for multiple comparisons (Figure 4a). In maps from individual infants, we observed some infants with bilateral auditory activations and some infants with no auditory activations (Figure 4b). An fROI analysis of stimuli > scanner across all subjects revealed significant auditory activations in AC (Figure 5a). Importantly, in independent data from AC, each stimulus condition produced a significant response relative to scanner noise (Figure 5b; music $t(35) = 3.66$, ci = 0.19–0.67, $p = 0.000$; mmMusic $t(35) = 3.50$, ci = 0.20–0.76, $p = 0.001$; speech $t(35) = 3.28$, ci = 0.20–0.86, $p = 0.002$; mmSpeech $t(35) = 2.21$, ci = 0.03–0.68, $p = 0.03$).

As a proxy for ensuring infants heard the stimuli while in the scanner, we decided a priori to only include infants in any analysis of music or speech selectivity if they had significant activations in AC in an fROI analysis when comparing the response to all stimuli to scanner noises. Of the 36 infants that had sufficient data, 19 had significant auditory activations (1.9–11.7 weeks, mean = 5.2 weeks) while 17 did not (2.0–11.5 weeks, mean = 4.6 weeks; Figure 5a). Age (sig. AC mean = 5.18 weeks, n.s. AC mean = 4.59 weeks; $t(34) = 0.80$, ci = −0.90 to 2.08, $p = 0.4$) and motion (measured as the proportion of volumes > 0.5 mm of translation or 0.5° of rotation) were similar between the two groups (sig. AC mean = 0.03, n.s. AC mean = 0.05, $t(34) = -1.26$, ci = −0.05 to 0.01, $p = 0.22$). There was a marginal difference in the quantity of data between the two groups such that the group that had significant auditory activations had less data (mean = 926.9 volumes) than the group with non-significant auditory activations (mean = 1335.4 volumes; $t(34) = -1.85$, ci = −856.9–39.94, $p = 0.07$). Thus, the absence of significant activation in AC is not due to insufficient data. Further, 5 out of 10 infants with data collected using the adult coil and 14 out of 26 infants with data collected using the infant coil had significant auditory responses, so the absence of significant activation of AC is not due to the type of coil used. All pre-planned analyses reflect results from the 19 infants with reliable activations in AC.

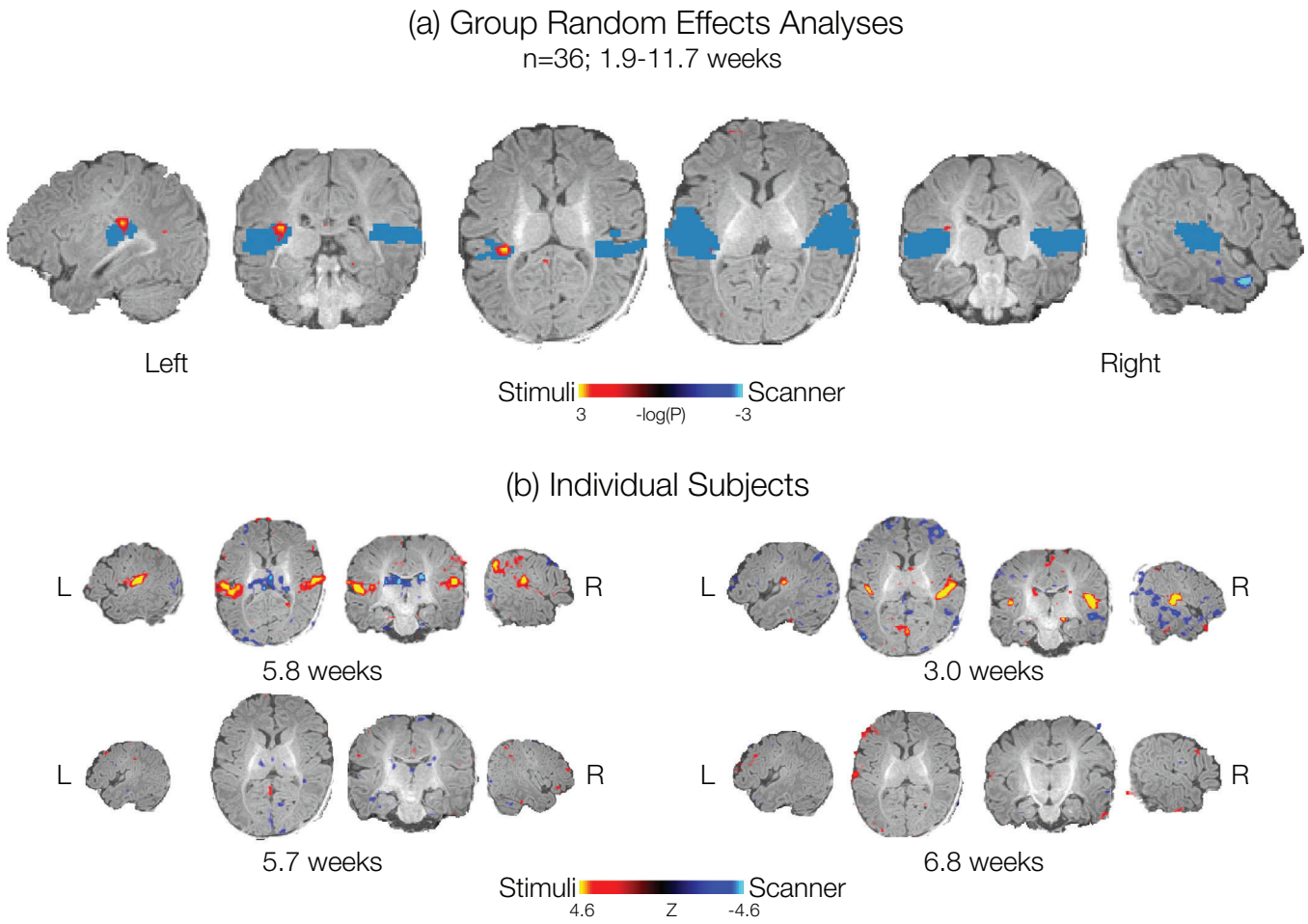

FIGURE 4 Auditory activation in infant cortex. (a) Group random effects analysis of stimuli > scanner noise across all subjects (threshold $p < 0.01$, uncorrected). Auditory cortex (AC) parcel (Heschl's Gyrus and non-primary auditory cortex combined) is highlighted in blue (b) Whole-brain statistical maps from individual infants (threshold $Z = 2.3$) that show clear activations in AC (top row) and absence of activations (bottom row). Statistical maps displayed on a non-representative, high-resolution infant anatomical image.

3.3 | Response to music and speech in infants with significant auditory responses

Given that we could detect auditory responses in sleeping infants, we next asked whether music or speech elicit different responses in those infants who showed significant auditory responses. In a group random effects analysis of music > speech, music and speech responses in auditory cortex were not significant after correcting for multiple comparisons. However, we observed significant music activations in ventral-temporal, ventral-prefrontal, and dorsal-prefrontal cortices and significant speech activations in cerebellum, occipto-parietal cortex, and somatomotor cortex (Figure S2). We did not further interrogate activations outside of auditory and language cortices. In individual infants, whole-brain maps revealed music and speech responses in both hemispheres within auditory cortices (Figure 6).

