Behavioral/Cognitive

Cortical Tracking of Speech-in-Noise Develops from Childhood to Adulthood

Marc Vander Ghinst,[1,2*] Mathieu Bourguignon,[1,3,5*] Maxime Niesen,[1,2] Vincent Wens,[1,4] Sergio Hassid,[2] Georges Choufani,[2] Veikko Jousmäki,[6,7] Riitta Hari,[8] Serge Goldman,[1,4] and Xavier De Tiège[1,4]

[1]Laboratoire de Cartographie fonctionnelle du Cerveau, UNI–ULB, Neuroscience Institute, [2]Service d'ORL et de chirurgie cervico-faciale, CUB Hôpital Erasme, Université Libre de Bruxelles (ULB), [3]Laboratoire Cognition Langage et Développement, UNI–ULB Neuroscience Institute, [4]Department of Functional Neuroimaging, Service of Nuclear Medicine, CUB Hôpital Erasme, Université libre de Bruxelles (ULB), 1070 Brussels, Belgium, [5]Basque Center on Cognition, Brain and Language (BCBL), 20009 Donostia, Spain, [6]Aalto NeuroImaging, Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, 00076 AALTO, Espoo, Finland, [7]Cognitive Neuroimaging Centre, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 636921, and [8]Department of Art, Aalto University School of Arts, Design, and Architecture, 00076 AALTO, Helsinki, Finland

In multitalker backgrounds, the auditory cortex of adult humans tracks the attended speech stream rather than the global auditory scene. Still, it is unknown whether such preferential tracking also occurs in children whose speech-in-noise (SiN) abilities are typically lower compared with adults. We used magnetoencephalography (MEG) to investigate the frequency-specific cortical tracking of different elements of a cocktail party auditory scene in 20 children (age range, 6–9 years; 8 females) and 20 adults (age range, 21–40 years; 10 females). During MEG recordings, subjects attended to four different 5 min stories, mixed with different levels of multitalker background at four signal-to-noise ratios (SNRs; noiseless, +5, 0, and −5 dB). Coherence analysis quantified the coupling between the time courses of the MEG activity and attended speech stream, multitalker background, or global auditory scene, respectively. In adults, statistically significant coherence was observed between MEG signals originating from the auditory system and the attended stream at <1, 1–4, and 4–8 Hz in all SNR conditions. Children displayed similar coupling at <1 and 1–4 Hz, but increasing noise impaired the coupling more strongly than in adults. Also, children displayed drastically lower coherence at 4–8 Hz in all SNR conditions. These results suggest that children's difficulties to understand speech in noisy conditions are related to an immature selective cortical tracking of the attended speech streams. Our results also provide unprecedented evidence for an acquired cortical tracking of speech at syllable rate and argue for a progressive development of SiN abilities in humans.

Key words: coherence analysis; magnetoencephalography; speech-in-noise

Significance Statement

Behaviorally, children are less proficient than adults at understanding speech-in-noise. Here, neuromagnetic signals were recorded while healthy adults and typically developing 6- to 9-year-old children attended to a speech stream embedded in a multitalker background noise with varying intensity. Results demonstrate that auditory cortices of both children and adults selectively track the attended speaker's voice rather than the global acoustic input at phrasal and word rates. However, increments of noise compromised the tracking significantly more in children than in adults. Unexpectedly, children displayed limited tracking of both the attended voice and the global acoustic input at the 4–8 Hz syllable rhythm. Thus, both speech-in-noise abilities and cortical tracking of speech syllable repetition rate seem to mature later in adolescence.

Introduction

Children often grow up and learn in noisy surroundings. Clamorous classrooms, rowdy playgrounds, and domestic sound disturbances indeed constitute adverse auditory scenes for a still immature auditory system.

Interestingly, speech-in-noise (SiN) perception in children appears strenuous but improves during late childhood (≥10 years) due to maturation of the auditory system and attentional abilities (Elliott, 1979; Moore et al., 2010; Sanes and Woolley, 2011; Thompson et al., 2017). Still, the neurophysiological mech-

Received July 10, 2018; revised Jan. 8, 2019; accepted Jan. 12, 2019.

Author contributions: M.V.G., M.B., S.H., G.C., S.G., and X.D.T. designed research; M.V.G., M.N., and X.D.T. performed research; M.B., M.N., V.W., V.J., and R.H. contributed unpublished reagents/analytic tools; M.V.G., M.B., M.N., V.W., and X.D.T. analyzed data; M.V.G., M.B., R.H., and X.D.T. wrote the paper.

This study and the MEG project at the CUB Hôpital Erasme were financially supported by the Fonds Erasme (research convention "Les Voies du Savoir," Fonds Erasme, Brussels, Belgium). M.V.G. and M.N. were supported by a research grant from the Fonds Erasme (Brussels, Belgium). M.B. was supported by the program Attract of Innoviris (Grant 2015-BB2B-10), by the Spanish Ministry of Economy and Competitiveness (Grant PSI2016-77175-P), and by

anisms accounting for the improvement in SiN perception observed from childhood to adulthood are unsettled. It has been hypothesized that, in adverse auditory scenes, children's auditory system would actually lack the capacity to segregate the attended auditory stream from the unattended noisy background (Sussman and Steinschneider, 2009; Sanes and Woolley, 2011). However, no study has so far confirmed this speculation in children.

