Adult-like processing of naturalistic sounds in auditory cortex by 3- and 9-month old infants

Conor J. Wild[a,*], Annika C. Linke[a], Leire Zubiaurre-Elorza[a], Charlotte Herzmann[a], Hester Duffy[a], Victor K. Han[b], David S.C. Lee[b], Rhodri Cusack[a,b,c]

[a] Brain and Mind Institute, Western University, London, Canada
[b] Children's Health Research Institute, London, Canada
[c] School of Psychology, Trinity College Dublin, Dublin, Ireland

ABSTRACT

Functional neuroimaging has been used to show that the developing auditory cortex of very young human infants responds, in some way, to sound. However, impoverished stimuli and uncontrolled designs have made it difficult to attribute brain responses to specific auditory features, and thus made it difficult to assess the maturity of feature tuning in auditory cortex. To address this, we used functional magnetic resonance imaging (fMRI) to measure the brain activity evoked by naturalistic sounds (a series of sung lullabies) in two groups of infants (3 and 9 months) and adults. We developed a novel analysis method – inter-subject regression (ISR) – to quantify the similarity of cortical responses between infants and adults, and to decompose components of the response due to different auditory features. We found that the temporal pattern of activity in infant auditory cortex shared similarity with adults. Some of this shared response could be attributed to simple acoustic features, such as frequency, pitch, envelope, but other parts were not, suggesting that even more complex adult-like features are represented in auditory cortex in early infancy.

Introduction

In their first months, human infants show surprisingly sophisticated auditory perception, with a predisposition to listen to speech (Vouloumanos and Werker, 2004, 2007) and the ability to discriminate subtle phonetic contrasts in native (Trehub and Rabinovitch, 1972; Trehub, 1973; Eimas, 1975) and in non-native languages (Werker and Tees, 1984). They can discriminate between the voices of different speakers, such as their mother's and a strangers' voice (DeCasper and Fifer, 1980), and even generalize speech sounds across talkers (Kuhl, 1979; Jusczyk et al., 1992). It is surprising that young infants can perform these complex feats, considering that the cortical auditory system – the network of brain regions that supports complex auditory processing in adults, and its afferent connections – is immature at the time of birth, and it is not until 4.5–6 postnatal months that the differentiation of cortical layers and the myelination of thalamocortical projections are visible in auditory cortex (Moore and Guan, 2001; Moore and Linthicum, 2007). From these observations, some researchers have suggested that early auditory abilities are facilitated mostly by subcortical auditory processing (Moore, 2002; Eggermont and Moore, 2012). This seems unlikely given that functional neuroimaging has

shown cortical responses evoked by sound in fetuses (Moore et al., 2001; Draganova et al., 2005; Holst et al., 2005; Eswaran et al., 2007; Jardri et al., 2008), infants born very prematurely (Mahmoudzadeh et al., 2013, 2016), newborns (Cheour-Luhtanen et al., 1995a; Cheour et al., 1998a; Anderson et al., 2001; Peña et al., 2003; Kotilahti et al., 2010; Perani et al., 2010, 2011; Baldoli et al., 2014) and 3-month olds (Dehaene-Lambertz et al., 2002, 2006, 2010; Blasi et al., 2011a). What has not been established, however, is the functional role of developing auditory cortex – the auditory features it represents, and the maturity of these responses, which remain unclear.

Some studies have asked whether the immature auditory cortex of very young infants is capable of speech-specific processing. They have found that responses differ in magnitude between speech and non-speech sounds (Dehaene-Lambertz et al., 2002, 2010; Peña et al., 2003; Homae et al., 2011; Minagawa-Kawai et al., 2011; Perani et al., 2011; Sato et al., 2011; Shultz et al., 2014), or between different kinds of speech sounds, such as different phonemes (Cheour-Luhtanen et al., 1995b; Cheour et al., 1998b; Mahmoudzadeh et al., 2013, 2016; Kuhl et al., 2014), or languages (Minagawa-Kawai et al., 2011; Sato et al., 2011; Vannasing et al., 2016). However, the conclusion that these differences in responses reflect speech specificity is confounded by the

fact that different sounds differ in their basic acoustic features such as acoustic envelope, pitch, and frequency. Adult auditory cortex is known to be exquisitely sensitive to these simple features (Giraud et al., 2000; Hall et al., 2002a; Patterson et al., 2002; Nourski et al., 2009; Linke et al., 2011a; Kubanek et al., 2013), and so neuroimaging researchers of adult speech perception often go to great lengths to create well controlled acoustic stimuli (Scott et al., 2000; Remez et al., 2001; Sohoglu et al., 2012; Wild et al., 2012a). Because of poor auditory control in infant studies (e.g., forwards vs. backwards speech, or even speech vs. silence), or presentation of a very few sounds in a highly stereotyped sequence (e.g., a train of '/ga' with the occasional '/ba'), we still cannot determine which acoustic features drive cortical responses in very young brains. Do these responses reflect differences in simple acoustic features, more complex auditory representations (e.g., pho-nemes), or tuning to a variety of auditory features?

Furthermore, does the feature tuning of infant auditory cortex appear adult-like? If presented with naturalistic sounds, is the infant cortex driven by the same kinds of auditory features, and in the same way, as the mature auditory cortex? One tantalizing observation from infant functional magnetic resonance imaging (fMRI) is that the spatial distribution of auditory activity resembles that seen in adults (Dehaene-Lambertz et al., 2002, 2006; Perani et al., 2011; Shultz et al., 2014). While one can conclude that similar cortical regions process sound in infants and adults, we cannot conclude that they function similarly. For example, It is well known that the functional mismatch response (MMR) – the most common tool for studying infant perception with electroencephalography (EEG) – undergoes significant change in the first six months after birth to become more adult-like (Trainor et al., 2003; He et al., 2007a), which suggests that cortical auditory responses in early infancy are immature, and perhaps quite distinct in function. Yet, alternatively the differences in the morphology of auditory evoked responses might reflect the influence of immature physiology of the coupling between neural activity and the measured signal (Trainor et al., 2003; Eggermont and Moore, 2012), rather than underlying differences in function.

We conducted an experiment to address these unresolved questions about auditory cortex function in early infancy. What acoustic features does auditory cortex process? Is there evidence for tuning to complex in addition to simple acoustic features? Do auditory responses in infant auditory cortex resemble adult cortical responses? We used fMRI to isolate and characterize, in infants (at 3- and 9-months of age) and adults, the responses in auditory cortex evoked by rich and naturalistic sounds that are engaging to infants – sung lullabies. These infant age groups were selected because around six months after birth is considered to be a turning point in terms of auditory perception (e.g., the beginning of a shift from universal to language-specific perception, Werker and Tees, 1984) and anatomical development (e.g., mylenation of thalmocortical projections, Moore and Guan, 2001). We therefore expected to see developmental differences in how auditory cortex responses to rich sounds; for example, perhaps these perceptual and anatomical changes are accompanied by more mature-like processing in auditory cortex. We developed a novel analysis technique, inter-subject regression (ISR), which specifically allowed us to answer these questions. ISR combines the hypothesis-driven general linear model (GLM) with the model-free approach of inter-subject correlation (ISC) (Hasson et al., 2004) in order decom-pose brain activity into components that reflect different aspects of naturalistic auditory processing. Brain responses driven by simple acoustic features, such as amplitude envelope, pitch, and frequency, were directly modelled from the stimulus (as in a conventional GLM), whereas more complex brain responses to abstract features were identified as the component of evoked response shared by all listeners that could not be attributed to the simple features. Importantly, ISR allowed us to directly measure the similarity of the timecourse of brain

activity between infants and adults, while accounting for age-related differences in the hemodynamic response function (HRF) (Arichi et al., 2012) that would make their evoked brain responses appear to be dissimilar. We hypothesized that adult auditory cortex would be sensitive to low-level acoustic features and show coding of more complex features of the rich acoustic stimuli. We also hypothesized that, despite immature cortical anatomy, we would observe reliable tuning to these features at 3 months of age, and that these responses would be somewhat similar to adults. Furthermore, we expected to observe maturational changes in the first year, so that by 9-months, auditory responses would be even more adult-like.

Materials and methods

Subjects

Two groups of infants at different ages (3- and 9-months old) were recruited to undergo MRI scanning. These ages were chosen because they fall before and after 6-months – a time that has been proposed to be associated with increasing cortical connectivity (Moore and Guan, 2001; Moore and Linthicum, 2007) and the beginning of significant changes in auditory perception (Kuhl et al., 2008; Werker and Hensch, 2014).

Twenty-four 3-month old infants were recruited, but only six useable fMRI data sets were obtained from these volunteers (mean corrected age at scan 3.5 months, standard deviation 0.5 months; 4 females, 2 males). Similarly, sixteen 9-month old infants were admitted to the study, but only seven yielded useable fMRI data (mean corrected age at scan 9.3 months, standard deviation 1.0 months; 1 female, 6 males). All infant participants were scanned unsedated during natural sleep to minimize subject motion. The seemingly low success rate (6/24 3-months, and 7/16 9-months) is a result of infants either: not falling asleep at all (and not entering the scanner); or waking before we had completed 2 fMRI sessions and at least one structural image. Of the 3-month old infants (N=6), two families were recruited from Western University's Developmental Psychology Pool, and four families were recruited from the Neonatal Intensive Care Unit (NICU) at Children's Hospital in London, Ontario. Families in the latter group were approached to participate in a larger study investigating the effects of suspected brain injury on neurodevelopmental outcomes. Similarly, four of the 9-month olds (N=7) were recruited from the Developmental Psychology Pool at Western University, and three of the families were approached in the NICU at Children's Hospital. Despite the inclusion of subjects recruited from the NICU, none of the final cohort had brain injury apparent on the structural MRI. The details of the infant participants are shown in Table 1. Parents or legal guardians gave informed consent for the infants to participate in this study, and the study protocol was examined and cleared by the Western University Health Sciences Research Ethics Board.