Next, in the pre-planned fROI analysis of subjects with significant auditory activations ($n = 19$), we asked if these subjects have a music-selective response in NPAC. First, from NPAC we selected the top 5% of voxels that responded more to music than speech. In independent, left-out data from these voxels, the response to music was significantly greater than the response to speech ($p = 0.03$) and each of the model-matched control conditions (mmMusic $p = 0.002$; mmSpeech $p = 0.04$) but the difference between the response to music and scanner noise did not reach significance ($p = 0.1$; Figure 5b, left; effect sizes and confidence intervals reported in Table 2). There were no significant differences in motion between conditions (music vs. speech $t(18) = 1.04$, $p = 0.31$; music vs. model-matched music $t(18) = 0.33$, $p = 0.74$; speech vs. model-matched speech $t(18) = 0.23$, $p = 0.82$; model-matched speech vs. model-matched music $t(18) = 0.79$, $p = 0.44$). As the voxels were selected only for a response to music that was greater than the response to speech, the significantly greater response to music than model-matched music provides strong evidence that the music response in young infants cannot be explained by the spectrotemporal modulation properties of the music stimuli. The lack of a significant difference between music and scanner noise is due to variability across participants in the overall magnitude of auditory responses, which is

TABLE 2 Responses to music and speech in infant auditory and language cortices.

Parcel	Contrast	Intercept[a]	Music[a]	mmMusic[a]	Speech[a]	mmSpeech[a]	age	sex	Motion[b]
HG	Music > Speech	0.17	n/a	−0.06	−0.08	**−0.20**	−0.01	−0.03	1.46
		(−0.05; 0.39)		(−0.15; 0.03)	(−0.18; 0.03)	(−0.21; 0.01)	(−0.13–0.10)	(−0.16–0.09)	(−0.24–3.16)
	Speech > Music	0.16	−0.14	−0.02	n/a	0.01	0.01	0.04	0.21
		(−0.19; 0.51)	(−0.34; 0.06)	(−0.16; 0.11)		(−0.12; 0.14)	(−0.18–0.21)	(−0.18–0.26)	(−2.78–3.21)
NPAC	Music > Speech	0.26	n/a	**−0.16**	**−0.16**	**−0.22**	−0.07	0.19	2.98
		(−0.17; 0.70)		(−0.33; 0.02)	(−0.33; 0.01)	(−0.36; −0.07)	(−0.32–0.18)	(−0.09–0.47)	(−0.82–6.77)
	Speech > Music	0.38	**−0.25**	−0.14	n/a	−0.14	−0.19	**0.47**	3.80
		(−0.34; 1.09)	(−0.54; 0.05)	(−0.31–0.03)		(−0.33;–0.06)	(−0.59–0.21)	**(0.02–0.91)**	(−2.23–9.84)
T. Lang.	Speech > Music	0.12	0.03	−0.03	n/a	0.03	−0.15	0.14	−0.92
		(−0.40; 0.63)	(−0.14; 0.20)	(−0.17–0.12)		(−0.07–0.13)	(−0.45–0.16)	(−0.20–0.49)	(−5.62–3.77)
F. Lang.	Speech > Music	−0.04	0.01	−0.02	n/a	0.07	−0.00	−0.00	0.66
		(−0.27; 0.19)	(−0.10; 0.13)	(−0.16–0.12)		(−0.03–0.17)	(−0.14–0.13)	(−0.15–0.15)	(−1.36–2.68)

[a]Parameter estimates from a linear mixed-effects model for the intercept and three control condition regressors. In the music > speech fROI analysis, the intercept indicates the effect size of the music response relative to scanner noise, the mmMusic value indicates the effect size of the model-matched music response relative to music, the speech value indicates the effect size of the speech response relative to music, and the mmSpeech value indicates the effect size of the model-matched speech response relative to music. In the speech > music fROI analysis, the intercept indicates the effect size of the speech response relative to scanner noise, the music value indicates the effect size of the music response relative to speech, the mmMusic value indicates the effect size of the model-matched music response relative to speech, and the mmSpeech value indicates the effect size of the model-matched speech response relative to speech. Confidence intervals are reported in parentheses and effect sizes with $p < 0.05$ are indicated in bold.
[b]Motion is the proportion of scrubbed voxels.