Accumulating evidence shows that adults' auditory system tracks the attended speech stream rather than the global auditory scene in a multitalker background (Mesgarani and Chang, 2012). In such adverse auditory background, the auditory system entrains to the slow amplitude modulations (i.e., the temporal envelope) of the attended speaker's voice rather than to modulations of the global auditory scene (Ding and Simon, 2012a; Zion Golumbic et al., 2013; Vander Ghinst et al., 2016). This coupling typically occurs at frequencies <10 Hz and declines with increasing noise level. Given that this frequency range matches with prosodic stress/phrasal/sentential (<1 Hz), word (1–4 Hz), and syllable (4–8 Hz) repetition rates, the corresponding cortical tracking of speech has been hypothesized to subserve the chunking of the continuous verbal flow into relevant segments used for further speech recognition, up to a certain noise level (Ding and Simon, 2012a, 2013a; Giraud and Poeppel, 2012; Vander Ghinst et al., 2016; Keitel et al., 2018).

How children's auditory system tracks connected speech has remained largely unknown. Until now, cortical tracking of speech in children has only involved dyslexic children older than 11 years, showing that they have impaired tracking at frequencies <2 Hz compared with age-matched healthy control subjects (Molinaro et al., 2016; Power et al., 2016). However, to the best of our knowledge, no study has so far investigated how this low-frequency cortical tracking might differ between typically developed children and adults, and whether noise differentially corrupts the coupling in these two populations. We therefore specifically test the hypothesis that children's poor SiN perception abilities (Elliott, 1979; Berman and Friedman, 1995; Wightman and Kistler, 2005) are related to inaccurate low-frequency cortical tracking of the attended speech stream in a noisy background.

To test this hypothesis, children (6–9 years of age) and adults (21–40 years of age) with normal SiN perception listened to speech recordings mixed with a cocktail party noise at different intensities in an ecological connected speech-listening paradigm. Based on magnetoencephalographic (MEG) recordings, we quantified the cortical tracking of the different elements of the auditory scene: (1) attended stream (i.e., reader's voice only), (2) unattended multitalker background only, and (3) global scene (i.e., the combination of the attended stream and the unattended multitalker background).

Materials and Methods

The methods used for MEG data acquisition, preprocessing, and analyses are derived from previous studies (Bourguignon et al., 2013; Vander Ghinst et al., 2016).

Participants

Twenty native French-speaking healthy children (mean age, 8 years; age range, 6–9 years; 8 females and 12 males) and 20 native French-speaking healthy adults (mean age, 30 years; age range, 21–40 years; 10 females and 10 males) without any history of neuropsychiatric or otologic disorders participated in this study. All subjects had normal hearing according to pure-tone audiometry [i.e., normal hearing thresholds (between 0 and 20 dB HL) for 250, 500, 1000, 2000, 4000, and 8000 Hz] and normal otomicroscopy. Subjects' auditory perception was assessed with three separate subtests of a validated and standardized French language central auditory battery: (1) a dichotic test, (2) a speech audiometry, and (3) a SiN audiometry (Demanez et al., 2003). In the two later tests, 30 monosyllabic words were presented with (3) or without (2) noise in a predetermined counterbalanced order, so that every word is presented once in silence and once in noise. A score was then obtained, corresponding to the number of words correctly repeated with and without noise. According to the tests (1–3), all subjects had normal dichotic perception, speech, and SiN perception for their age (Demanez et al., 2003). Children and adults were all right handed according to the Edinburgh handedness inventory (Oldfield, 1971). The study had prior approval by the ULB-Hôpital Erasme Ethics Committee. Participants gave written informed consent before participation.

Experimental paradigm

During MEG recordings, the subjects sat comfortably in the MEG chair with the arms resting on a table positioned in front of them. They underwent four listening conditions and one rest condition, each lasting 5 min. The order of the five conditions was randomized for each subject.

Subjects were told before the task that questions on the content of the story would be asked after each listening condition. Children were given a clue about the content of the text they were about to listen to so as to ensure that they selected the attended auditory stream straightaway (e.g., "you are going to listen to the story of two little princes"). During the listening conditions, subjects listened to four different stories in French recorded by different native adult French speakers. The recordings were randomly selected from a set of four stories (readers' sex ratio: 1:1) obtained from a French audiobook database (http://www.litteratureaudio.com) after written authorization from the readers. Children and adults listened to different stories adapted to their age to maximize their implication in the task and their comprehension of the stories. This approach was particularly important for stories used in children as it has been previously demonstrated that reading stories aloud from books exposes children to a linguistic and cognitive complexity typically not found in child-directed or adult-directed speech (Massaro, 2017). Special care was therefore taken to select stories with comprehensible vocabulary and content. Phrasal, word, and syllable rates, assessed as the number of phrases, words, or syllables divided by the corrected duration of the audio recording, were comparable in children (mean phrasal, word, and syllable rates across different stories 0.45, 3.6, and 5.54 Hz) and in adults (0.49, 3.39, and 5.56 Hz, respectively). For phrases, the corrected duration was (trivially) the total duration of the audio recording. For words and syllables, the corrected duration was the total time during which the speaker was actually speaking, that is the total duration of the audio recording (here 5 min) minus the sum of all silent periods when the speech amplitude was below a tenth of the mean amplitude for at least 10 ms. A specific speech signal-to-noise ratio (SNR; where signal was the attended reader's voice, and noise was the multitalker background) was randomly assigned to each story: a noiseless condition, and three SiN conditions (with SNRs of +5, 0, and −5 dB), leading to four different SNR conditions. This randomization procedure prevented any systematic association between stories and SNRs. The noise (Fonds Sonores version 1.0; Perrin and Grimault, 2005) was a continuous cocktail party noise obtained by mixing the voices of six native French speakers talking simultaneously in French (three females and three males). This configu-