To characterize the response of the mature auditory system, a group of 16 adults (11 females, mean age 22 years 7 months) were recruited from the students and staff at Western University, and were paid for their participation in this study. All participants spoke fluent English, were right-handed, with no reported hearing deficits, and no known language or neurological impairments. Participants were screened for compliance with magnetic resonance imaging safety standards: they reported no prior surgeries involving metallic implants, devices, or objects. They gave informed consent according to ethical guidelines laid down by Western University.

Data acquisition

All scanning was performed on a Siemens (Erlangen, Germany) MAGNETOM Prisma 3T MRI system located at the Robart's Centre for

Table 1
Details of infant participants included in fMRI analyses. Asterisks indicate estimated GA at birth because medical details were not available for these infants, as they were not recruited through the NICU. N/A – data available.

ID	Sex	GA at Birth (weeks)	Weight at Birth (grams)	(Corrected) Age at Scan (days)	Observations
3_1	M	29	800	92	Preterm (also subject 9_6)
3_2	F	28	1160	92	Preterm
3_3	F	41*	N/A	89	N/A
3_4	M	41	3275	91	Hypoxia, stroke, seizures
3_5	F	41	4070	92	Stroke, white matter hemorrhages
3_6	F	40*	N/A	92	N/A
9_1	M	40*	N/A	273	N/A
9_2	M	40*	N/A	274	N/A
9_3	M	40*	N/A	285	N/A
9_4	M	27	1070	273	Preterm
9_5	M	27	N/A	273	IVH grade II
9_6	M	29	800	273	Preterm (also subject 3_1)
9_7	M	25	N/A	334	IVH grade II

Functional and Metabolic Mapping at Western University in London, Ontario, Canada. Adult subjects were scanned while awake with a Siemens 32-channel head coil. We acquired in every adult subject a T1-weighted structural image and four runs of functional MRI data (with auditory stimulation) that were approximately 7 min and 30 s in length. All fMRI data were acquired using the Centre for Magnetic Resonance Research multiband EPI sequence (Release 10b, VD13D, http://www.cmrr.umn.edu/multiband/). Acquisition was transverse oblique, angled away from the eyes, and in most cases covered the whole brain; in a very few subjects, slice positioning excluded the top of the superior parietal lobule.

Following discussion with parents on what would best help their infant sleep, some 3-month-old subjects were swaddled and comforted in a vacuum cushion (MedVac Infant Vacuum Immobilization Bag, Contour Fabricators Inc., Fenton). A Siemens 20-channel head coil was required to accommodate the added cushioning and large over-the-ear headphones used for infants. An MR-compatible infra-red camera (MRC Systems M12 camera) and noise cancelling microphone (Optoacoustics FORMI-III noise-cancelling microphone system) placed on the coil allowed experimenters and parents to see and hear the infants in the scanner; if an infant awoke and was distressed, scanning was immediately stopped. A NICU nurse that had attended a pediatric advanced life support (PALS) course was present at all infant scanning sessions, in case of an unexpected medical occurrence. No such emergencies occurred.

We attempted to obtain from every infant T1- and T2-weighted structural images and two sessions of fMRI data (7 min and 30 s each, with auditory stimulation), along with other structural MRI sequences not intended for this study. Total scanning time for infants was less than one hour. Multiband acceleration (Feinberg et al., 2010; Xu et al., 2013) was used to increase sampling rate and reduce sensitivity to movement (36 slices of 64×64 matrix size with 3×3 mm in-plane resolution, and slice thickness 3 mm, multiband factor 4). The echo-time was adjusted for the 3-month group (TE=40 ms) to reflect the longer T2* relaxation due to increased water content in brain tissue (Rivkin et al., 2004), but used a typical value in the 9-month and adult group (TE=30 ms). There was slight variation in protocol between subjects, reflecting ongoing optimization (3 month: slice gap=0–0.3 mm, TR=776–861 ms. Adults and 9 month: slice gap=0–0.3 mm; TR=686–861 ms). Both T1 and T2* weighted structural images were acquired (36 oblique slices of 3 mm thickness, 64 × 64 matrix, voxel size 3 × 3 x 3 mm3, TR = 4000 ms, TE = 120 ms) at the beginning of the MRI testing.

Auditory stimulation

To evoke brain responses to naturalistic sounds, during fMRI sessions all participants heard auditory stimulation in the form of lullabies sung by a female vocalist (a native speaker of North American English). Five different songs were recorded: *You are my sunshine*, *Rock-a-bye Baby*, *Brahm's Lullaby*, and *Hush, Little Baby*. Each lullaby was processed to manipulate different acoustic features, yielding four acoustic conditions: 1) natural singing with speech and melody (i.e., the original recording); 2) a version without words, only "la la la"; 3) a re-synthesized version in which individual phonemes were reversed; and 4) a re-synthesized version in which the pitch trajectory of the original vocal was reconstructed using a single sine-wave, embedded in a background of speech-shaped noise at a signal-to-noise ratio of 12 dB. Synthesized stimuli were created using TANDEM-STRAIGHT, a tool for manipulating and creating naturalistic speech-like stimuli (Kawahara et al., 2008). However, these manipulations were implemented for another study, and will not be discussed further in this manuscript.

The recordings were edited to yield 15-s long clips of singing, which were normalized to have the same root-mean-square amplitude. The sound clips were presented in a pseudo-random order in a block design with 15 s of singing followed by 11 s of silence. Critically for our model free analysis method (described later), the song presentation order was *the same* across all participants. Only sixteen clips (out of all 20 possibilities) were presented during each fMRI session to reduce the length of the scanning session, and each session contained a slightly different selection of songs. The song sequence was triggered to begin after 10 "dummy" scans at the start of each fMRI session. Auditory presentation was controlled from a Windows XP laptop running Matlab 2013b (www.mathworks.com), using Psychtoolbox Version 3 (psychtoolbox.org). Sounds were played through a MOTU Microbook II USB soundcard (http://www.motu.com/) into a Pyle 30-Watt amplifier that drove Sensimetrics S14 insert earphones (http://www.sens.com/) at a comfortable listening level.

A slightly different auditory presentation setup was used for infants, to ensure safe hearing levels (e.g., to prevent a scenario in which an in-ear headphone falls out) and to better separate auditory stimulation from the background noise of the scanner. A custom built piezo-electric speaker system (driven by the same soundcard and amplifier) was placed at the foot of the MRI bed, and plastic tubing directed the sound into customized ear defenders that were placed around the infants' ears. The infants were also equipped with earplugs and minimuffs (http:scanmedics.com/mini-muffs/) to provide additional ear protection (up to 30 dB and 7 dB, respectively). Again, sound presentation levels were set to levels that were comfortable by the experimenters (using the same equipment – including earplugs and minimuffs) in pilot scanning sessions.

fMRI data preprocessing

Functional MRI data were processed using automaticanalysis (version 4.1; Cusack et al., 2015a): a data processing and analysis pipeline that integrates commonly used routines from Statistical Parametric Mapping (SPM8; Wellcome Centre for Neuroimaging, London, UK) with custom data processing modules written in MATLAB. Ten "dummy" scans were excluded from the beginning of every fMRI data set to allow longitudinal magnetization to reach equilibrium. Initial preprocessing steps included: rigid realignment of each EPI volume to the first image of the series to correct for motion, generation of a mean EPI volume, reslicing of all EPI volumes, and co-registration of the structural image (T1 for adults, T2 and/or T1 for infants) to the mean EPI.

We estimated the non-rigid warp of each subjects' structural image to the appropriate group template using SPM8's unified segment-and-normalize routine (Ashburner and Friston, 2005). Adults were normalized to MNI space using the default SPM8 tissue probability maps, whereas infants (both 3- and 9-month olds) were normalized to the University of North Carolina (UNC) 1-year-old template (Shi et al., 2011). Fig. 1 shows examples of normalized structural images, and the corresponding templates. A transformation from MNI space to UNC 1-year-old space, and its inverse transformation, was estimated using a pair-wise non-rigid registration between skull-stripped template images in each space. These normalization transformations were used to warp the regions of interest (ROIs; see below) to the individual subjects' space.

The fMRI time series was extracted from each region of interest (ROI; see next section) by averaging across all voxels in the region, and each summary time series was resampled to match the adult TR. Temporal resampling was done using spline interpolation of the time varying signal. Signal spikes in the resulting ROI time series were clamped to 2.5 standard deviations away from the mean signal (for each ROI, for each subject and session). Finally, the fMRI time series were high-pass filtered with a cut-off of 120 s by regressing out a discrete cosine basis set.

Regions of interest (ROIs)

We defined auditory cortex and the broader regions of a "speech perception" network using Neurosynth (www.neurosynth.org; Yarkoni et al., 2011) - a large-scale meta-analytic tool with access to 11406 fMRI studies – to identify brain regions that are selectively involved in processing speech in adults. Given the search term "speech", Neurosynth provided a reverse inference brain map of z-scores; higher voxel z-scores indicate that, given activation in that voxel, there is a greater likelihood that it came from a study containing the key term "speech".

We used an adaptive thresholding technique to parcellate this continuous z-score map into discrete brain regions. First, we smoothed the three-dimensional reverse inference map with a Gaussian kernel with a half-width at half-maximum of 14 mm, and identified all the peak voxels in the smoothed image (i.e., voxels where the spatial gradient of z-scores was of opposite sign on either side, in each dimension). This process yielded 15 peak voxels. Next, we determined a unique z-score threshold for each peak that cleanly separated the cluster around that peak from all other clusters. Fig. 2 shows the resulting 15 ROIs overlaid on the adult and 1-year-old template structural images. Given our specific interest in the development of auditory cortex, most of our analyses focus on that particular region, which includes Heschl's Gyrus (the anatomical landmark of primary auditory cortex) and extends along the length of the superior temporal gyrus and sulcus.

Inter-subject regression (ISR)

Our study was designed to address two complementary questions: 1) do rich and naturalistic sounds evoke similar functional responses in infant and adult auditory cortex; and, 2) are these responses driven by simple acoustic features, such as the amplitude envelope, or by higher-order features? The first question could be tested using the method of inter-subject correlation – a model-free analysis method that allows researchers to study how the brain processes rich and naturalistic stimuli (Hasson et al., 2004). The detection of shared brain responses to the same rich stimulus across subjects implies that there is some shared signal or representation in a brain region evoked by the common stimulus. However, it can remain unclear exactly what features of the stimulus are driving that shared signal. For example, in an auditory presentation paradigm, inter-subject synchrony might be driven by a response to the sound envelope or something much more complex and difficult to quantify, such as an emotional response to the prosody. On the other hand, conventional fMRI analysis utilizing the general linear model (GLM) can be used to specifically test whether brain responses are driven by well-characterized stimulus features, such as the amplitude envelope.