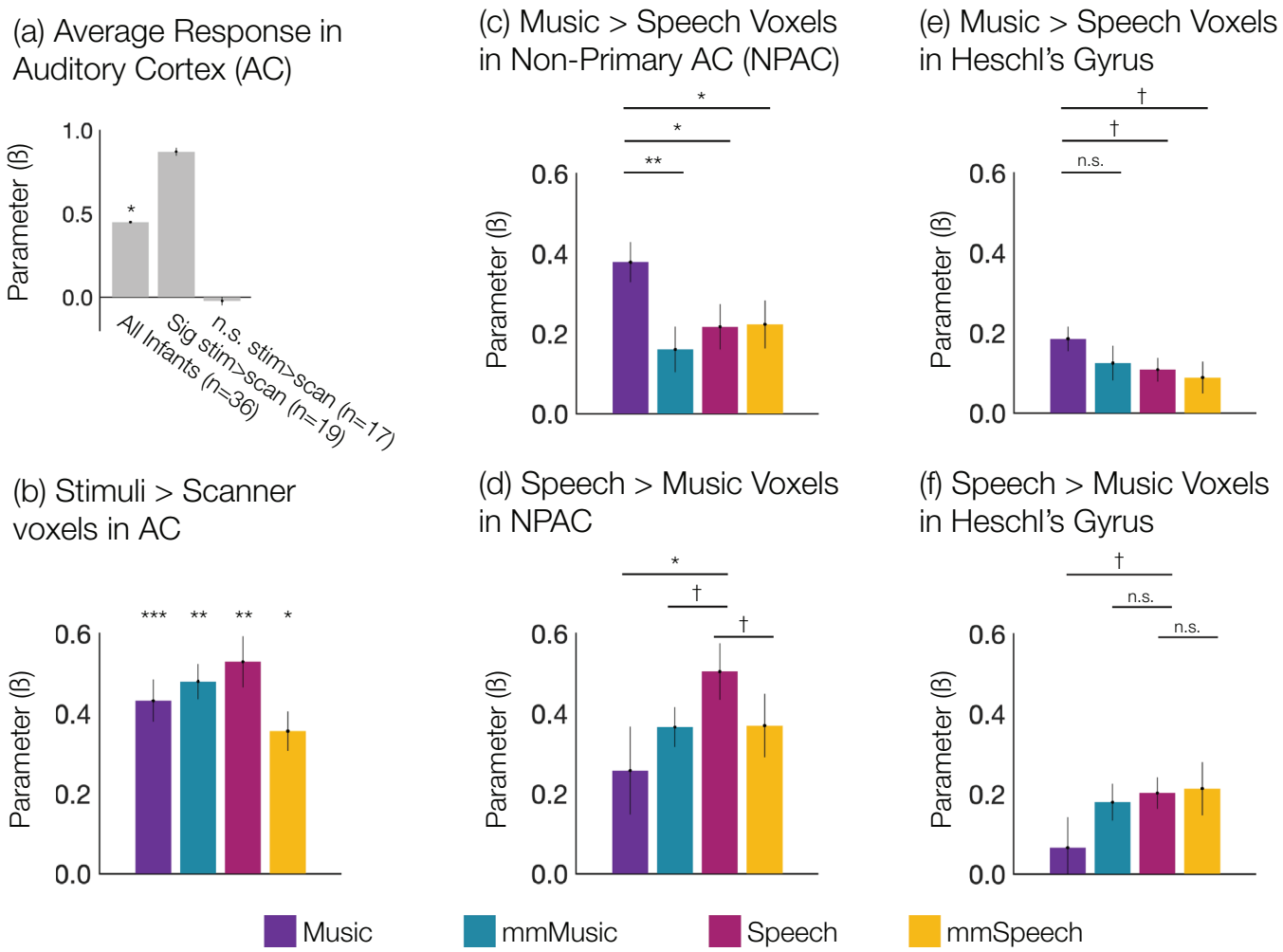

FIGURE 5 Functional organization of infant auditory cortex (AC). Across all infants ($n = 36$) we observed a significant response to all sounds compared to scanner noise in AC (a, includes Heschl's Gyrus (HG) and non-primary auditory cortex (NPAC)) with each condition significantly activating AC (b). The average magnitude response in these voxels was greater for subjects that had significant auditory activations (a, middle bar, $n = 19$) and was at baseline for subjects that did not have significant activations (a, right bar, $n = 17$). (c) In our planned fROI analysis of subjects with significant auditory activation ($n = 19$), voxels that were selected for a stronger response to music compared to speech in NPAC had a music response in left-out data that was significantly greater than the response to all other conditions. (d) In the same subjects, the speech > music fROI analysis of NPAC indicated that the speech response was numerically greater than the response to music and model-matched music but not significantly different than the response to model-matched speech. (e) For the fROI analysis of music > speech in HG, the response to music was greater than model-matched speech but was not significantly different than the response to speech and model-matched music. (f) The speech > music fROI analysis of HG indicated that the speech response was not significantly different than the response to any other condition. Bar charts show the average response in each fROI to music (purple), model-matched music (teal), speech (pink), and model-matched speech (yellow) in data independent of that used to define the fROI. Error bars indicate within-subject SE (Cousineau, 2005). Symbols used to report one-tailed statistics: $^{n.s.}p > 0.1$; $^{\dagger}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$, $^{***}p < 0.001$.

evident in the correlated variability between conditions across subjects (e.g., the Pearson correlation between music and model-matched music was 0.70; $p = 0.0008$), making it easier to detect a significant difference between conditions than between each condition and baseline. This overall variation in response magnitudes is unsurprising and could be due to a variety of neural and non-neural factors (e.g., vascularization).

In the fROI analysis for speech ($n = 19$), we selected the top 5% of voxels in NPAC that responded more to speech than music. In independent, left-out data we found that the speech response was not statistically greater than the response to any of the other stimulus conditions (all $ps \geq 0.05$; Figure 5b, right; Table 2). Although our planned analyses do not provide evidence for speech selectivity, a variety of unplanned analyses do (Figure S4).

To test whether the response of the NPAC voxels selected for their music response differed significantly from the NPAC voxels selected for their speech response, we conducted a repeated-measures analysis of variance (ANOVA) with fROI (voxels preferring music or speech) and condition (music, model-matched music, speech, and model-matched speech) as factors and motion as a covariate. This analysis found a main

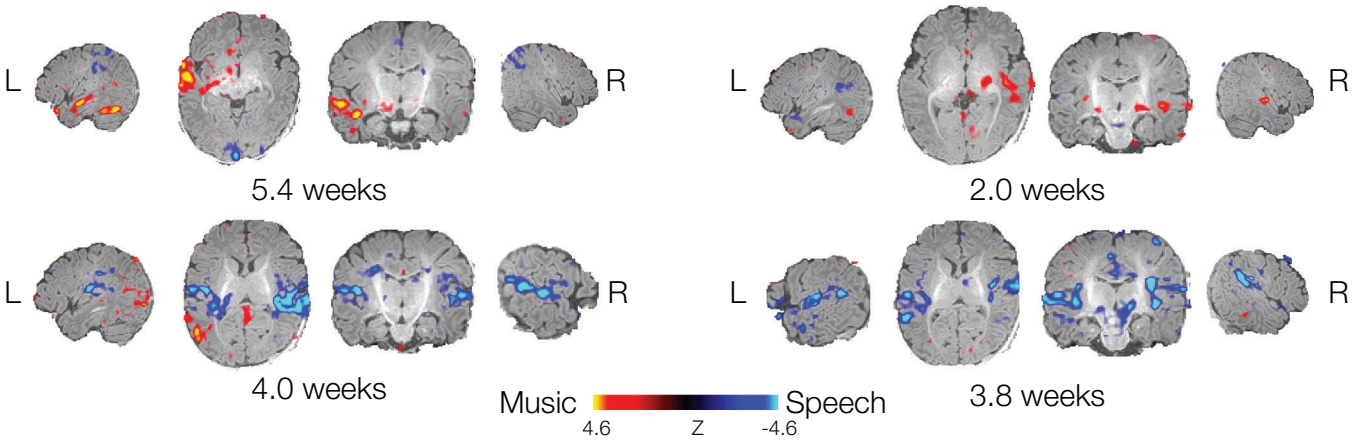

FIGURE 6 Music and speech activations in infant auditory cortex. Whole-brain statistical maps from representative individual infants (threshold $Z < 2.3$) show music (hot) and speech (cool) activations in both hemispheres of auditory cortex. Statistical maps displayed on a non-representative, high-resolution infant anatomical image. Results from group random effects analysis reported in Figure S2.

effect of fROI ($F(1,51) = 4.625$, $p = 0.046$) and a significant fROI by condition interaction ($F(3,51) = 4.344$, $p = 0.008$) supporting the inference that the music fROI and speech fROI differ from each other in their responses across stimulus conditions.

To test the possibility that we biased our results by arbitrarily choosing the top 5% of voxels rather than some other number, we conducted growing window analyses and observed the same pattern of response for both music and speech (see Supplemental Materials).

3.4 | Speech responses in putative language areas

A variety of studies have suggested that infants have speech responses in frontal and lateral language regions prior to language acquisition. Could our inability to find strong evidence for speech selectivity arise because we are looking in auditory regions rather than language regions? To address this possibility, we conducted an fROI analysis of speech responses in temporal and frontal language parcels (Table 2). In independent data the speech responses in the lateral and frontal language parcels were not statistically different than the response to any other condition or scanner noise (all $ps > 0.3$). Thus, we do not find evidence for speech-selective responses in infant brain regions that correspond to language cortex in adults.

3.5 | Functional organization of infant Heschl's Gyrus

All of our pre-planned analyses utilized AC anatomical constraint parcels derived from the organization of adult AC because infant anatomical atlases do not account for variation in functional location. However, it is possible that adult-derived anatomical constraint parcels are also not appropriate for characterizing infant AC function. To determine if the functional organization of infant AC corresponds to that of adults, we conducted an fROI analysis testing whether music- and speech-selectivity is absent in HG (as it is in adults).

Do infants have category-selective responses to music and speech in HG? To answer this question, we conducted an fROI analysis of music responses in the adult derived HG parcel. HG voxels that were selected for their higher response to music than speech showed, in independent, left-out data, a response to music that was numerically but not significantly greater than the response to scanner noise ($p = 0.07$), model-matched music ($p = 0.10$), or speech ($p = 0.07$), but was significantly greater than the response to model-matched speech ($p = 0.047$; Figure 5e; Table 2). However, we did not find a significant interaction between fROI (music in NPAC and music in HG) and condition ($F(3,51) = 0.76$, $p = 0.52$), so we cannot conclude that the music response in NPAC differs from the music response in HG.