the Marie Skłodowska-Curie Action of the European Commission (Grant 743562). V.J. was supported by a research grant from the Institut d'Encouragement de la Recherche Scientifique et de l'Innovation de Bruxelles ("Brains back to Brussels," Brussels, Belgium). X.D.T. is Postdoctorate Clinical Master Specialist at the Fonds de la Recherche Scientifique (FRS-FNRS, Brussels, Belgium). We thank Nicolas Grimault and Fabien Perrin (Centre de Recherche en Neuroscience, Lyon, France) for providing all of the audio material. We also thank Brice Marty for providing help with data acquisition.

*M.V.G. and M.B. contributed equally to this work.

The authors declare no competing financial interests.

Correspondence should be addressed to Marc Vander Ghinst at Marc.Vander.Ghinst@erasme.ulb.ac.be.

https://doi.org/10.1523/JNEUROSCI.1732-18.2019

ration of cocktail party noise was selected because it accounts for both energetic and informational masking at phonetic and lexical levels (Simpson and Cooke, 2005; Hoen et al., 2007). Sound recordings were played using a VLC Media Player (VideoLAN Project, GNU General Public License) running on a MacBook Pro computer (Apple). Sound signals were transmitted to a MEG-compatible 60 × 60 cm ^{2} high-quality flat panel loudspeaker (Panphonics SSH sound shower, Panphonics Oy) placed 2.4 m away, in front of the subjects. The average sound intensity was set to 60 dB, as assessed by a sound level meter (Sphynx Audio System). Subjects were asked to attend to the reader's voice and to gaze at a fixation point on the wall of the magnetically shielded room facing them. During the Rest condition, subjects were instructed to relax, not to move, and to gaze at the same fixation point. At the end of each listening condition, subjects were asked to score the intelligibility of the attended reader's voice on a visual analog scale (VAS) ranging from 0 to 10 (0 = totally unintelligible; 10 = perfectly intelligible) and were also asked 16 (adults) or 8 (children) yes/no forced-choice questions exploring the salience and explicitness of the heard story by analogy with what is required for the clinical diagnosis of text comprehension deficits (Ferstl et al., 2005).

Data acquisition

Cortical neuromagnetic signals were recorded at CUB Hôpital Erasme using a whole scalp-covering, 306-channel MEG device (for 15 adults and 12 children; Elekta Neuromag Vectorview, Elekta Oy; and otherwise Elekta Neuromag Triux, MEGIN) installed in a lightweight magnetically shielded room (Maxshield, MEGIN), the characteristics of which have been described previously (De Tiège et al., 2008; Carrette et al., 2011). The MEG device has 102 sensor chipsets, each comprising one magnetometer and two orthogonal planar gradiometers. MEG signals were bandpass filtered through 0.1–330 Hz and sampled at 1 kHz. Four head-tracking coils monitored subjects' head position inside the MEG helmet. The locations of the coils and at least 150 head surface (on scalp, nose, and face) points with respect to anatomical fiducials were digitized with an electromagnetic tracker (Fastrack, Polhemus). Electro-oculogram (EOG), electrocardiogram (EKG), and the audio signals presented to the subjects were recorded simultaneously with MEG signals (passband, 0.1–330 Hz for EOG and EKG; low-pass at 330 Hz for audio signals). The recorded audio signals were used for synchronization between MEG and the transmitted audio signals, the latter being bandpass filtered at 50–22,000 Hz and sampled at 44.1 kHz. High-resolution 3D-T1 cerebral magnetic resonance (MR) images were acquired on a 1.5 T MRI (Intera, Philips).

Data preprocessing

Continuous MEG data were first preprocessed off-line using the temporal extension of the signal-space-separation method (correlation limit, 0.9; segment length, 20 s) to suppress external inferences and correct for head movements (Taulu et al., 2005; Taulu and Simola, 2006). For the subsequent coherence (Coh) analyses used to quantify the cortical tracking of speech, continuous MEG and audio signals were split into 2048 ms epochs with 1638 ms epoch overlap, leading to a frequency resolution of ~0.5 Hz (Bortel and Sovka, 2007). MEG epochs exceeding 3 pT (magnetometers) or 0.7 pT/cm (gradiometers) were excluded from further analysis to avoid contamination of the data by eye movement artifacts, muscle activity, or artifacts in the MEG sensors. The mean ± SD number of artifact-free epochs was 695 ± 66 (across subjects and conditions) in the adult group and 621 ± 93 in the children's group.

A two-way repeated-measures ANOVA (factors, age group and condition; dependent variable, number of epochs used to compute coherence) revealed a significant effect of group on the number of epochs ($F_{(1,114)} = 4.42, p = 0.042$) but no effect of SNR ($F_{(3,114)} = 0.03, p = 0.99$) or interaction ($F_{(3,114)} = 1.46$, $p = 0.23$). To avoid a possible methodological bias in our results due to differences between age groups in the accuracy of speech-tracking estimation, we threw away epochs in adults' data so as to equalize the number of epochs in both groups.