For the purposes our study, we developed a novel technique – intersubject regression –that combines the model-free approach of ISC with the GLM, in order to estimate the contribution of well-characterized auditory features to a shared complex brain response. This allows us to assess not only whether brain responses to a given stimulus are similar, but also what aspects of the stimulus drive the brains' responses. Furthermore, by accounting for confounding or simple
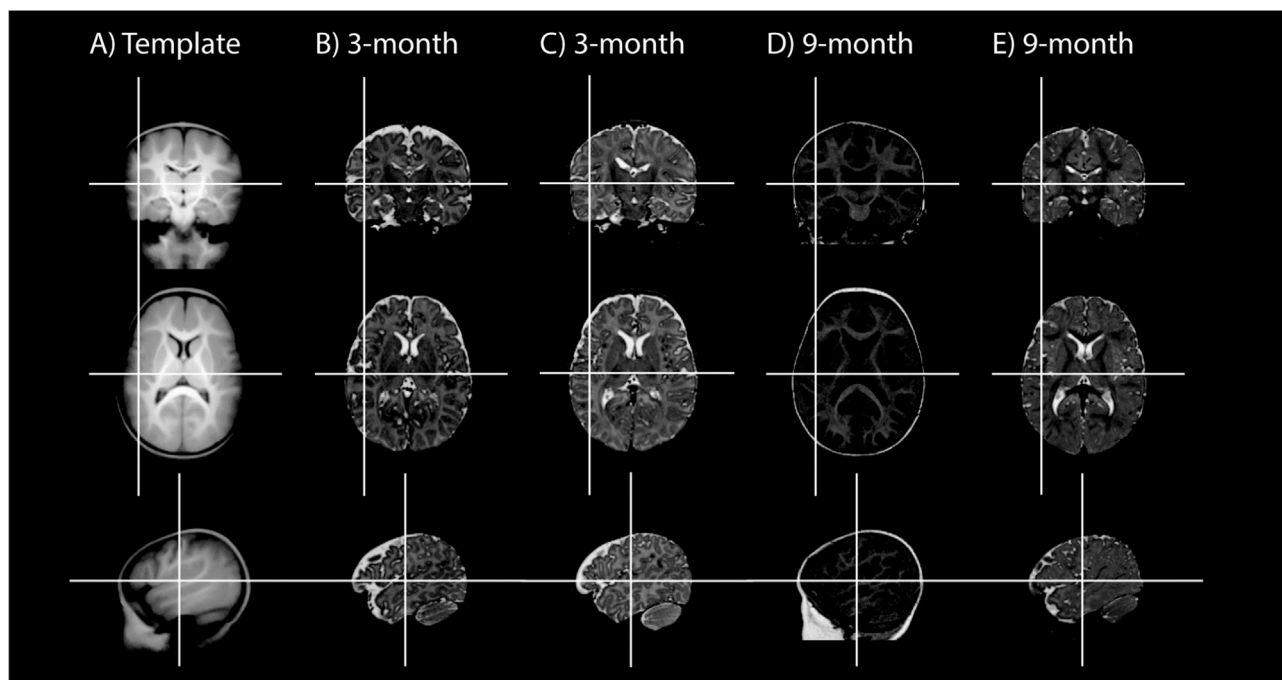

Fig. 1. Examples of normalized structural images for two 3-month-olds (B, C) and two 9-month olds (D, E). The UNC 1-year-old is shown for reference (A).
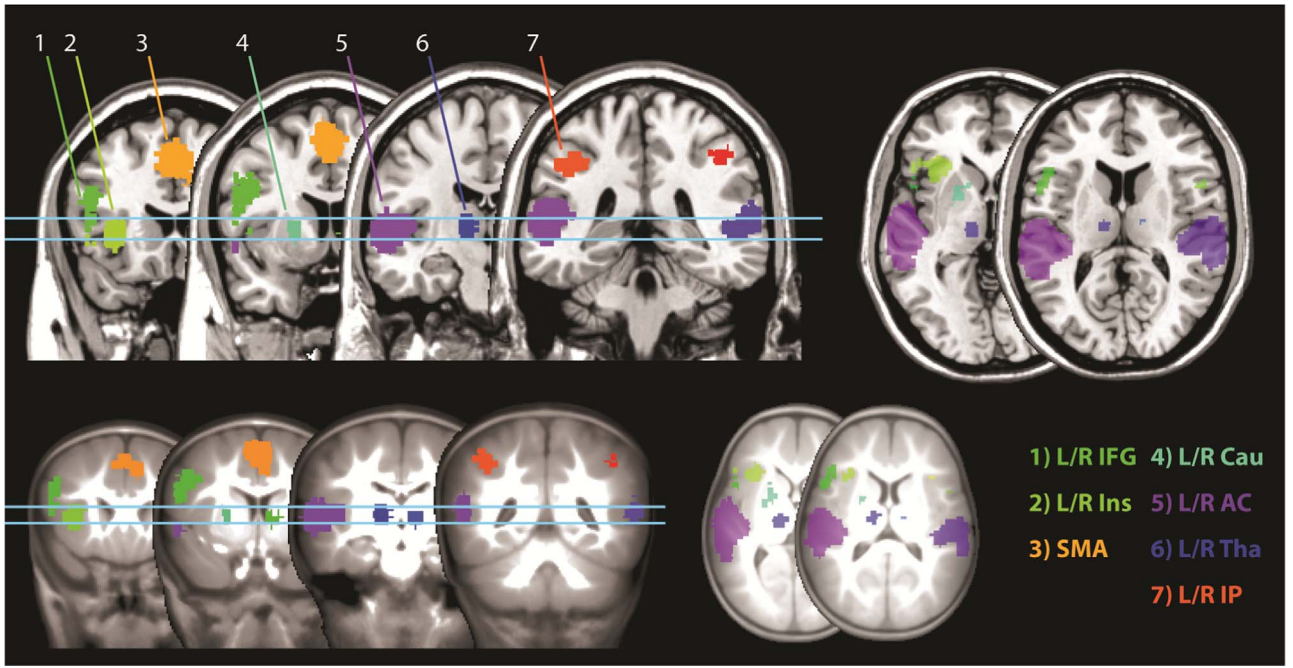
Fig. 2. Regions of Interest (ROIs) derived from a Neurosynth reverse inference map of the term "speech", overlaid on the adult MNI template T1-weighted structural image (top), and the 1-year old UNC infant template (bottom). Horizontal blue lines across coronal slices indicate the position of the axial slices. L/R – Left/Right; IFG – Inferior Frontal Gyrus; Ins – Insula; SMA – Supplementary Motor Area; Cau – Caudate; AC – Auditory Cortex; Tha – Thalamus; IP – Inferior Parietal. Not shown: left and right cerebellum.

acoustic effects, we can be more sure that any shared responses reflect higher-order processing.

ISR is a regression model (Fig. 3), where the goal is to estimate for each subject, l, the contribution, β_i, of each of p regressors of interest, $X_i(t)$, to the fMRI signal in a region, $Y_l(t)$:

$$Y_l(t) = \sum_{i=1}^{p} \beta_i X_i(t) + \varepsilon \qquad (1)$$

For adult participants, one of the regressors of interest in $X(t)$ was the average fMRI time series from all other subjects in that group (i.e., the mean N-1 signal; Eq. (3)), allowing us to estimate the similarity of brain responses between adult subjects. For infants, the model instead included the average adult fMRI time series from the same region, in order to estimate the similarity between infant and adult brain responses. Other regressors were included in $X(t)$ to explain responses that could be directly modelled from the stimulus (e.g., the auditory envelope, see the following section: acoustic predictors). Importantly, the mean N-1 signal (or the adult mean signal) was orthogonalized with respect to all other predictors in $X(t)$, with any shared variance attributed to the other predictors, so that it represented the component of the shared signal that couldn't be accounted for by any of the acoustic predictors.

To account for the delay between stimulation and the evoked BOLD signal, and differences in the shapes and/or latency of the response between infants and adults, $X(t)$ was convolved with a set of regressor- and region-specific hemodynamic response functions (HRFs, see below), $H(t')$:

$$Y_l(t) = \sum_{i=1}^{p} \beta_i H_i(t') \otimes X_i(t) + \varepsilon \qquad (2)$$

Note that in the case of adult subjects, the mean N-1 predictor in $X(t)$ was not convolved with any HRF because we assumed a similar response between all subjects in the group.

It has been shown that the HRF can differ between brain regions (Aguirre et al., 1998), and that its shape changes dramatically during early development (Arichi et al., 2012); critically, using an incorrect HRF model can reduce sensitivity to evoked brain responses (Cusack
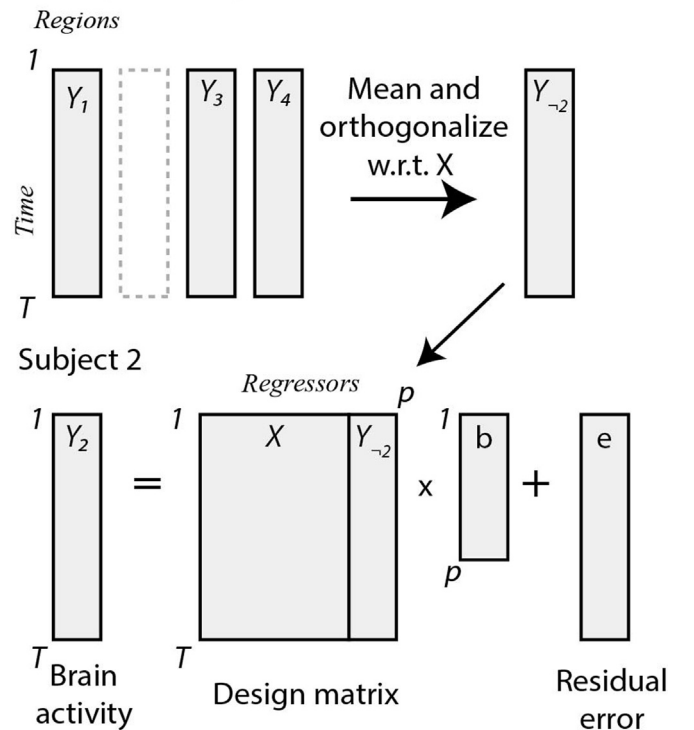

Fig. 3. A schematic diagram of the inter-subject regression (ISR) method. The fMRI time series for a given region for a subject, (e.g., Y_2) is modelled using the general linear model. The design matrix includes regressors derived from the auditory stimulus (X), and the mean fMRI signal from all other subjects (e.g., $Y_{\neg 2}$). Y = fMRI time series; T = number of time points; X = auditory regressors; p = number of regressors in the design matrix; b = parameter estimates (betas); e = residual error.

et al., 2015b). Therefore, estimation of each H_i for each brain region reduced the possibility of an HRF mismatch, which would lead to a poor fit of the acoustic predictors, and hence an overestimation of the

remaining shared brain signal. The set of hemodynamic responses $H(t')$ were considered unknown, and therefore estimated using leave-one-subject-out cross validation.