For the voxels in HG that were selected for their greater response to speech than music, responses from independent, left-out data indicated that the response to speech was not statistically different than the response to any other condition (all $ps > 0.08$; Figure 5f; Table 2). These results indicate infant HG does not have voxels that are significantly music- or speech-selective, consistent with prior findings for adults.

In sum, we have preliminary evidence for music-selective responses in infant NPAC (not PAC), but we lack clear evidence for speech-selective responses.

4 | DISCUSSION

To answer the fundamental question of whether neural populations selective for speech and music are present in young infants, we designed control stimuli that unconfound responses to higher-level acoustic structure from responses to their lower-level spectrotemporal modulation statistics, providing a more stringent test than has been attempted before. Using these stimuli, we find evidence for music selectivity in NPAC but not PAC, but no strong evidence for speech

selectivity in sleeping infants. While these findings will need to be replicated, and several properties of the response need to be better understood, they provide suggestive evidence that selectivity for music arises very early in human cortical development. Further, these findings provide some of the strongest evidence to date for the presence of any category selective responses in infant cortex, and suggest that specialized cortical responses to music, in particular, develops within about a month of birth.

What can we conclude about auditory responses more generally in sleeping infants? Consistent with previous fMRI research, we observed significant auditory activation in infants' AC (Anderson et al., 2001; Benavides-Varela & Gervain, 2017; Blasi et al., 2011; Bortfeld et al., 2007; Cristia et al., 2014; Dehaene-Lambertz et al., 2002, 2010; Giordano et al., 2021; Homae et al., 2006, 2007, 2012; Kotilahti et al., 2010; Lloyd-Fox et al., 2012; Nishida et al., 2008; Peña et al., 2003; Perani et al., 2010; Sato et al., 2010, 2012; Shultz et al., 2014; Telkemeyer et al., 2011). However, we found that only a subset of infants ($n = 19$) had significant activations in AC while the remaining subjects that contributed data ($n = 17$) did not. The two groups were not different in age, motion, or data quantity. Perhaps the 17 infants without significant auditory activations were unable to hear the stimuli because the speakers that presented sound to infants' ears shifted during scanning. However, given that we observed significant speech activation (relative to scanner noise and model-matched speech) in this group of infants, this possibility is unlikely. Alternatively, it is possible that infants' sleep state influenced whether they had significant auditory activations. As we could not monitor speaker location during scanning and did not monitor sleep state, we cannot distinguish between these two possibilities.

Our finding of music selectivity in infant cortex is subject to several caveats. First, although most aspects of the main analysis presented here were planned in advance, the specific anatomical parcels used were not, giving us experimenter degrees of freedom (see explanation in Table S1). Due to structural and functional neuroanatomical variability across participants, future studies should focus on functionally defining primary and non-primary auditory cortices in infants to increase the rigor of mapping infant auditory cortex. Second, although we found a significantly higher response to music than to the three control stimulus types, the response to music was only numerically and not significantly higher than baseline (scanner noise, $p = 0.1$). One possible explanation for this finding is that the response to scanner noise alone is variable, potentially due to infants' sleep state or physiological differences. Third, although music selectivity reached significance in NPAC and not in HG, the interaction of region by condition did not reach significance. For these reasons, a replication of our finding of music selectivity, and an understanding of the weak statistics concerning the baseline response, will be important. If future methodological advances make possible the collection of a larger amount of data within each participant, such that music-selective activations can be seen in whole-brain contrast maps in most individual participants, that will provide stronger evidence than current methods allow. Further, the simple monophonic instrumental lullabies we used in the present study represent a very small portion of the wide range of music seen within and across cultures. Thus, future research should determine if music responses in infant NPAC is selective to all music genres.

However, the methodological challenges of this work are not small. Notably, the infants in the present study were asleep. In adults, cortical responses to sounds, including speech, are dampened by sleep state (Dang-Vu et al., 2011; Makov et al., 2017; Schabus et al., 2012; Song & Tagliazucchi, 2020) (although see (Davis et al., 2007)). A study of awake infants could address this concern, though likely would cause other problems such as increased motion (Ellis et al., 2020; Ghotra et al., 2021; Kosakowski et al., 2022), and challenges associated with keeping one-month-olds awake. More generally, fMRI with infants is challenging because infants' small HRF (Arichi et al., 2012) and lower tolerance of the scanner environment make it very hard to collect sufficiently high-quality data. New headphones, a new coil, and other innovations only partially address these challenges.

Our finding of early music selectivity may relate to behavioral evidence that infants encode, remember, and have preferences for specific melodies (e.g., a melody sung by a parent rather than a toy (Mehr et al., 2016)), from the first month after birth. Two previous neuroimaging investigations of early music perception failed to find a response to music that was significantly different than the response to speech in neonates (Kotilahti et al., 2010) and three-month-old infants (Dehaene-Lambertz et al., 2010). Why did these studies fail to observe music-selective responses? As we demonstrated, music responses are selective in infants that had significant auditory activations but weaker when collapsed across all infants (Table S2). Thus, perhaps previous studies did not observe music-selective responses because they had a small number of subjects and did not confirm AC activation in each subject. In addition to a small number of subjects, previous studies had a small number of stimuli from each stimulus condition that were played repeatedly. Thus, another possibility is that music responses were weak due to neural habituation (Dehaene-Lambertz et al., 2010). In our study, we had 60 instances of each stimulus category, so no infant ever heard the same stimulus more than once. Additionally, both previous studies had lower spatial resolution than ours—the first used fNIRS, which has a spatial resolution on the order of centimeters—and the second study used fMRI with a voxel volume of ~4 mm^3. In contrast, we used 2 mm^3 voxels. So, another possibility is that previous research blurred together music and speech responses in a single measurement. Taken together, previous studies may have failed to find music responses because AC was not adequately stimulated, because of a lack in power due to small sample sizes, small amounts of data from individual infants and/or neural habituation, or because music and speech responses were blurred together by the imaging modality.

Concerning speech-selective responses, previous studies have a very mixed pattern of results for infant AC and/or language regions (Arimitsu et al., 2011; Benavides-Varela & Gervain, 2017; Blasi et al., 2011; Bortfeld et al., 2007; Cristia et al., 2014; Dehaene-Lambertz et al., 2002, 2010; Fló et al., 2019; Gervain et al., 2008; Kotilahti et al., 2010; Lloyd-Fox et al., 2012; May et al., 2011, 2018; Minagawa-Kawai et al., 2010; Nishida et al., 2008; Peña et al., 2003; Sato et al., 2010, 2012; Shultz et al., 2014). We found evidence for speech-selective responses in NPAC in the unplanned analysis that included

infants without significant AC activation (see Supplemental Materials). However, we failed to observe evidence for speech-selectivity in NPAC for the planned analysis of infants with significant auditory activation or in putative language cortex. One possible explanation for these results is that we did not have enough power to measure robust speech responses in infants. Alternativity, speech selectivity may emerge later in development than music selectivity. Finally, perhaps speech responses are weaker and harder to measure in sleeping infants. Regardless, the inconsistency of speech responses in infants' brains is also found in the previous literature.