Coherence analyses in sensor space

For each listening condition, synchronization between the temporal envelope of wide-band (50–22,000 Hz) audio signals and artifact-free MEG epochs (2048 ms long) was assessed with coherence analysis in sensor space at frequencies in which speech temporal envelope is critical for speech comprehension (i.e., 0.1–20 Hz; Drullman et al., 1994). Coherence is an extension of the Pearson correlation coefficient to the frequency domain. It quantifies the degree of coupling between two signals [say $x(t)$ and $y(t)$], providing a number between 0 (no linear dependency) and 1 (perfect linear dependency) for each frequency bin (Halliday et al., 1995). Coherence was computed as follows:

$$Coh_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f)\,P_{yy}(f)},$$

where $P_{xx}(f) = \Sigma_k|\hat{x}_k(f)^2|$, $P_{yy}(f) = \Sigma_k|\hat{y}_k(f)^2|$, and $P_{xy}(f) = \Sigma_k|\hat{x}_k(f)\hat{y}^*_k(f)|$, and where $\hat{x}_k(f)$ [respectively $\hat{y}_k(f)$] is the Fourier coefficient of the k^{th} epoch of signal $x(t)$ [respectively $y(t)$] at frequency bin f, and * denotes the complex conjugate. In our application, coherence analysis provides a quantitative assessment of the cortical tracking of speech.

For the three SiN conditions (+5, 0, and −5 dB), coherence was separately computed between MEG signals and three acoustic elements of the auditory scene: (1) the global scene (attended stream + multitalker background), leading to $\mathrm{Coh}_{\mathrm{global}}$; (2) the attended (att) stream only (i.e., the reader's voice), leading to $\mathrm{Coh}_{\mathrm{att}}$; and (3) the multitalker background (bckgr) only, leading to $\mathrm{Coh}_{\mathrm{bckgr}}$. Sensor-level coherence maps were obtained using gradiometer signals only, and signals from gradiometer pairs were combined as was done in the study by Bourguignon et al. (2015).

Previous studies have demonstrated statistically significant coupling between acoustic and brain signals at frequencies corresponding to phrases, words, and syllables (Ding and Simon, 2012b; Bourguignon et al., 2013; Peelle et al., 2013; Clumeck et al., 2014; Koskinen and Seppä, 2014; Vander Ghinst et al., 2016, Keitel et al., 2018). Accordingly, sensor-level coherence maps were produced separately for phrases (frequency bin corresponding to 0.5 Hz), words (average across the frequency bins falling in 1–4 Hz), and syllables (4–8 Hz). Note that coherence at the frequency bin corresponding to 0.5 Hz actually reflects coupling in a frequency range of ~0.5 Hz, with a sensitivity profile proportional to the Fourier transform of a boxcar function: $\mathrm{sinc}(\pi(f - 0.5\text{ Hz})/0.5\text{ Hz})$. The <1, 1–4, and 4–8 Hz frequency ranges are henceforth referred to as frequency bands of interest.

Coherence analyses in source space

Individual MR images were first segmented using Freesurfer software (Martinos Center for Biomedical Imaging, Boston, MA; RRID: SCR_001847; Reuter et al., 2012). MEG and segmented MRI coordinate systems were then coregistered using the three anatomical fiducial points for initial estimation and the head-surface points to manually refine the surface coregistration. The MEG forward model was computed for triplets of orthogonal current dipoles, placed on a homogeneous 5 mm grid source space covering the whole brain, using MNE suite (Martinos Centre for Biomedical Imaging; RRID:SCR_005972; Gramfort et al., 2014). The forward model was then reduced to its two first principal components. This procedure is justified by the insensitivity of MEG to currents radial to the skull, and hence, this dimension reduction leads to considering only the tangential sources. As a preliminary step, to simultaneously combine data from planar gradiometers and magnetometers for source estimation, sensor signals (and the corresponding forward-model coefficients) were normalized by their noise root mean square, estimated from the Rest data filtered through 1–195 Hz. Coherence maps obtained for each subject, listening condition (noiseless, +5, 0, and −5 dB), audio signal (global scene, attended stream, multitalker background) and frequency bands of interest (<1, 1–4, and 4–8 Hz) were finally produced using the dynamic imaging of coherent sources approach (Gross et al., 2001) with minimum-norm estimates inverse solution (Dale and Sereno, 1993). Noise covariance was estimated from the rest data filtered through 1–195 Hz, and the regularization parameter was fixed in terms of MEG sensor noise level, as done by Hämäläinen et al. (2010).

Group-level analyses in source space

A nonlinear transformation from individual MR images to the standard Montreal Neurological Institute (MNI) brain was first computed using the spatial normalization algorithm implemented in Statistical Parametric Mapping (SPM8, Wellcome Department of Cognitive Neurology, London, UK; RRID:SCR_007037; Ashburner et al., 1997; Ashburner and Friston, 1999) and then applied to individual MR images and every coherence map. The adult MNI template was used in both children and adults despite the fact that spatial normalization may fail for brains of small size when using an adult template (Reiss et al., 1996). However, this risk is negligible for the population studied here. Indeed, the brain volume does not change substantially from the age of 5 years to adulthood (Reiss et al., 1996). This assumption has been confirmed by a study that specifically addressed this question in children aged >6 years (Muzik et al., 2000).