We first calculated the mean time-course $Y_{\neg l}(t)$ of the other N-1 subjects in the same age group as subject l:

$$Y_{\neg l}(t) = \frac{1}{N-1} \sum_{k=1, k \neq l}^{N} Y_k(t) \tag{3}$$

The hemodynamic responses $H_i(t')$ were represented as the sum of a set of 6 finite impulse response (FIR) basis functions $F_j(t')$ with unknown weights h_{ij}:

$$H_i(t') = \sum_{j=1}^{6} h_{ij} F_j(t') \tag{4}$$

which were determined using an ordinary least squares fit to the mean N-1 signal, $Y_{\neg l}$:

$$Y_{\neg l}(t) = \sum_{i=1}^{p} H_i(t') \otimes X_i(t) + \varepsilon$$
$$= \sum_{i=1}^{p} \sum_{j=1}^{6} h_{ij} F_j(t') \otimes X_i(t) + \varepsilon \tag{5}$$

These values of h_{ij} were used to generate the hemodynamic response using (4) and to solve for β_i in Eq. (2). Supplementary Fig. S1 shows the estimated hemodynamic responses. As in a conventional GLM analysis, the model parameters β_i were estimated separately for, and averaged over, repeated fMRI sessions, and statistically tested in a random effects analysis at the group level.

Acoustic predictors

In this study, we used a series of ISR analyses that show how synchronized brain activity in response to a common auditory stimulus can be attributed to different features of the sound. We began with a simple ISR model that included only the mean N-1 signal (or mean adult signal, for infants) and one other predictor, and then added more regressors to model to determine what aspects of any shared brain responses could be explained by acoustic factors.

First, we included a predictor that modelled the sound of the scanner turning on. Even though all participants wore ear protection, it is probable that they could still hear the onset of the scanner at the start of each session. A build-up of signal in auditory cortex elicited by the sound of the scanner might still be detectable after the removal of dummy scans – and would create artefactual inter-subject synchrony. So, we wanted to conservatively account for any such signal drift at the start of our fMRI time series. The hemodynamic response in auditory cortex to a block of sound peaks around ten seconds after the onset of that sound in adults (Hall et al., 1999), and it is likely to be even longer in infants given that their HRF has a longer time to peak than the adults'. To create a flexible model of this onset effect, we convolved a boxcar (onset at the first dummy volume) with our FIR basis set and then removed the first ten rows of the resulting regressor matrix. By including these regressors in the ISR model, we could capture any shape at the start of each fMRI session that was consistent across subjects within a group.

Second, we expected that the auditory envelope – which includes both the overall block effect of sound turning on and off, in addition to intensity variations within the 15 s song clips – would drive a significant portion of auditory cortex activity (Giraud et al., 2000; Millman et al., 2013). To model these amplitude fluctuations, the envelope of the complete 7-min song sequence was extracted by: half-wave rectifying the audio files; applying the Hilbert transform; and resampling the resulting time series to the sampling rate of the fMRI data.

Third, it seemed possible that the time varying pitch of the songs could drive fMRI activity in auditory cortex, as has been observed by other researchers (e.g., Patterson et al., 2002; Hall and Plack, 2009). To create a representation of the time varying pitch signal, we extracted the fundamental frequency (f0), in Hz, of the sequence of songs using TANDEM-STRAIGHT – a tool for the analysis, warping, and synthesis of speech signals (Kawahara et al., 2008). The resulting f0 signal was resampled to match the fMRI sampling rate, and periods of unvoiced singing and silence were replaced with the mean pitch. The resulting signal was mean centred.

Finally, we tested the sensitivity of auditory cortex to frequency content of sounds (e.g., Talavage et al., 2000; Linke et al., 2011b). We constructed a signal that measured the "brightness" of the auditory stimuli by calculating the spectral centroid of the auditory signal (i.e., the weighted mean of the frequencies in the signal) at each point in time. Periods of silence were replaced with the mean brightness, and the signal was mean centred and resampled to the fMRI sampling rate.

Group level statistics

Statistical testing of the parameter estimates from the single subject ISR models consisted of a random effects analysis using two-tailed one-sample t-tests at the group level. Each ROI was tested individually. Non-parametric statistics were required to overcome the covariance between measures from different subjects introduced by the leave-one subject out nature of ISR. Every subject's data (i.e., the same data) were used to estimate the similarity of each subject to the rest of the group, and regressor-specific HRFs, which makes the standard error a poor estimator of the volatility of the mean across repeated experiments (see Supplementary Fig. S2)

We created synthetic fMRI time series for each group in order to create null distributions of group-level contrast values (a separate null distribution for each contrast of interest, for each age group). The real group-level contrast values were then compared to null distributions, and if they lay outside 95% of the null were considered statistically significant. Simulated fMRI data were generated using parametric bootstrap resampling in the wavelet domain to ensure that the fMRI noise was well characterized for each group of subjects (Bullmore et al., 2001; Cusack et al., 2015b). First, we performed a discrete wavelet transform of each subjects' fMRI time course, using Debauchies wavelets of order 1 with 7 levels of detail. Wavelet coefficients for each level were pooled across subjects, and a normal distribution was fit to the histogram of these coefficients. We then randomly sampled the required number of coefficients from the normal distributions at each level, and performed a wavelet reconstruction to create a time course of the same length as the original fMRI data. A null fMRI time course was synthesized for each subject in the group, and these data were fed into the ISR pipeline with the acoustic predictors to create null parameter estimates, and a null group-level contrast value (i.e., the mean parameter estimate). This process was repeated 5000 times to create null distributions.

For between-group comparisons, the mean parameter estimate from one group was subtracted from the mean parameter estimate from the other group to create a between-group contrast value. This was compared to a null distribution that was created by performing the same calculation for 5000 null iterations. Again, real contrast values were considered statistically significant if they were outside 95% of the null (2.5% each tail).

Results

We first examined whether the temporal pattern of activity in auditory cortex evoked by a series of 15-second lullabies was consistent across adult individuals. All listeners heard the same lullabies in the same order, so that we could assess the inter-subject synchronization in the evoked fMRI signal. A multiple regression model (i.e., the inter-subject regression model) was constructed for each subject (and each

hemisphere), using leave-one-subject-out cross validation to assess the similarity of the response between people in a model-free way. The time course of activity in each subject's auditory cortex was modelled using a regressor comprising the average fMRI time series from the same region in all other subjects (i.e., the mean N-1 subjects' fMRI time series), and an additional regressor to account for the effect of the scanner onset. As the cross-validation folds are not independent from one another, a non-parametric one-sample test was used to assess whether the parameter estimates of these regressors were on average (across subjects) significantly different from zero. We found that the onset of the scanner drove cortical auditory responses in adults (Fig. 4C; p's < 0.001), and even when this effect was controlled for, auditory responses to the naturalistic stimulus were similar between adult subjects (Fig. 4C; p's < 0.001).

We then tested whether the infant response was in part mature, by comparing whether auditory responses in the infant groups were similar to the adults (3-months, Fig. 4A; 9-months, Fig. 4B). To do so, the ISR model for each infant comprised the average adult fMRI time series and a scanner onset regressor. Like the adults, the scanner onset drove responses in both 3- and 9-month old auditory cortex (Figs. 4A, 4B; p < 0.01). Importantly, though, beyond this the infant fMRI response was similar to the adults' (all p's < 0.001). The 3- and 9-month infant groups did not differ significantly in the magnitude of their responses (all p's > 0.5), and there was no significant difference between hemispheres for any group.

It is conceivable that these temporal patterns of signal fluctuation in auditory cortex were driven by the mere presence of sound (i.e., the on/off block design of auditory stimulation) or by amplitude variation within the songs. Such an effect could be the underlying reason why

infant and adult auditory responses appear to be similar. To test this hypothesis, we expanded the ISR model to include the envelope of auditory stimulation as another regressor of auditory cortex activity. We found that both 3- and 9-month old subjects' responses were in part driven by the envelope (Figs. 5A and B, all p's < 0.001), in addition to the scanner onset (p's < 0.001). Interestingly, there was a residual component of auditory-evoked activity that was not explained by these acoustic factors that was consistent between 3-month olds and adults (Fig. 5A; left auditory cortex p < 0.05, right auditory cortex p < 0.01) and between 9-month olds adults (Fig. 5B; left p < 0.01 and right p < 0.05), and there was no difference between 3- and 9-month olds in terms of how well any of these predictors explained activity in auditory cortex (all p's > 0.2). Again, the adult auditory responses in left and right auditory cortex were consistent across subjects (Fig. 5C; p's < 0.05), even when controlling for the fact that activity was driven by the auditory envelope (p's < 0.001) and the onset of the scanner (p's < 0.001).