Although category selectivity has been observed for some categories that are evolutionarily old and phylogenetically preserved (e.g., faces and bodies (Downing et al., 2001; Kanwisher et al., 1997; Pinsk et al., 2005; Tsao et al., 2003)), it has also been observed for categories that emerged recently in human history (e.g., orthographies (Cohen & Dehaene, 2004)). Thus, the presence (or absence) of a selective response to a perceptual category cannot be used to support, or refute, the phylogenetic origins of the category. Yet, the finding of music selectivity so soon after birth, after only a modest amount of acoustic exposure, and before speech selectivity can be detected, argues against the idea that music selectivity results from the developmental co-option of a neural system that originally arose to support speech perception (James, 1890; Lieberman & Billingsley, 2021; Pinker, 1997). Rather, our results indicate that music- and speech-selectivity emerge independently in infant auditory cortex.

Philosophers and scientists have long debated the origins of musicality. Is some aspect of music innate in humans? Our study strengthens the case that domain-specific mechanisms for music perception arise extremely early in development, apparently independent of, and perhaps even before, domain-specific mechanisms for speech perception.

AUTHOR CONTRIBUTIONS

Heather L. Kosakowski, Samuel Norman-Haignere, and Nancy Kanwisher designed the study. Heather L. Kosakowski and Samuel Norman-Haignere created the stimuli with input and supervision from Nancy Kanwisher. Heather L. Kosakowski and Anna Mynick collected the data with technical support from Atsushi Takahashi. Heather L. Kosakowski analyzed the data with input and supervision by Nancy Kanwisher and Rebecca Saxe. Heather L. Kosakowski prepared figures and drafted the manuscript; Heather L. Kosakowski, Samuel Norman-Haignere, Rebecca Saxe, and Nancy Kanwisher edited and revised the manuscript. All authors provided feedback on the final version.

ACKNOWLEDGMENTS

This research was carried out at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT. The authors thank Ellen Grant and her RAs Catherine Vu, Barbora Zvarova, Clarissa Carruthers, Michaela Sisitsky, Mickayla Royer, Abigail Thomas, and Elizabeth Martin for help with the earliest stages of recruiting and scanning for this project. Lilla Zollei for help with anatomical brain extraction for the template brain. We also thank Joanne Shih and Jonathan Lee for help with melody arrangements; the mothers and musicians that helped with the recordings; Steven Shannon and members of Saxe and Kanwisher Labs for help during recruitment and data collection; Hannah LeBlanc for all the things; and all the infants and their families. We gratefully acknowledge support of this project by the NIH (#1F99NS124175 to HLK; #8K00DA058542 to HLK; #NIDCD-K99-DC018051 to SVN-H; #NIDCD-R00-DC018051 to SVN-H; #R21-HD090346-02 to RS; #DP1HD091947 to NK; shared instrumentation grant S10OD021569 for the MRI scanner), the McGovern Institute for Brain Research at MIT, and the Center for Brains, Minds and Machines (CBMM), funded by an NSF STC award (CCF-1231216) and the Kristin R. Pressman and Jessica J. Pourian '13 Fund at MIT.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data and stimuli that support the findings of this study are openly available in OSF at https://osf.io/8ty34/.

ENDNOTE

[1] Human vocal cords are distinct from other species in that they support various human-specific vocalizations as such singing. The ability to produce vocal song may have emerged ~6 million years ago with the evolutionarily loss of the vocal membranes (Nishimura et al., 2022), or when we evolved better breath control and tongue flexibility, or possibly when the larynx descended, ~50,000 years ago. Conversely, the oldest known instrument, a flute, is ~40,000 years old (Conard et al., 2009). Thus, we make the conjecture that vocal song arose earlier in human evolution than instrumental music.

REFERENCES

Akkermann, M., Can Akkaya, U., Demirel, C., Pflüger, D., & Dresler, M. (2021). Sound sleep: Lullabies as a test case for the neurobiological effects of music. *Behavioral and Brain Sciences*, *44*, e96. https://doi.org/10.1017/S0140525X20001259

Anderson, A. W., Marois, R., Colson, E. R., Peterson, B. S., Duncan, C. C., Ehrenkranz, R. A., Schneider, K. C., Gore, J. C., & Ment, L. R. (2001). Neonatal auditory activation detected by functional magnetic resonance imaging. *Magnetic Resonance Imaging*, *19*, 1–5. https://doi.org/10.1016/S0730-725X(00)00231-9

Angulo-Perkins, A., Aubé, W., Peretz, I., Barrios, F. A., Armony, J. L., & Concha, L. (2014). Music listening engages specific cortical regions within the temporal lobes: Differences between musicians and non-musicians. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior*, *59*, 126–137. https://doi.org/10.1016/j.cortex.2014.07.013

Arichi, T., Fagiolo, G., Varela, M., Melendez-Calderon, A., Allievi, A., Merchant, N., Tusor, N., Counsell, S. J., Burdet, E., Beckmann, C. F., & Edwards, A. D. (2012). Development of BOLD signal hemodynamic responses in the human brain. *Neuroimage*, *63*, 663–673. https://doi.org/10.1016/j.neuroimage.2012.06.054

Arimitsu, T., Uchida-Ota, M., Yagihashi, T., Kojima, S., Watanabe, S., Hokuto, I., Ikeda, K., Takahashi, T., & Minagawa-Kawai, Y. (2011). Functional hemispheric specialization in processing phonemic and prosodic auditory changes in neonates. *Frontiers in Psychology*, *2*, 1–10. https://doi.org/10.3389/fpsyg.2011.00202

Bainbridge, C. M., Bertolo, M., Youngers, J., Atwood, S., Yurdum, L., Simson, J., Lopez, K., Xing, F., Martin, A., & Mehr, S. A. (2021). Infants relax in response to unfamiliar foreign lullabies. *Nature Human Behaviour*, *5*, 256–264. https://doi.org/10.1038/s41562-020-00963-z

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of*

Memory and Language, *68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Baruch, C., & Drake, C. (1997). Tempo discrimination in infants. *Infant Behavior and Development*, *20*, 573–577. https://doi.org/10.1016/S0163-6383(97)90049-7

Benavides-Varela, S., & Gervain, J. (2017). Learning word order at birth: A NIRS study. *Developmental Cognitive Neuroscience*, *25*, 198–208. https://doi.org/10.1016/j.dcn.2017.03.003

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Pnas*, *109*, 3253–3258. https://doi.org/10.1073/pnas.1113380109

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology General*, *117*, 21–33. https://doi.org/10.1037/0096-3445.117.1.21

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. *Current Biology*, *21*, 1220–1224. https://doi.org/10.1016/j.cub.2011.06.009

Boebinger, D., Norman-Haignere, S. V., Mcdermott, J. H., & Kanwisher, N. (2021). Music-selective neural populations arise without musical training. *Journal of Neurophysiology*, *125*, 2237–2263. https://doi.org/10.1152/jn.00588.2020

Bortfeld, H., Wruck, E., & Boas, D. A. (2007). Assessing infants' cortical response to speech using near-infrared spectroscopy. *Neuroimage*, *34*, 407–415. https://doi.org/10.1016/j.neuroimage.2006.08.010

Brown, S., & Jordania, J. (2013). Universals in the world's musics. *Psychology of Music*, *41*, 229–248. https://doi.org/10.1177/0305735611425896

Chang, H.-W., & Trehub, S. E. (1977). Auditory processing of relational information by young infants. *Journal of Experimental Child Psychology*, *24*, 324–331. https://doi.org/10.1016/0022-0965(77)90010-8