To produce coherence maps at the group level, we computed across subjects the generalized f mean of normalized maps, according to $f(\cdot) = \text{arctanh}(\sqrt{\cdot})$, namely, the Fisher z-transform of the square root. This procedure transforms the noise on the coherence estimate into an approximately normally distributed noise (Rosenberg et al., 1989). Thus, the computed coherence is an unbiased estimate of the mean coherence at the group level. In addition, this averaging procedure avoids an overcontribution of subjects characterized by high coherence values to the group analysis (Bourguignon et al., 2012). The resulting subject- and group-level coherence maps are henceforth referred to as the audio maps.

Experimental design and statistical analyses

Sample size was based on a previous study from our group with a similar design, which included 20 healthy adults (Vander Ghinst et al., 2016). Accordingly, we set the sample size to 20 per age group.

Comparison of SiN perception in adults versus children. Children's and adults' capacities to understand speech and SiN—as measured with speech and SiN audiometry—were compared with a t test.

Effect of SNR on the comprehension and the intelligibility of the attended stream. A two-way repeated-measures ANOVA was used to assess the effects of the multitalker background noise level (within-subject factor; noiseless, +5, 0, and −5 dB) and of the age group (between-subjects factor; adults, children) on the comprehension scores and intelligibility ratings separately. The distribution of the residues of the ANOVAs was then tested for normality using the Lilliefors (1967) test.

Of note, we acknowledge that the interpretability of these analyses could be limited for two reasons. First, adults and children listened to different texts and had to answer questions where the difficulty was adapted to their age. Second, the intelligibility ratings by children and adults may also differ due to differences in the VASs: explicit visual support was provided for the children (more or less happy faces) to facilitate the evaluation.

Significance of individual subjects' coherence values. The statistical significance of individual subjects' coherence values (for each listening condition, audio signal, and frequency band of interest) was assessed with surrogate data-based maximum statistics. This statistical assessment was performed on sensor–space coherence values, and it tested the null hypothesis that the brain does not track audio signals more than other plausible unrelated (surrogate) signals. This method was chosen because it intrinsically deals with the issue of multiple comparisons across sensors, and because it takes into account the temporal autocorrelation within signals. For each subject, 1000 surrogate sensor-level coherence maps were computed as was done for genuine coherence maps but with audio signals replaced by Fourier transform surrogate audio signals (Faes et al., 2004). The maximum coherence value across all sensors was extracted for each surrogate simulation, and the 95th percentile of this distribution of maximum coherence values yielded the significance threshold at $p < 0.05$.

Significance of group-level coherence values. The statistical significance of coherence values in group-level audio maps was assessed for each hemisphere separately with a nonparametric permutation test (Nichols and Holmes, 2002). In practice, subject- and group-level rest coherence maps were computed as done for the audio maps, but with MEG signals in listening conditions replaced by rest MEG signals and sound signals unchanged. Group-level difference maps were obtained by subtracting f-transformed audio and rest group-level coherence maps. Under the null hypothesis that coherence maps are the same whatever the experimental condition, the labeling audio and rest are exchangeable at the subject-level before group-level difference map computation (Nichols and Holmes, 2002). To reject this hypothesis and to compute a threshold of statistical significance for the correctly labeled difference map for each hemisphere separately, the permutation distribution of the maximum absolute value of the difference map in each hemisphere was computed for 10,000 permutations. The test assigned a p value to each voxel in the group-level audio map, equal to the proportion of surrogate values exceeding the difference value of the corresponding voxel (Nichols and Holmes, 2002).

We further identified the coordinates of local maxima in group-level coherence maps. Such local coherence maxima are sets of contiguous voxels displaying higher coherence values than all neighboring voxels. We only report statistically significant local coherence maxima, disregarding the extents of these clusters. Indeed, cluster extent is hardly interpretable in view of the inherent smoothness of MEG source reconstruction (Hämäläinen and Ilmoniemi, 1994; Wens et al., 2015; Bourguignon et al., 2018).

Cortical processing of the auditory scene in SiN conditions. To identify cortical areas wherein activity reflects more the attended stream than the global scene, we compared Coh_{att} to $\text{Coh}_{\text{global}}$ maps using the same nonparametric permutation test described above, but with the labels global scene and attended stream instead of audio and rest, leading to the $\text{Coh}_{\text{att}} - \text{Coh}_{\text{global}}$ difference maps.

Comparison of the tracking in adults versus children. To identify cortical areas showing stronger tracking in adults than children (and vice versa), we compared the corresponding coherence maps using the above described permutation test, but with the labels $\text{Coh}_{\text{att,adults}}$ and $\text{Coh}_{\text{att,children}}$ instead of Coh_{att} and $\text{Coh}_{\text{global}}$, leading to the $\text{Coh}_{\text{att,adults}} - \text{Coh}_{\text{att,children}}$ difference maps.

Effect of the SNR, age group, and hemispheric lateralization on cortical tracking of speech in noise. In this between-subject design, we used a three-way repeated-measures ANOVA to compare cortical tracking of speech between $n = 20$ children and $n = 20$ adults with additional factors of hemisphere (left vs right) and four different SNR conditions (noiseless, +5, 0, and −5 dB). The dependent variable was the maximal Coh_{att} value within a sphere of 10 mm radius around the maximum of the group-level difference map in each hemisphere. As the Lilliefors test revealed that the distribution of the residues of the ANOVAs deviated statistically significantly from the normal distribution at 1–4 and 4–8 Hz Coh_{att} values (p values < 0.05), we repeated the ANOVAs on the coherence values transformed with the transformation used to average source coherence maps ($f(\cdot) = \text{arctanh}(\sqrt{\cdot})$). After such transformation, the residues did not deviate significantly from a normal distribution (p values >0.05). Therefore, we report only on the results of the later ANOVAs. *Post hoc* comparisons were performed with pairwise t tests.