Next, we aimed to determine whether the auditory-evoked fMRI responses in infants and adults were driven by other basic acoustic factors: the pitch trajectory (i.e., the fundamental frequency of the vocal), and brightness (i.e., a measure of the amount high-frequency energy in the sounds), which were extracted using the TANDEM-STRAIGHT software (Kawahara et al., 2008). We found that responses in adult auditory cortex were sensitive to both pitch (Fig. 6C; left p < 0.05, right p < 0.01) and brightness (p's < 0.001), in addition to the auditory envelope (p's < 0.001) and scanner onset (p's < 0.001). There was still some component of the shared auditory response not explained by these acoustic features (p's < 0.001). Three-month-old infants showed a similar pattern to the adults, such that auditory cortex
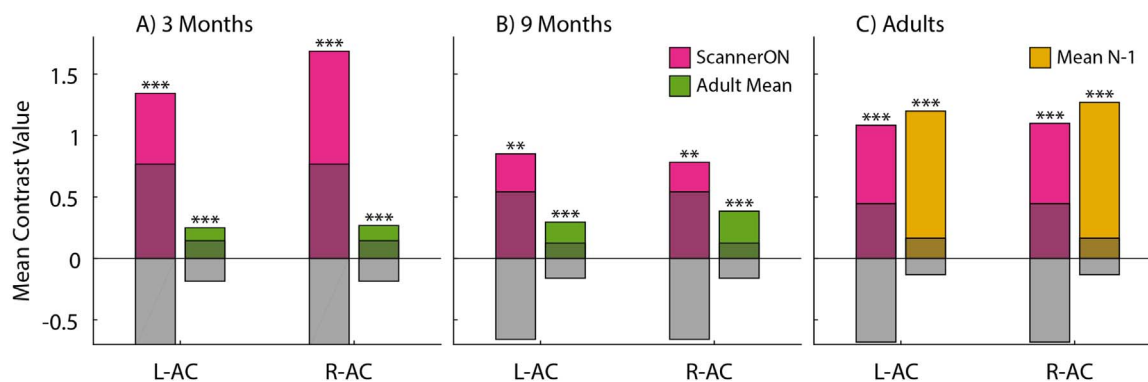

Fig. 4. ISR results in left and right auditory cortex (L-AC and R-AC, respectively), for A) 3 months old subjects; B) 9 month old subjects; and C) Adult subjects. For all subjects, the ISR model included the scannerON predictor (pink bar). The mean adult fMRI time series (green bar) was used as a predictor in the ISR model for all infants, whereas the mean N-1 adult signal (yellow) was included in the adult ISR model. Gray bars indicate the two-tailed 95% of the null distribution for each contrast (and each group, and ROI). Bars that extend beyond the gray are statistically different from the null: * p < 0.05; ** p < 0.01; *** p < 0.001.
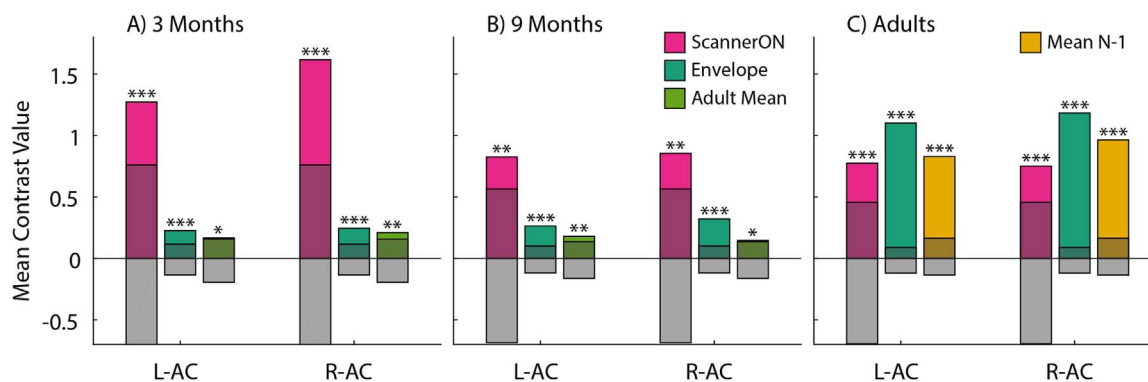

Fig. 5. ISR results in left and right auditory cortex (L-AC and R-AC), for A) 3 month, B) 9 month, and C) adult subjects. The ISR model also included the envelope of auditory stimulation (teal bar). Gray shaded regions indicate the two-tailed 95% of the null distribution for each contrast (and each group, and ROI). Effects that extend beyond gray bars are statistically different than the null: * p < 0.05; ** p < 0.01; *** p < 0.001.
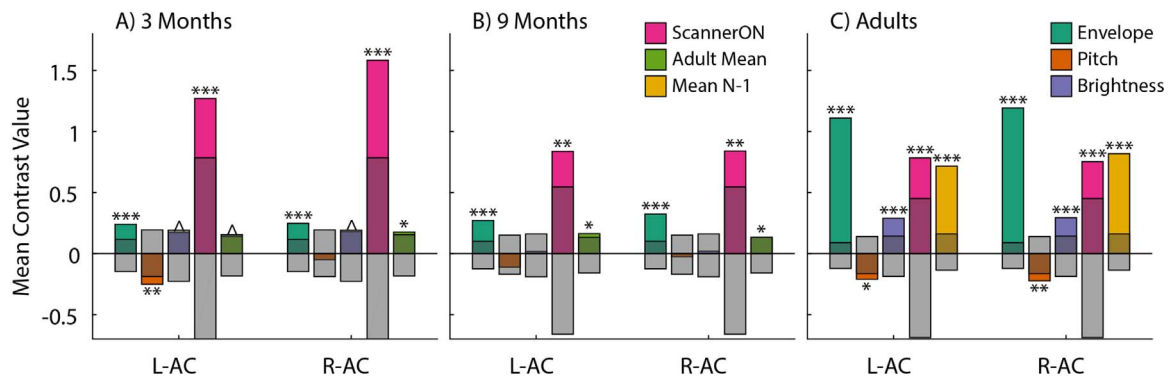
Fig. 6. ISR results in left and right auditory cortex, for A) 3 months old subjects; B) 9 month old subjects; and C) Adult subjects. This ISR model included two additional auditory *a priori* predictors – pitch (orange) and brightness (purple). Gray shaded regions indicate the two-tailed 95% of the null distribution for each contrast (and each group, and ROI). Effects that extend beyond the gray are statistically different than the null: ^ p < 0.10 (trend); * p < 0.05; ** p < 0.01; *** p < 0.001.

was sensitive to pitch (Fig. 6A; left p < 0.01; right p > 0.4), scanner onset (p's < 0.001), and envelope (p's < 0.001). There was a trend that 3-month cortex was sensitive to the auditory brightness (p's < 0.1). Nine-month old auditory cortex did not demonstrate sensitivity to pitch or brightness (Fig. 6B, all p's > 0.1), yet there was a still a significant response to both the envelope (p's < 0.001) and scanner onset (p's < 0.01). Despite the limited evidence that 3-month auditory responses were sensitive to pitch and brightness, and 9 months were not, we did not detect a significant difference between infant groups in how well these features predicted auditory cortex activity (all p's > 0.1). Therefore, we cannot conclude that sensitivity to these acoustic features is different between 3- and 9-months of age. However, we observed yet again that infant auditory responses were significantly similar to the adults', even when controlling for simple acoustic features (all p's < 0.05). This shows that infant auditory cortex is already tuned to some more complex acoustic features in common with adults.

Finally, we performed an exploratory analysis to investigate whether brain areas beyond auditory cortex in infants respond to, and process, sound similarly to adults. Functional neuroimaging studies of adult speech perception have revealed a distributed and hierarchical network of areas that process spoken language, including auditory regions in the temporal lobe (Scott et al., 2000; Davis and Johnsrude, 2003), and depending on the stimuli or experimental context, can include inferior frontal, motor, parietal, and subcortical regions (Davis and Johnsrude, 2003; Hickok and Poeppel, 2007; Peelle et al., 2010; Wild et al., 2012b). We used ISR to assess whether any of these regions responded to sound. We found in 3- and 9-month subjects, that prefrontal cortex, supplementary motor area, and inferior parietal areas exhibited synchronized auditory-evoked responses across subjects; however, ISR shows that this synchronized response is correlated with the scanner onset (Figs. 7A and B). Therefore, it is not clear from this result whether frontal regions were processing sound, or whether the synchronized activity reflects some confounding factor at scanner onset. This result at least highlights how care should be taken when interpreting inter-subject correlations, as they might be driven (at least in part) by confounding effects. In contrast to the infant results, adult listeners showed synchronized responses driven by the auditory envelope in the thalamus, cerebellum, prefrontal cortex, insula, and supplementary motor area (Fig. 7C). In adult prefrontal cortex, we found synchronized responses that could not be explained by any acoustic factors – this could represent higher-level linguistic processing of the lyrical content of the lullabies.

Discussion

These results show that infant auditory cortex at 3- and 9-months of age responds to sequences of rich and naturalistic sounds (sung

lullabies) in a way that is similar to adult auditory cortex, and thus that it is capable of performing some mature complex auditory processing despite the apparent immaturity of its structure and afferent connectivity. Importantly, we found that the similarity in these temporal patterns of auditory-evoked activity was not driven by trivial aspects of the listening experience, such as the sound of the scanner starting, or the fact that songs were presented in an on/off block design; therefore, the activity we observed in infant auditory cortex reflected stimulus-specific processing.