Chi, T., Ru, P., & Shamma, S. A. (2005). Multiresolution spectrotemporal analysis of complex sounds. *Journal of the Acoustical Society of America*, *118* 887–906. https://doi.org/10.1121/1.1945807

Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: The case for the visual word form area. *Neuroimage*, *22*, 466–476. https://doi.org/10.1016/j.neuroimage.2003.12.049

Conard, N. J., Malina, M., & Münzel, S. C. (2009). New flutes document the earliest musical tradition in southwestern Germany. *Nature*, *460*, 737–740. https://doi.org/10.1038/nature08169

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology TQMP*, *1*, 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Cristia, A., Minagawa, Y., & Dupoux, E. (2014). Responses to vocalizations and auditory controls in the human newborn brain. *PLoS ONE*, *9*, 1–23. https://doi.org/10.1371/journal.pone.0115162

Da Costa, S., Van Der Zwaag, W., Marques, J. P., Frackowiak, R. S. J., Clarke, S., & Saenz, M. (2011). Human primary auditory cortex follows the shape of Heschl's Gyrus. *Journal of Neuroscience*, *31*, 14067–14075. https://doi.org/10.1523/JNEUROSCI.2000-11.2011

Dang-Vu, T. T., Bonjean, M., Schabus, M., Boly, M., Darsaud, A., Desseilles, M., Degueldre, C., Balteau, E., Phillips, C., Luxen, A., Sejnowski, T. J., & Maquet, P. (2011). Interplay between spontaneous and induced brain activity during human non-rapid eye movement sleep. *PNAS*, *108*, 15438–15443. https://doi.org/10.1073/pnas.1112503108

Davis, M. H., Coleman, M. R., Absalom, A. R., Rodd, J. M., Johnsrude, I. S., Matta, B. F., Owen, A. M., & Menon, D. K. (2007). Dissociating speech perception and comprehension at reduced levels of awareness. *PNAS*, *104*, 16032–16037. https://doi.org/10.1073/pnas.0701309104

Deen, B., Richardson, H., Dilks, D. D., Takahashi, A., Keil, B., Wald, L. L., Kanwisher, N., & Saxe, R. (2017). Organization of high-level visual cortex in human infants. *Nature Communications*, *8*, 1–10. https://doi.org/10.1038/ncomms13995

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science (80-)*, *298*, 2013–2015. https://doi.org/10.1126/science.1077066

Dehaene-Lambertz, G., Montavont, A., Jobert, A., Allirol, L., Dubois, J., Hertz-Pannier, L., & Dehaene, S. (2010). Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain and Language*, *114*, 53–65. https://doi.org/10.1016/j.bandl.2009.09.003

Demany, L., Mckenzie, B., & Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature*, *266*, 718–719. https://doi.org/10.1038/266718a0

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience and Biobehavioral Reviews*, *81*, 181–187. https://doi.org/10.1016/j.neubiorev.2017.02.011

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science (80-)*, *293*, 2470–2473. https://doi.org/10.1126/science.1063414

Ellis, C. T., Skalaban, L. J., Yates, T. S., Bejjanki, V. R., Córdova, N. I., & Turk-Browne, N. B. (2020). Re-imagining fMRI for awake behaving infants. Nature Communications, *11*. https://doi.org/10.1038/s41467-020-18286-y

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *PNAS*, *108*, 16428–16433. https://doi.org/10.1073/pnas.1112937108

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Current Biology*, *22*, 2059–2062. https://doi.org/10.1016/j.cub.2012.09.011

Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. *Journal of Neurophysiology*, *104*, 1177–1194. https://doi.org/10.1152/jn.00032.2010

Fedorenko, E., Mcdermott, J. H., Norman-Haignere, S., & Kanwisher, N. (2012). Sensitivity to musical structure in the human brain. *Journal of Neurophysiology*, *108*, 3289–3300. https://doi.org/10.1152/jn.00209.2012

Feierabend, J. M. (2000). *The Book of Lullabies: Wonderful Songs and Rhymes Passed Down from Generation to Generation*. GIA Publications.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*, 279–293. https://doi.org/10.1016/0163-6383(87)90017-8

Fló, A., Brusini, P., Macagno, F., Nespor, M., Mehler, J., & Ferry, A. L. (2019). Newborns are sensitive to multiple cues for word segmentation in continuous speech. *Developmental Science*, *22*, 1–16. https://doi.org/10.1111/desc.12802

Gasparini, L., Langus, A., Tsuji, S., & Boll-Avetisyan, N. (2021). Quantifying the role of rhythm in infants' language discrimination abilities: A meta-analysis Loretta. *Cognition*, *213*, 104757. https://doi.org/10.1016/j.cognition.2021.104757

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *PNAS*, *105*, 14222–14227. https://doi.org/10.1073/pnas.0806530105

Ghotra, A., Kosakowski, H. L., Takahashi, A., Etzel, R., May, M. W., Scholz, A., Jansen, A., Wald, L. L., Kanwisher, N., Saxe, R., & Keil, B. (2021). A size-adaptive 32-channel array coil for awake infant neuroimaging at 3 Tesla MRI. *Magnetic Resonance in Medicine*, *86*, 1773–1785. https://doi.org/10.1002/mrm.28791

Giordano, V., Alexopoulos, J., Spagna, A., Benavides-Varela, S., Peganc, K., Kothgassner, O. D., Klebermass-Schrehof, K., Olischar, M., Berger, A., &

Bartha-Doering, L. (2021). Accent discrimination abilities during the first days of life: An fNIRS study. *Brain and Language*, *223*, 105039. https://doi.org/10.1016/j.bandl.2021.105039

Goss-Sampson, M. A. (2022). Statistical Analysis in JASP 0.16.1: A Guide for Students. https://jasp-stats.org/wp-content/uploads/2022/04/Statistical-Analysis-in-JASP-A-Students-Guide-v16.pdf

Graven, S. N., & Browne, J. V. (2008). Auditory development in the fetus and infant. *Newborn and Infant Nursing Reviews*, *8*, 187–193. https://doi.org/10.1053/j.nainr.2008.10.010

Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology*, *24*, 14–20.

Hannon, E. E., & Johnson, S. P. (2005). Infants use meter to categorize rhythms and melodies: Implications for musical structure learning. *Journal of Cognitive Psychology*, *50*, 354–377. https://doi.org/10.1016/j.cogpsych.2004.09.003

Hannon, E. E., & Trehub, S. E. (2005a). Tuning in to musical rhythms: Infants learn more readily than adults. *PNAS*, *102*, 12639–12643. https://doi.org/10.1073/pnas.0504254102

Hannon, E. E., & Trehub, S. E. (2005b). Metrical categories in infancy and adulthood. *Psychological Science*, *16*, 48–55. https://doi.org/10.1111/j.0956-7976.2005.00779.x

Homae, F., Watanabe, H., Nakano, T., Asakawa, K., & Taga, G. (2006). The right hemisphere of sleeping infant perceives sentential prosody. *Neuroscience Research*, *54*, 276–280. https://doi.org/10.1016/j.neures.2005.12.006

Homae, F., Watanabe, H., Nakano, T., & Taga, G. (2007). Prosodic processing in the developing brain. *Neuroscience Research*, *59*, 29–39. https://doi.org/10.1016/j.neures.2007.05.005

Homae, F., Watanabe, H., Nakano, T., & Taga, G. (2012). Functional development in the infant brain for auditory pitch processing. *Human Brain Mapping*, *33*, 596–608. https://doi.org/10.1002/hbm.21236

Humphries, C., Liebenthal, E., & Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *Neuroimage*, *50*, 1202–1211. https://doi.org/10.1016/j.neuroimage.2010.01.046

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*, 645–646. https://doi.org/10.1016/S0140-6736(99)02901-3

Jacoby, N., Undurraga, E. A., Mcpherson, M. J., Valdés, J., Ossandón, T., & McDermott, J. H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. *Current Biology*, *29*, 3229–3243.e12. https://doi.org/10.1016/j.cub.2019.08.020

James, W. (1890) *The principles of psychology*. H. Holt & Co.