Based on our results in the noiseless condition, we conducted an additional analysis by computing Pearson correlation between children's ages and their maximum coherence values, separately for both hemispheres in the three frequency bands of interest.

Results

In this study, children and adults were listening to connected speech embedded in a multitalker background with different SNR conditions. Ensuing cortical tracking of speech (i.e., the coupling between brain activity and audio signals) was quantified with coherence analysis. The specific aim was to compare this tracking between children and adults.

SiN perception in adults versus children

Speech perception in silence did not differ ($t_{(38)} = 1.27$; $p = 0.211$) between children (28.35 ± 0.88; mean ± SD) and adults (28.7 ± 0.86). SiN perception quantified with SiN audiometry was significantly ($t_{(38)} = 3.35$; $p = 0.0018$) poorer in children (25.75 ± 1.33) than in adults (27.1 ± 1.21).

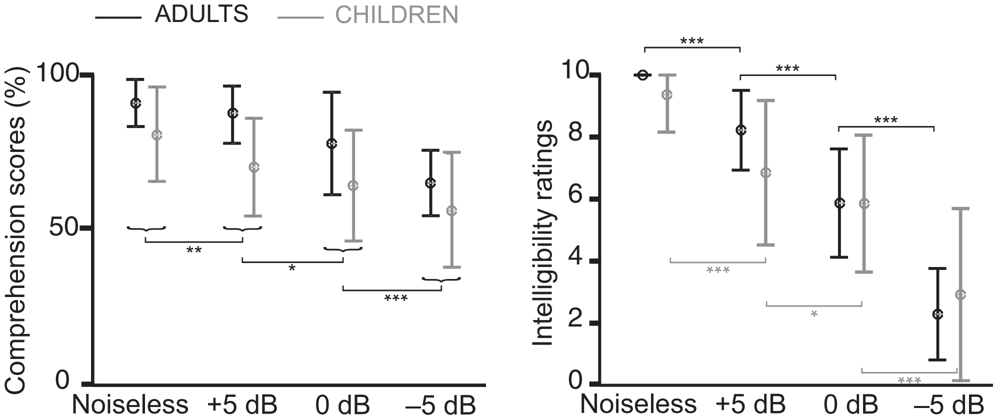

Figure 1. Comprehension scores (left graph) and intelligibility ratings (right graph) in adults (black) and children (gray). Bars indicate the mean and range. Comprehension scores are reported as the percentage of questions (16 for adults and 8 for children) answered correctly, and intelligibility ratings ranged from 0 (totally unintelligible) to 10 (perfectly intelligible). The comprehension and intelligibility of the attended stream decreased significantly as SNR decreased. Horizontal brackets indicate the outcome of *post hoc* paired *t* tests between adjacent conditions (***$p < 0.001$, **$p < 0.01$, *$p < 0.05$).

Effect of SNR on the comprehension and the intelligibility of the attended stream

Figure 1 displays the comprehension scores and intelligibility ratings in the different SNR conditions in both groups.

In the noiseless condition, all adult participants gave the maximum intelligibility rating (10), leading to a null variance. For this reason, the ANOVA for the intelligibility ratings was computed only with the three other conditions (+5, 0, and −5 dB). Doing otherwise would have violated the homoscedasticity assumption of the ANOVA.

The ANOVA performed on the intelligibility ratings revealed a statistically significant effect of SNR ($F_{2,76} = 119$; $p < 0.0001$) and a significant interaction between SNR and age group ($F_{2,76} = 4.94$; $p = 0.0096$), but no significant effect of age group ($F_{1,38} = 0.05$; $p = 0.83$). The Lilliefors test showed that the distribution of the residuals did not deviate significantly from a normal distribution ($p = 0.15$). *Post hoc* comparisons performed with pairwise *t* tests between adjacent conditions demonstrated that intelligibility ratings decreased statistically significantly from noiseless to +5 dB (adults, $t_{19} = 6.28$, $p < 0.0001$; children, $t_{19} = 5.06$, $p < 0.0001$), from +5 to 0 dB (adults, $t_{19} = 7.19$, $p < 0.0001$; children, $t_{19} = 2.11$, $p = 0.048$) and from 0 to −5 dB (adults, $t_{19} = 9.48$, $p < 0.0001$; children, $t_{19} = 5.09$, $p < 0.0001$). Comparison between adults and children revealed that children gave lower intelligibility ratings than adults in the noiseless ($t_{38} = 2.26$, $p = 0.030$) and +5 dB conditions ($t_{(38)} = 2.10$, $p = 0.042$) but not in the two other noisiest conditions (p values > 0.05).