Our inter-subject regression method allowed us to attribute components of these shared and complex auditory responses to different features of the auditory stimulation, such as the amplitude envelope, pitch, and the amount of high frequency energy. We found that auditory cortex at 3 months was sensitive to pitch and envelope, and there was some evidence for spectral (frequency) sensitivity. The pitch sensitivity found is consistent with EEG evidence of cortical mismatch responses (MMRs) to pitch deviants in young infants (e.g., He et al., 2007a), and similar EEG work has shown evidence for frequency (Bisiacchi et al., 2009) and timbre (Trainor et al., 2011) sensitivity. The results further show that extraction of theses auditory features happens in the context of much more complex stimuli, where changes in pitch are intertwined with dramatic changes in spectral shape and amplitude. Also, to our knowledge, this work is the first to show that activity in infant auditory cortex correlates with the amplitude envelope of complex sounds, as it does in adults (Nourski et al., 2009; Kubanek et al., 2013). This finding reinforces the importance of accounting or controlling for physical differences between acoustic stimuli. Previous studies of processing in infant auditory cortex have contrasted responses to speech with those evoked by different control sounds, such as: hummed or flattened speech (Perani et al., 2011), backwards speech (Dehaene-Lambertz et al., 2002), non-voice sounds (Blasi et al., 2011b), and music (Perani et al., 2010). However, auditory brain responses observed in these experiments might reflect differences in amplitude envelope or pitch. Our findings extend beyond these studies by showing that cortical responses evoked by rich acoustic stimulation are driven by acoustic features such as amplitude, in addition to more complex, but unspecified, auditory features. This, of course, begs the question: what does the component of the shared auditory-evoked signal that is unaccounted for by these *a priori* acoustic predictors tell us about auditory processing?

That infants show synchronized cortical auditory responses with adults (and adults show synchronized responses with each other) when controlling for simple acoustic features suggests that this residual shared signal represents some higher-level aspect of auditory processing that is not predicted from the physical stimulus. One particularly enticing explanation is that it could indicate the processing of more complex acoustic features, such as spectro-temporal modulations (Hall et al., 2002b; Hullett et al., 2016), which are critical for speech
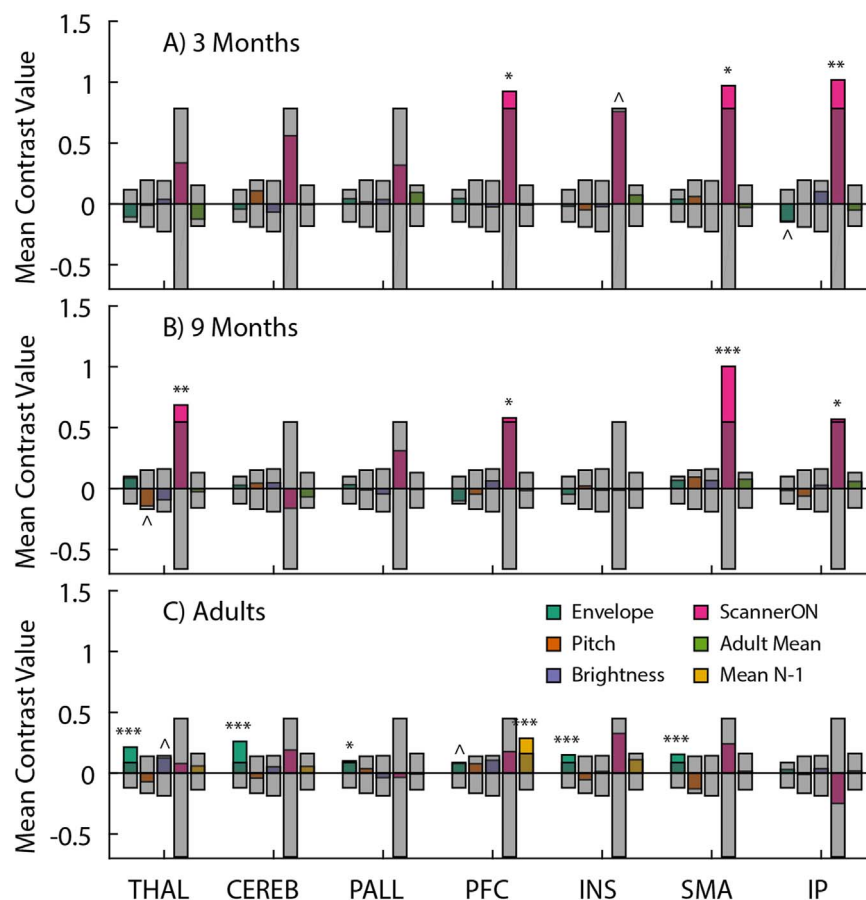
Fig. 7. ISR results for brain regions beyond auditory cortex – thalamus (THA), cerebellum (CEREB), pallidum (PALL), prefrontal cortex (PFC), insula (INS), supplementary motor area (SMA), and inferior parietal lobule (IP). Results are averaged across left and right hemispheres. Gray bars indicate the two-tailed 95% of the null distribution for each contrast (and each group, and ROI). Effects that extend beyond the gray are statistically different from the null: ^ $p < 0.05$ (trend, one-tailed); * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

recognition (Pasley et al., 2012), or even phonemes (Chang et al., 2010; Mesgarani et al., 2014). It is possible that higher-order non-speech related features underlie these shared responses, such as pitch strength, harmonics-to-noise ratio, the temporal variability of amplitude (Leaver and Rauschecker, 2010), or features of the music such as familiarity. It was not possible in this study to eliminate every potential acoustic feature; rather, we wanted to investigate whether there is any similarity in how the developing (i.e. very young) and mature cortex respond to natural sounds, and if so, whether this similarity could be explained by basic auditory features that are often not controlled for in the neuroimaging literature on infant speech perception. Future work could attempt to model other complex auditory features to further specify what sources of auditory information – specifically speech-related or otherwise – drive these responses that are shared by very young and mature brains. Nonetheless, our results and methods provide a stepping-stone from which we can build a better characterization of the cortical representations of naturalistic sounds, and how they mature.

One potential limitation of this study is that infants were scanned during natural sleep, whereas adults were scanned while awake. While the overall magnitude of auditory-evoked responses in temporal lobe auditory regions can be attenuated by sleep (Czisch et al., 2002), sedation (Dueck et al., 2005), or distraction (Wild et al., 2012b), the sensitivity of these lower-level auditory areas to different kinds of sounds and acoustic features does not depend on the listeners' state (Davis et al., 2007; Issa and Wang, 2008; Wild et al., 2012b). The fact that we observed synchrony between awake adult and sleeping infant listeners only reinforces this point. On the other hand, the processing

of ecologically salient sounds (e.g., speech) in frontal and prefrontal regions has been shown to depend on the state of the listener (Davis et al., 2007; Wild et al., 2012b; Adapa et al., 2014), and so one plausible explanation for the lack of coupling between adults and infants in frontal regions is that infants were asleep, whereas adults were not.

We did not observe any differences in auditory processing between 3- and 9-month old listeners. This could be interpreted as inconsistent with the idea these ages occur before and during a time of significant developmental change in terms of anatomy (Moore, 2002; Moore and Linthicum, 2007), behaviour (Kuhl et al., 1992; Kuhl, 2010; Maurer and Werker, 2014), and electrophysiology (Pang et al., 1998; Trainor et al., 2003; He et al., 2007b). On the other hand, by six months of age infants already possess some sophisticated speech perception, such as the ability to recognize a limited repertoire of words (Bergelson and Swingley, 2012, 2013; Vouloumanos et al., 2014). Our data support a model in which auditory cortex functioning is already established in these early months, tuned to a variety of acoustic features like the adult auditory cortex, and hence might support complex auditory behaviours. Therefore, it could be the case that any developmental changes in functional processing over the period we observed (from 3 to 9 months of age) were too subtle to detect with our paradigm, stimuli, and/or sample size. Future studies will need to run larger cohorts to further investigate whether there are differences that were too small to be detected by our current study. Nonetheless, whereas most behavioural and electrophysiological studies of infant development emphasize the differences between age groups, we have shown that it is possible to quantify their similarities, and that this provides novel

information about the maturity of auditory processing in very young infants – specifically, that functional responses in auditory cortex evoked by rich and naturalistic sounds are measurably adult-like.

Other neuroimaging studies have suggested that infant auditory cortex displays adult-like functional activations (Dehaene-Lambertz et al., 2006; Perani et al., 2011), but their conclusions were based on the observation that similar brain regions respond to sound. Again, our results go further than these studies by showing that the overall temporal pattern of rich auditory-evoked activity is similar between infants and adults. At the least, this suggests that activity in auditory cortex in infants and adults is driven by the same parts of the songs, and therefore that they show tuning to similar kinds of features. It is not clear at what timescale this similarity is present and whether it reflects segmental or suprasegmental features, which can be detected with rapid fMRI (Lewis et al., 2016).

That we observed adult-like auditory tuning at such an early age reinforces the potential importance of early auditory experience. A lack of rich auditory stimulation in the late gestational and perinatal period might inhibit the development of these cortical representations, and subsequently impair spoken language development. Indeed, it has been shown that a quiet NICU auditory environment affects the development of auditory cortex of prematurely born infants (Pineda et al., 2014; Webb et al., 2015), and that increased exposure to language during this time improves their language outcomes (Caskey et al., 2014; Pineda et al., 2014). It is worth noting that many of our infants (6/13) were in fact born very prematurely. The implication is that these infants had longer exposure to an ex-utero acoustic environment than their term-born peers. However, it not clear what effect we might expect this exposure to have on our results. Is the pre-term auditory system more mature because of earlier exposure to this environment? Or is it impaired (or developing differently) compared to the full term auditory system (e.g., due to the aforementioned effect of a too-quiet acoustic environment)? Future work should examine whether the markers of functional auditory processing that we observed are affected by premature birth and early auditory experience.

In fact, our study is not the first to examine the inter-subject synchrony in brain responses between young subjects and adults. Cantlon and Li (2013) found that a measure of neural maturity in 4–11 year old children – that is, how similar the childrens' brain activity (recorded with fMRI) was to adults' while viewing Sesame Street videos – predicted the their mathematical and language abilities. Though we could not do so with our dataset, our method and paradigm might similarly be used to predict speech and language outcomes following adverse birth events that increase the risk of neurodevelopmental delay, such as very preterm birth (Lee et al., 2011; Vohr, 2013).

In conclusion, we have developed a novel and powerful analysis technique – inter-subject regression – that revealed that the immature infant auditory cortex in the first 6 months is capable of performing adult-like auditory processing. Unlike other infant neuroimaging studies we were able to rule out less interesting causes of these brain responses, and can therefore could conclude that they reflect stimulus-specific processing. We hope to push this work further in future studies by better characterizing exactly what sound features evoke these adult-like brain responses, how they develop in tandem with structural maturation, how they might be disrupted by neurological disorders, and whether this information can be used to predict neurodevelopmental outcomes.