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. *International Journal of Developmental Neuroscience*, *30*, 159–161. https://doi.org/10.1016/j.ijdevneu.2011.11.002

Kanwisher, N., Mcdermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Kosakowski, H. L., Cohen, M. A., Takahashi, A., Keil, B., Kanwisher, N., & Saxe, R. (2022). Selective responses to faces, scenes, and bodies in the ventral visual pathway of infants. *Current Biology*, *32*, 265–274.e5. https://doi.org/10.1016/j.cub.2021.10.064

Kotilahti, K., Nissilä, I., Näsi, T., Lipiäinen, L., Noponen, T., Meriläinen, P., Huotilainen, M., & Fellman, V. (2010). Hemodynamic responses to speech and music in newborn infants. *Human Brain Mapping*, *31*, 595–603. https://doi.org/10.1002/hbm.20890

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831–843. https://doi.org/10.1038/nrn1533

Landemard, A., Bimbard, C., Demené, C., Shamma, S., Norman-Haignere, S., & Boubenec, Y. (2021). Distinct higher-order representations of natural sounds in human and ferret auditory cortex. *Elife*, *10*, 1–30. https://doi.org/10.7554/eLife.65566

Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *Journal of Neuroscience*, *30*, 7604–7612. https://doi.org/10.1523/JNEUROSCI.0296-10.2010

Lieberman, D., & Billingsley, J. (2021). If it quacks like a duck: The by-product account of music still stands. *Behavioral and Brain Sciences*, *44*, 502–509. https://doi.org/10.1017/S0140525%D7;20000990

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. *Society for Neuroscience*, *7*, 317–330. https://doi.org/10.1080/17470919.2011.614696

Makov, S., Sharon, O., Ding, N., Ben-Shachar, M., Nir, Y., & Zion Golumbic, E. (2017). Sleep disrupts high-level speech parsing despite significant basic auditory processing. *Journal of Neuroscience*, *37*, 7772–7781. https://doi.org/10.1523/JNEUROSCI.0168-17.2017

May, L., Byers-Heinlein, K., Gervain, J., & Werker, J. F. (2011). Language and the newborn brain: Does prenatal language experience shape the neonate neural response to speech? *Frontiers in Psychology*, *2*, 1–9. https://doi.org/10.3389/fpsyg.2011.00222

May, L., Gervain, J., Carreiras, M., & Werker, J. F. (2018). The specificity of the neural response to speech at birth. *Developmental Science*, *21*, e12564. https://doi.org/10.1111/desc.12564

Mcdermott, J. H., & Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis. *Neuron*, *71*, 926–940. https://doi.org/10.1016/j.neuron.2011.06.032

Mcdermott, J. H., & Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis. *Neuron*, *71*, 926–940. https://doi.org/10.1016/j.neuron.2011.06.032

Mcdermott, J. (2008). The evolution of music. *Nature*, *453*, 287–288. https://doi.org/10.1038/453287a

McDermott, J. H. (2014). Audition. In K. N. Ochsner & S. M. Kosslyn (Eds.), *The Oxford handbook of cognitive neuroscience, Vol. 1. Core topics* (pp. 135–170). Oxford University Press.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143–178. https://doi.org/10.1016/0010-0277(88)90035-2

Mehr, S. A., Krasnow, M. M., Bryant, G. A., & Hagen, E. H. (2021). Origins of music in credible signaling. Behavioral and Brain Sciences, *44*, e60. https://doi.org/10.1017/S0140525X20000345

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., Howard, R. M., Hartshorne, J. K., Jennings, M. V., Simson, J., Bainbridge, C. M., Pinker, S., O'donnell, T. J., Krasnow, M. M., & Glowacki, L. (2019). Universality and diversity in human song. *Science (80-)*, *366*, eaax0868. https://doi.org/10.1126/science.aax0868

Mehr, S. A., Singh, M., York, H., Glowacki, L., & Krasnow, M. M. (2018). Form and function in human song article form and function in human song. *Current Biology*, *28*, 356–368.e5. https://doi.org/10.1016/j.cub.2017.12.042

Mehr, S. A., Song, L. A., & Spelke, E. S. (2016). For 5-month-old infants, melodies are social. *Psychological Science*, *27*, 486–501. https://doi.org/10.1177/0956797615626691

Minagawa-Kawai, Y., Cristià, A., & Dupoux, E. (2011). Cerebral lateralization and early speech acquisition: A developmental scenario. *Developmental Cognitive Neuroscience*, *1*, 217–232. https://doi.org/10.1016/j.dcn.2011.03.005

Minagawa-Kawai, Y., Van Der Lely, H., Ramus, F., Sato, Y., Mazuka, R., & Dupoux, E. (2010). Optical brain imaging reveals general auditory and language-specific processing in early infant development. *Cerebral Cortex*, *21*, 254–261. https://doi.org/10.1093/cercor/bhq082

Mineroff, Z., Blank, I. A., Mahowald, K., & Fedorenko, E. (2018). A robust dissociation among the language, multiple demand, and default mode networks: Evidence from inter-region correlations in effect size. *Neuropsychologia*, *119*, 501–511. https://doi.org/10.1016/j.neuropsychologia.2018.09.011

Montagu, J. (2017). How music and instruments began: A brief overview of the origin and entire development of music, from its earliest stages. *Frontiers in Sociology*, *2*, 1–12. https://doi.org/10.3389/fsoc.2017.00008

Moore, R. J., Vadeyar, S., Fulford, J., Tyler, D. J., Gribben, C., Baker, P. N., James, D., & Gowland, P. A. (2001). Antenatal determination of fetal brain activity in response to an acoustic stimulus using functional magnetic resonance imaging. *Human Brain Mapping*, *12*, 94–99. https://doi.org/10.1002/1097-0193(200102)12:2<94::AID-HBM1006>3.0.CO2-E

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J.-B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage*, *25*, 653–660. https://doi.org/10.1016/j.neuroimage.2004.12.005

Nishida, T., Kusaka, T., Isobe, K., Ijichi, S., Okubo, K., Iwase, T., Kawada, K., Namba, M., Imai, T., & Itoh, S. (2008). Extrauterine environment affects the cortical responses to verbal stimulation in preterm infants. *Neuroscience Letters*, *443*, 23–26. https://doi.org/10.1016/j.neulet.2008.07.035

Nishimura, T., Tokuda, I. T., Miyachi, S., Dunn, J. C., Herbst, C. T., Ishimura, K., Kaneko, A., Kinoshita, Y., Koda, H., Saers, J. P. P., Imai, H., Matsuda, T., Larsen, O. N., Jürgens, U., Hirabayashi, H., Kojima, S., & Fitch, W. T. (2022). Evolutionary loss of complexity in human vocal anatomy as an adaptation for speech. *Science (80-)*, *377*, 760–763. https://doi.org/10.1126/science.abm1574