The ANOVA performed on comprehension scores—converted to percentage correct—revealed a significant effect of SNR ($F_{3,114} = 27.6$; $p < 0.0001$), a significant effect of age group ($F_{1,38} = 19.4$; $p < 0.0001$), and no significant interaction ($F_{3,114} = 0.66$; $p = 0.58$). The Lilliefors test showed that the distribution of the residuals did not deviate significantly from a normal distribution ($p = 0.054$). Comprehension scores were higher in adults (80.5 ± 15.2%; mean ± SD across conditions and participants) than in

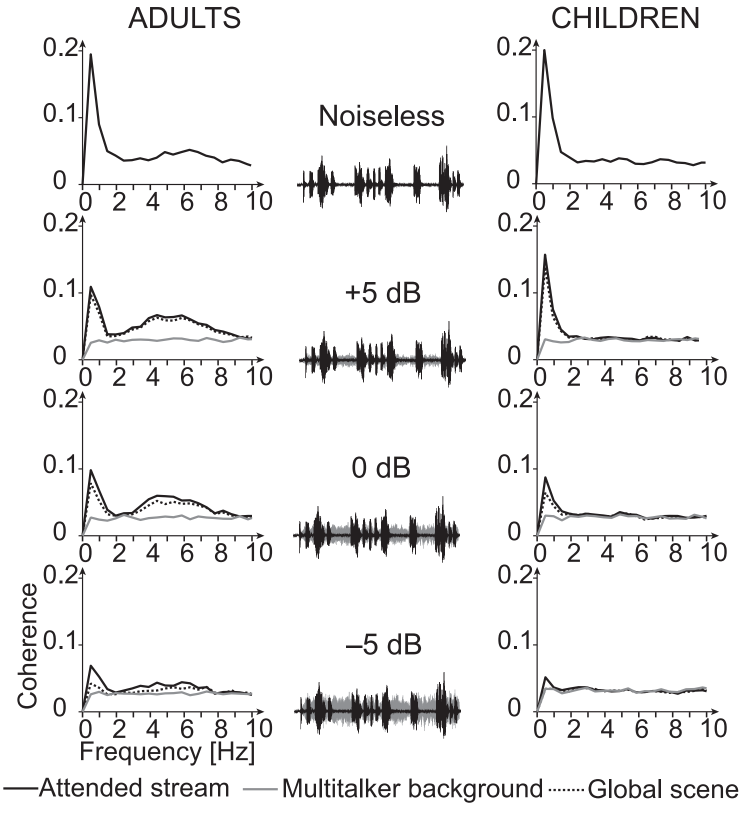

Figure 2. Spectra of cortical tracking of speech in the four speech-in-noise conditions and corresponding sound excerpts. Group-averaged coherence spectra are shown separately for adults (left column) and children (right column), and when estimated with the temporal envelope of the different components of the auditory scene: the attended stream (black connected trace), the multitalker background (gray connected trace), and the global scene (gray dotted trace). Each spectrum represents the mean across subjects (20 in each age group) of the maximum coherence across all sensors. The sound excerpts showcase the attended stream (black traces) and the multitalker background (gray traces) and their relative amplitude depending on the signal-to-noise ratio.