Conflict of interest

The authors have no conflicts of interest to disclose.

Acknowledgements

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.neuroimage.2017.06.038.

References

Adapa, R.M., Davis, M.H., Stamatakis, E.A., Absalom, A.R., Menon, D.K., 2014. Neural correlates of successful semantic processing during propofol sedation. Hum. Brain Mapp. 35, 2935–2949.

Aguirre, G.K., Zarahn, E., D'esposito, M., 1998. The variability of human, BOLD hemodynamic responses. Neuroimage 8, 360–369.

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. Magn. Reson Imaging 19, 1–5.

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. NeuroImage development of BOLD signal hemodynamic responses in the human brain. Neuroimage 63, 663–673.

Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26, 839–851.

Baldoli, C., Scola, E., Della Rosa, P.A., Pontesilli, S., Longaretti, R., Poloniato, A., Scotti, R., Blasi, V., Cirillo, S., Iadanza, A., Rovelli, R., Barera, G., Scifo, P., 2014. Maturation of preterm newborn brains: a fMRI-DTI study of auditory processing of linguistic stimuli and white matter development. Brain Struct. Funct..

Bergelson, E., Swingley, D., 2012. At 6-9 months, human infants know the meanings of many common nouns. Proc. Natl. Acad. Sci. USA. 109, 3253–3259.

Bergelson, E., Swingley, D., 2013. The acquisition of abstract words by young infants. Cognition 127, 391–397.

Bisiacchi, P.S., Mento, G., Suppiej, A., 2009. Cortical auditory processing in preterm newborns: an ERP study. Biol. Psychol. 82, 176–185.

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., 2011a. Early specialization for voice and emotion processing in the infant brain. Curr. Biol. 21, 1220–1224.

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., 2011b. Early specialization for voice and emotion processing in the infant brain. Curr. Biol. 21, 1220–1224.

Bullmore, E., Long, C., Suckling, J., Fadiwj, J., Calvertl, G., Zelaya, F., Carpenter, A., Brammers, M., 2001. Colored noise and computational inference in fMRI time series analysis: resampling methods in time and wavelet domains Slnstitute of Psychiatry KCL, London UK. Hum. Brain Mapp. 12, 2001.

Cantlon, J.F., Li, R., 2013. Neural activity during natural viewing of sesame street statistically predicts test scores in early childhood. PLoS Biol., 11.

Caskey, M., Stephens, B., Tucker, R., Vohr, B., 2014. Adult talk in the NICU with preterm infants and developmental outcomes. Pediatrics 133, e578–e584.

Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M., Knight, R.T., 2010. Categorical speech representation in human superior temporal gyrus. Nat. Neurosci. 13, 1428–1432.

Cheour, M., Ceponiene, R., Lehtokoski, A., 1998a. Development of language-specific phoneme representations in the infant brain. Nat. …. 1, 351–353.

Cheour, M., Ceponiene, R., Lehtokoski, A., 1998b. Development of language-specific phoneme representations in the infant brain. Nat. …. 1, 351–353.

Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., Aaltonen, O., Eerola, O., Näätänen, R., 1995a. Mismatch negativity indicates vowel discrimination in newborns. Hear Res. 82, 53–58.

Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., Aaltonen, O., Eerola, O., Näätänen, R., 1995b. Mismatch negativity indicates vowel discrimination in newborns. Hear Res. 82, 53–58.

Cusack, R., Vicente-Grabovetsky, A., Mitchell, D.J., Wild, C.J., Auer, T., Linke, A.C., Peelle, J.E., 2015a. Automatic analysis (aa): efficient neuroimaging workflows and parallel processing using Matlab and XML. Front Neuroinform. 8, 1–13.

Cusack, R., Wild, C., Linke, A.C., Arichi, T., Lee, D.S.C., Han, V.K., 2015b. Optimizing stimulation and analysis protocols for neonatal fMRI. PLoS One, 10.

Czisch, M., Wetter, T.C., Kaufmann, C., Pollmacher, T., Holsboer, F., Auer, D.P., 2002. Altered Processing of Acoustic Stimuli during Sleep: reduced Auditory Activation and Visual Deactivation Detected by a Combined fMRI/EEG Study. Neuroimage 16, 251–258.

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. Proc. Natl. Acad. Sci. 104, 16032.

Davis, M.H., Johnsrude, I.S., 2003. Hierarchical processing in spoken language comprehension. J. Neurosci. 23, 3423.

DeCasper, A., Fifer, W., 1980. Of human bonding: newborns prefer their mothers' voices.

Sci. (80-). 208, 1174–1176.

Dehaene-Lambertz, G., Dehaene, S., Hertz-Pannier, L., 2002. Functional neuroimaging of speech perception in infants. Science 298, 2013–2015.

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Mériaux, S., Roche, A., Sigman, M., Dehaene, S., 2006. Functional organization of perisylvian activation during presentation of sentences in preverbal infants. Proc. Natl. Acad. Sci. USA 103, 14240–14245.

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 Lang. 114, 53–65.

Draganova, R., Eswaran, H., Murphy, P., Huotilainen, M., Lowery, C., Preissl, H., 2005. Sound frequency change detection in fetuses and newborns, a magnetoencephalographic study. Neuroimage 28, 354–361.

Dueck, M.H., Petzke, F., Gerbershagen, H.J., Paul, M., Heßelmann, V., Girnus, R., Krug, B., Sorger, B., Goebel, R., Lehrke, R., Sturm, V., Boerner, U., 2005. Propofol attenuates responses of the auditory cortex to acoustic stimulation in a dose-dependent manner: a FMRI study. Acta Anaesthesiol. Scand. 49, 784–791.

Eggermont, J.J., Moore, J.K., 2012. Morphological and Functional Development of the Auditory Nervous System. Human Auditory Development.

Eimas, P.D., 1975. Auditory and phonetic coding of the cues for speech: discrimination of the [r-l] distinction by young infants. Percept. Psychophys. 18, 341–347.

Eswaran, H., Haddad, N.I., Shihabuddin, B.S., Preissl, H., Siegel, E.R., Murphy, P., Lowery, C.L., 2007. Non-invasive detection and identification of brain activity patterns in the developing fetus. Clin. Neurophysiol. 118, 1940–1946.

Feinberg, D. a., Moeller, S., Smith, S.M., Auerbach, E., Ramanna, S., Gunther, M., Glasser, M.F., Miller, K.L., Ugurbil, K., Yacoub, E., 2010. Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. PLoS One 5, e15710.

Giraud, A.L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., Kleinschmidt, A., 2000. Representation of the temporal envelope of sounds in the human brain. J. Neurophysiol. 84, 1588–1598.

Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.R., Gurney, E.M., Bowtell, R.W., 1999. "sparse" temporal sampling in auditory fMRI. Hum. Brain Mapp. 7, 213–223.

Hall, D.A., Johnsrude, I.S., Haggard, M.P., Palmer, A.R., Akeroyd, M.A., Summerfield, A.Q., 2002a. Spectral and Temporal Processing in Human Auditory Cortex. Cereb. Cortex. 12, 140–149.

Hall, D.A., Johnsrude, I.S., Haggard, M.P., Palmer, A.R., Akeroyd, M.A., Summerfield, A.Q., 2002b. Spectral and temporal processing in human auditory cortex. Cereb. Cortex. 12, 140.

Hall, D.A., Plack, C.J., 2009. Pitch processing sites in the human auditory brain. Cereb. Cortex. 19, 576–585.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject synchronization of cortical activity DuringNatural vision. Science (80-) 303, 1634–1640.

He, C., Hotson, L., Trainor, L.J., 2007a. Mismatch responses to pitch changes in early infancy. J. Cogn. Neurosci. 19, 878–892.

He, C., Hotson, L., Trainor, L.J., 2007b. Mismatch responses to pitch changes in early infancy. J. Cogn. Neurosci. 19, 878–892.

Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402.

Holst, M., Eswaran, H., Lowery, C., Murphy, P., Norton, J., Preissl, H., 2005. Development of auditory evoked fields in human fetuses and newborns: a longitudinal MEG study. Clin. Neurophysiol. 116, 1949–1955.

Homae, F., Watanabe, H., Nakano, T., Taga, G., 2011. Large-scale brain networks underlying language acquisition in early infancy. Front Psychol. 2, 93.

Hullett, P.W., Hamilton, X.L.S., Mesgarani, N., Schreiner, X.C.E., Chang, E.F., 2016. Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli. J. Neurosci. 36, 2014–2026.

Issa, E.B., Wang, X., 2008. Sensory responses during sleep in primate primary and secondary auditory cortex. J. Neurosci. 28, 14467–14480.

Jardri, R., Pins, D., Houfflin-Debarge, V., Chaffiotte, C., Rocourt, N., Pruvo, J.-P., Steinling, M., Delion, P., Thomas, P., 2008. Fetal cortical activation to sound at 33 weeks of gestation: a functional MRI study. Neuroimage 42, 10–18.

Jusczyk, P.W., Pisoni, D.B., Mullennix, J., 1992. Some consequences of stimulus variability on speech processing by 2-month-old infants. Cognition 43, 253–291.

Kawahara H., Morise M., Takahashi T., Nisimura R., Irino T., Banno H., 2008. Tandem-straight: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. ICASSP, IEEE International Conference Acoust Speech Signal Process - Proceedings 3933–3936.

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. Hum. Brain Mapp. 31, 595–603.

Kubanek, J., Brunner, P., Gunduz, A., Poeppel, D., Schalk, G., 2013. The tracking of speech envelope in the human cortex. PLoS One, 8.

Kuhl, P., Williams, K., Lacerda, F., Stevens, K., Lindblom, B., 1992. Linguisitic experience alters Phonetic perception in infants by 6 months of age. Sci. (80-). 255, 606–608.

Kuhl, P.K., 1979. Speech perception in early infancy: perceptual constancy for spectrally dissimilar vowel categories. J. Acoust. Soc. Am. 66, 1668–1679.

Kuhl, P.K., 2010. Brain Mechanisms in Early Language Acquisition. Neuron 67, 713–727.

Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., Nelson, T., 2008. Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). Philos. Trans. R. Soc. Lond. B Biol. Sci. 363,

979–1000.

Kuhl, P.K., Ramirez, R.R., Bosseler, a., Lin, J.-F.L., Imada, T., 2014. Infants' brain responses to speech suggest analysis by Synthesis. Proc. Natl. Acad. Sci..

Leaver, A.M., Rauschecker, J.P., 2010. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J. Neurosci. 30, 7604–7612.

Lee, E.S., Yeatman, J.D., Luna, B., Feldman, H.M., 2011. Specific language and reading skills in school-aged children and adolescents are associated with prematurity after controlling for IQ. Neuropsychologia 49, 906–913.

Lewis, L.D., Setsompop, K., Rosen, B.R., Polimeni, J.R., 2016. Fast fMRI can detect oscillatory neural activity in humans. Proc. Natl. Acad. Sci..

Linke, A.C., Vicente-Grabovetsky, A., Cusack, R., 2011a. Stimulus-specific suppression preserves information in auditory short-term memory. Proc. Natl. Acad. Sci. USA 108, 12961–12966.

Linke, A.C., Vicente-Grabovetsky, A., Cusack, R., 2011b. Stimulus-specific suppression preserves information in auditory short-term memory. Proc. Natl. Acad. Sci. USA 108, 12961–12966.

Mahmoudzadeh, M., Dehaene-Lambertz, G., Fournier, M., Kongolo, G., Goudjil, S., Dubois, J., Grebe, R., Wallois, F., 2013. Syllabic discrimination in premature human infants prior to complete formation of cortical layers. Proc. Natl. Acad. Sci. 110, 1–6.

Mahmoudzadeh, M., Wallois, F., Kongolo, G., Goudjil, S., Dehaene-Lambertz, G., 2016. Functional Maps at the Onset of Auditory Inputs in Very Early Preterm Human Neonates. Cereb. Cortex. bhw1, 03.

Maurer, D., Werker, J.F., 2014. Perceptual narrowing during infancy: a comparison of language and faces. Dev. Psychobiol. 56, 154–178.

Mesgarani, N., Cheung, C., Johnson, K., Chang, E.F., 2014. Phonetic feature encoding in human superior temporal gyrus. Science 343, 1006–1010.

Millman, R.E., Prendergast, G., Hymers, M., Green, G.G.R., 2013. Representations of the temporal envelope of sounds in human auditory cortex: can the results from invasive intracortical "depth" electrode recordings be replicated using non-invasive MEG "virtual electrodes"? Neuroimage 64, 185–196.

Minagawa-Kawai, Y., Van Der Lely, H., Ramus, F., Sato, Y., Mazuka, R., Dupoux, E., 2011. Optical brain imaging reveals general auditory and language-specific processing in early infant development. Cereb. Cortex. 21, 254–261.

Moore, J.K., 2002. Maturation of human auditory cortex: implications for speech perception. Ann. Otol. Rhinol. Laryngol. Suppl..

Moore, J.K., Guan, Y.-L., 2001. Cytoarchitectural and axonal maturation in human auditory cortex. JARO - J. Assoc. Res Otolaryngol. 2, 297–311.

Moore, J.K., Linthicum, F.H., 2007. The human auditory system: a timeline of development. Int J. Audiol. 46, 460–478.

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, 99. pp. 94–99.

Nourski, K.V., Reale, R. a., Oya, H., Kawasaki, H., Kovach, C.K., Chen, H., Howard, M. a., Brugge, J.F., 2009. Temporal envelope of time-compressed speech represented in the human auditory cortex. J. Neurosci. 29, 15564–15574.

Pang, E.W., Edmonds, G.E., Desjardins, R., Khan, S.C., Trainor, L.J., Taylor, M.J., 1998. Mismatch negativity to speech stimuli in 8-month-old infants and adults. Int J. Psychophysiol. 29, 227–236.

Pasley, B.N., David, S.V., Mesgarani, N., Flinker, A., Shamma, S. a., Crone, N.E., Knight, R.T., Chang, E.F., 2012. Reconstructing speech from human auditory cortex. PLoS Biol. 10, e1001251.

Patterson, R.D., Uppenkamp, S., Johnsrude, I.S., Griffiths, T.D., 2002. The processing of temporal pitch and melody information in auditory cortex. Neuron 36, 767–776.

Peelle J.E., Johnsrude I.S., Davis M.H., 2010. Hierarchical Processing for Speech in Human Auditory Cortex and Beyond, 4.

Peña, M., Maki, A., Kovačić, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., Mehler, J., 2003. Sounds and silence: an optical topography study of language recognition at birth. Proc. Natl. Acad. Sci. USA 100, 11702–11705.

Perani, D., Saccuman, M.C., Scifo, P., Anwander, A., Awander, A., Spada, D., Baldoli, C., Poloniato, A., Lohmann, G., Friederici, A.D., 2011. Neural language networks at birth. Proc. Natl. Acad. Sci. 108, 1–6.

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. Proc. Natl. Acad. Sci. USA 107, 4758–4763.

Pineda, R.G., Neil, J., Dierker, D., Smyser, C.D., Wallendorf, M., Kidokoro, H., Reynolds, L.C., Walker, S., Rogers, C., Mathur, A.M., Van Essen, D.C., Inder, T., 2014. Alterations in brain structure and neurodevelopmental outcome in preterm infants hospitalized in different neonatal intensive care unit environments. J. Pediatr. 164 (52–60), e2.

Remez, R.E., Pardo, J., Piorkowski, R., Rubin, P., 2001. On the Bistability of Sine wave analogues of speech. Psychol. Sci. 12, 24–29.

Rivkin, M.J., Wolraich, D., Als, H., McAnulty, G., Butler, S., Conneman, N., Fischer, C., Vajapeyam, S., Robertson, R.L., Mulkern, R.V., 2004. Prolonged T*2 values in newborn versus adult brain: implications for fMRI studies of newborns. Magn. Reson Med. 51, 1287–1291.

Sato, H., Hirabayashi, Y., Tsubokura, H., Kanai, M., Ashida, T., Konishi, I., Uchida-Ota, M., Konishi, Y., Maki, A., 2011. Cerebral hemodynamics in newborn infants exposed to speech sounds: a whole-head optical topography study. Hum. Brain Mapp. 00, 1–12.

Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain A J. Neurol. 123 (Pt 12), 2400–2406.

Shi, F., Yap, P.-T., Wu, G., Jia, H., Gilmore, J.H., Lin, W., Shen, D., 2011. Infant brain atlases from neonates to 1- and 2-year-olds. PLoS One 6, e18746.

Shultz, S., Vouloumanos, A., Bennett, R.H., Pelphrey, K., 2014. Neural specialization for

speech in the first months of life. Dev. Sci. 5, 766–774.

Sohoglu, E., Peelle, J.E., Carlyon, R.P., Davis, M.H., 2012. Predictive Top-Down Integration of Prior Knowledge during Speech Perception. J. Neurosci. 32, 8443–8453.

Talavage, T.M., Ledden, P.J., Benson, R.R., Rosen, B.R., Melcher, J.R., 2000. Frequency-dependent responses exhibited by multiple regions in human auditory cortex. Hear Res. 150, 225–244.

Trainor, L., McFadden, M., Hodgson, L., Darragh, L., Barlow, J., Matsos, L., Sonnadara, R., 2003. Changes in auditory cortex and the development of mismatch negativity between 2 and 6 months of age. Int J. Psychophysiol. 51, 5–15.

Trainor, L.J., Lee, K., Bosnyak, D.J., 2011. Cortical plasticity in 4-month-old infants: specific effects of experience with musical timbres. Brain Topogr. 24, 192–203.

Trehub, S.E., 1973. Infants' sensitivity to vowel and tonal contrasts. Dev. Psychol. 9, 91–96.

Trehub, S.E., Rabinovitch, M.S., 1972. Auditory-linguistic sensitivity in early infancy. Dev. Psychol. 6, 74–77.

Vannasing, P., Florea, O., González-Frankenberger, B., Tremblay, J., Paquette, N., Safi, D., Wallois, F., Lepore, F., Béland, R., Lassonde, M., Gallagher, A., 2016. Distinct hemispheric specializations for native and non-native languages in one-day-old newborns identified by fNIRS. Neuropsychologia 84, 63–69.

Vohr, B., 2013. Speech and language outcomes of very preterm infants. Semin Fetal Neonatal Med..

Vouloumanos, A., Martin, A., Onishi, K.H., 2014. Do 6-month-olds understand that speech can communicate? Dev. Sci. 6, 872–879.

Vouloumanos, A., Werker, J.F., 2004. Tuned to the signal: the privileged status of speech for young infants. Dev. Sci. 7, 270–276.

Vouloumanos, A., Werker, J.F., 2007. Listening to language at birth: evidence for a bias for speech in neonates. Dev. Sci. 10, 159–164.

Webb, A.R., Heller, H.T., Benson, C.B., Lahav, A., 2015. Mother's voice and heartbeat sounds elicit auditory plasticity in the human brain before full gestation. Proc. Natl. Acad. Sci. USA 112, 3152–3157.

Werker, J.F., Hensch, T.K., 2014. Critical periods in speech perception: new directions. Annu Rev. Psychol., 1–24.

Werker, J.F., Tees, R.C., 1984. Cross language speech perception: evidence for perceptual reorganization during the first year of life. Infant Behav. Dev. 7, 49–63.

Wild, C.J., Davis, M.H., Johnsrude, I.S., 2012a. Human auditory cortex is sensitive to the perceived clarity of speech. Neuroimage 60, 1490–1502.

Wild, C.J., Yusuf, A., Wilson, D.E., Peelle, J.E., Davis, M.H., Johnsrude, I.S., 2012b. Effortful listening: the processing of degraded speech depends critically on attention. J. Neurosci. 32, 14010–14021.

Xu, J., Moeller, S., Auerbach, E.J., Strupp, J., Smith, S.M., Feinberg, D. a., Yacoub, E., Uğurbil, K., 2013. Evaluation of slice accelerations using multiband echo planar imaging at 3 T. Neuroimage.

Yarkoni, T., Poldrack, R. a., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods 8, 665–670.