Norman-Haignere, S., Kanwisher, N. G., & Mcdermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition article distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron*, *88*, 1281–1296. https://doi.org/10.1016/j.neuron.2015.11.035

Norman-Haignere, S. V., Feather, J., Boebinger, D., Brunner, P., Ritaccio, A., Mcdermott, J. H., Schalk, G., & Kanwisher, N. (2022). A neural population selective for song in human auditory cortex. *Current Biology*, *32*, 1470–1484.e12. https://doi.org/10.1016/j.cub.2022.01.069

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 Biology*, *16*, e2005127. https://doi.org/10.1371/journal.pbio.2005127

Overath, T., Mcdermott, J. H., Zarate, J. M., & Poeppel, D. (2015). The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. *Nature Neuroscience*, *18*, 903–911. https://doi.org/10.1038/nn.4021

Overath, T., & Paik, J. H. (2021). From acoustic to linguistic analysis of temporal speech structure: Acousto-linguistic transformation during speech perception using speech quilts. *Neuroimage*, *235*, 117887. https://doi.org/10.1016/j.neuroimage.2021.117887

Peña, M., Maki, A., Kovacĭć, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: An optical topography study of language recognition at birth. *PNAS*, *100*, 11702–11705. https://doi.org/10.1073/pnas.1934290100

Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S. F., Dupoux, E., Fazio, F., & Mehler, J. (1996). Brain processing of native and foreign languages. *Neuroreport*, *7*, 2439–2444. https://doi.org/10.1097/00001756-199611040-00007

Perani, D., Saccuman, M. C., Scifo, P., Spada, D., Andreolli, G., Rovelli, R., Baldoli, C., & Koelsch, S. (2010). Functional specializations for music processing in the human newborn brain. *PNAS*, *107*, 4758–4763. https://doi.org/10.1073/pnas.0909074107

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*, 688–691. https://doi.org/10.1038/nn1083

Pinker, S. (1997). *How the mind works*. Norton.

Pinsk, M. A., Desimone, K., Moore, T., Gross, C. G., & Kastner, S. (2005). Representations of faces and body parts in macaque temporal cortex: A functional MRI study. *PNAS*, *102*, 6996–7001. https://doi.org/10.1073/pnas.0502605102

Plantinga, J., & Trainor, L. J. (2009). Melody recognition by two-month-old infants. *Journal of the Acoustical Society of America*, *125*, EL58–EL62. https://doi.org/10.1121/1.3049583

Reybrouck, M., & Podlipniak, P. (2019). Preconceptual spectral and temporal cues as a source of meaning in speech and music. *Brain Sciences*, *9*, 53 https://doi.org/10.3390/brainsci9030053

Sato, H., Hirabayashi, Y., Tsubokura, H., Kanai, M., Ashida, T., Konishi, I., Uchida-Ota, M., Konishi, Y., & Maki, A. (2012). Cerebral hemodynamics in newborn infants exposed to speech sounds: A whole-head optical topography study. *Human Brain Mapping*, *33*, 2092–2103. https://doi.org/10.1002/hbm.21350

Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Development of hemispheric specialization for lexical pitch-accent in Japanese infants. *Journal of Cognitive Neuroscience*, *22*, 2503–2513. https://doi.org/10.1162/jocn.2009.21377

Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2021). Music as a coevolved system for social bonding. Behavioral and Brain Sciences, *44*, e59. https://doi.org/10.1017/S0140525X20000333

Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and conquer: A defense of functional localizers. *Neuroimage*, *30*, 1088–1096. https://doi.org/10.1016/j.neuroimage.2005.12.062

Schabus, M., Dang-Vu, T. T., Heib, D. P. J., Boly, M., Desseilles, M., Vandewalle, G., Schmidt, C., Albouy, G., Darsaud, A., Gais, S., Degueldre, C., Balteau, E., Phillips, C., Luxen, A., & Maquet, P. (2012). The fate of incoming stimuli during NREM sleep is determined by spindles and the phase of the slow oscillation. *Frontiers in Neurology*, *APR*, 1–11. https://doi.org/10.3389/fneur.2012.00040

Shultz, S., Vouloumanos, A., Bennett, R. H., & Pelphrey, K. (2014). Neural specialization for speech in the first months of life. *Developmental Science*, *17*, 766–774. https://doi.org/10.1111/desc.12151

Song, C., & Tagliazucchi, E. (2020). Linking the nature and functions of sleep: Insights from multimodal imaging of the sleeping brain. *Current Opinion in Physiology*, *15*, 29–36. https://doi.org/10.1016/j.cophys.2019.11.012

Sulpizio, S., Doi, H., Bornstein, M. H., Cui, J., Esposito, G., & Shinohara, K. (2018). fNIRS reveals enhanced brain activation to female (versus male) infant directed speech (relative to adult directed speech) in young human infants. *Infant Behavior and Development*, *52*, 89–96. https://doi.org/10.1016/j.infbeh.2018.05.009

Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., Obrig, H., & Wartenburger, I. (2009). Sensitivity of newborn auditory cortex to the temporal structure of sounds. *Journal of Neuroscience*, *29*, 14726–14733. https://doi.org/10.1523/JNEUROSCI.1246-09.2009

Telkemeyer, S., Rossi, S., Nierhaus, T., Steinbrink, J., Obrig, H., & Wartenburger, I. (2011). Acoustic processing of temporally modulated sounds in infants: Evidence from a combined near-infrared spectroscopy and EEG study. *Frontiers in Psychology*, *2*, 1–14. https://doi.org/10.3389/fpsyg.2011.00062

Tierney, A., Dick, F., Deutsch, D., & Sereno, M. (2013). Speech versus song: Multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cerebral Cortex*, *23*, 249–254. https://doi.org/10.1093/cercor/bhs003

Trevor, C., & Frühholz, S. (2021). The evolutionary benefit of less-credible affective musical signals for emotion induction during storytelling. *Behavioral and Brain Sciences*, *44*, 127–129. https://doi.org/10.1017/S0140525X20001004

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. H. (2003). Faces and objects in macaque cerebral cortex. *Nature Neuroscience*, *6*, 989–995. https://doi.org/10.1038/nn1111

Werker, J. F., & Gervain, J. (2013). Speech perception in infancy: A foundation for language acquisition. In P. D. Zelazo (Ed.), *The Oxford Handbook of Developmental Psychology, Vol. 1: Body and Mind* (2013; online edn, Oxford Academic, 16 Dec. 2013). Oxford Library of Psychology. https://doi.org/10.1093/oxfordhb/9780199958450.013.0031

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology*, *66*, 173–196. https://doi.org/10.1146/annurev-psych-010814-015104

Wild, C. J., Linke, A. C., Zubiaurre-Elorza, L., Herzmann, C., Duffy, H., Han, V. K., Lee, D. S. C., & Cusack, R. (2017). Adult-like processing of naturalistic sounds in auditory cortex by 3- and 9-month old infants. *Neuroimage*, *157*, 623–634. https://doi.org/10.1016/j.neuroimage.2017.06.038

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kosakowski, H. L., Norman-Haignere, S., Mynick, A., Takahashi, A., Saxe, R., & Kanwisher, N. (2023). Preliminary evidence for selective cortical responses to music in one-month-old infants. *Developmental Science*, *26*, e13387. https://doi.org/10.1111/desc.13387