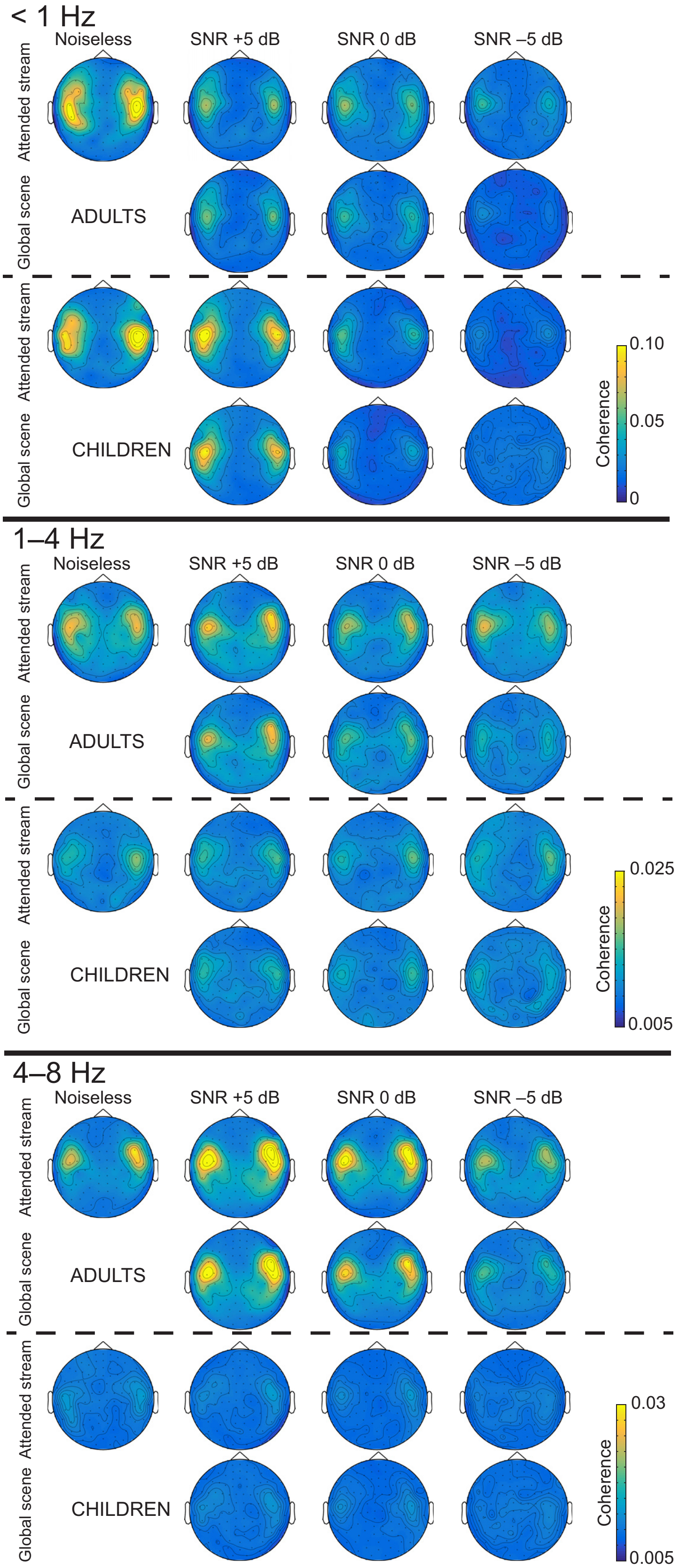

Figure 3. Cortical tracking of speech in the sensor space. Plots display the group-averaged coherence distributions obtained based on gradiometer data, implying that maximum coherence should peak right above generating brain sources. There is one distribution map for each frequency band of interest (<1, 1–4, and 4–8 Hz), age group (adults, children) component of the auditory scene (attended stream, global scene), and SNR condition (noiseless, and +5, 0, and −5 dB).

children (68.6 ± 18.7%), and decreased statistically significantly from noiseless to +5 dB (t_{39} = 3.13, p = 0.0033), from +5 to 0 dB ($t_{(39)}$ = 2.13, p = 0.040), and from 0 to −5 dB (t_{39} = 4.00, p = 0.0003).

Cortical tracking of speech in the noiseless condition

Figures 2, 3, and 4 display group-averaged coherence spectra (Fig. 2), sensor distribution (Fig. 3), and source distribution (Fig. 4) in all conditions and in both groups.

Table 1 provides the number of children and adults showing significant sensor space $\mathrm{Coh}_{\mathrm{global}}$, $\mathrm{Coh}_{\mathrm{att}}$, and $\mathrm{Coh}_{\mathrm{bckgr}}$ at <1, 1−4, and 4−8 Hz frequencies.

In adults, statistically significant $\mathrm{Coh}_{\mathrm{att}}$ was observed at <1 Hz in 20 of 20 adults, at 1−4 Hz in 17 of 20 adults, and at 4−8 Hz in 17 of 20 adults in MEG sensors covering the temporal lobe in the noiseless condition (Figs. 2, 3). In source space, group-level coherence at <1 Hz peaked at bilateral superior temporal sulcus (STS; left hemisphere MNI coordinates, [−66, −16, −1], p < 0.0001; right hemisphere MNI coordinates, [66, −26, 5], p < 0.0001); left inferior frontal gyrus (IFG), [−61, −6, 34], p = 0.0037); and right central sulcus, [59, −5, 41], p = 0.034). In addition, $\mathrm{Coh}_{\mathrm{att}}$ peaked at bilateral supratemporal auditory cortices (ACs) at 1–4 Hz (left AC, [−65, −16, 7], p < 0.0001; right AC, [64, −5, 6], p < 0.0001) and 4−8 Hz (left AC, [−65, −15, 9], p < 0.0001; right AC, [64, −8, 5], p < 0.0001; Fig. 4).

In children, statistically significant $\mathrm{Coh}_{\mathrm{att}}$ was observed at <1 Hz in 20 of 20 children, at 1–4 Hz in 10 of 20 children, and at 4−8 Hz in 11 of 20 children in temporal lobe MEG sensors in the noiseless condition (Figs. 2, 3). In source space, group-level coherence at <1 Hz peaked at bilateral STS (left STS, [−63, −10, 6], p < 0.0001; right STS, [63, −20, −3], p < 0.0001). It peaked at bilateral supratemporal AC at 1–4 Hz (left AC, [−64, −11, 13], p = 0.0016; right AC, [62, −28, 4], p < 0.0001), and at left supratemporal AC at 4−8 Hz ([−65, −17, 10], p = 0.018). Of note, coherence did peak in the right supratemporal AC at 4−8 Hz, but this local maximum did not reach statistical significance ([64, −14, 5], p = 0.11).

The above results suggest that speech tracking differed somewhat in adults and children at 1–4 and 4−8 Hz, but not at <1 Hz. Indeed, a smaller proportion of children than of adults showed significant tracking at 1–4 Hz (10 of 20 vs 17 of 20; p = 0.041, Fisher exact test). A similar but statistically nonsignificant trend was observed at 4–8 Hz (11 of 20 vs 17 of 20; $p = 0.082$). Also, contrast between adults and children did reveal stronger tracking in adults than in children in bilateral supratemporal AC at 1–4 Hz (left AC, [−65, −16, 7], $p = 0.0013$; right [64, −5, 6], $p = 0.0041$) and 4–8 Hz (left AC, [−65, −15, 9], $p < 0.0001$; right AC, [64, −8, 5], $p < 0.0001$), but not at <1 Hz.

Correlation between Coh_{att} values and children's age was statistically significant in the right STS at <1 Hz ($r = 0.47$, $p = 0.039$), and in the right supratemporal AC at 4–8 Hz ($r = 0.50$, $p = 0.025$; Fig. 5). No other correlations between Coh_{att} values and children's age reached statistical significance (p values > 0.3).

Cortical tracking of speech in SiN conditions

In adults, group-level Coh_{att} and $\text{Coh}_{\text{global}}$ maps at <1 Hz displayed statistically significant (p values < 0.05) local maxima at bilateral STS at every SNR, in the left IFG at 0 dB, and in the left temporal pole at −5 dB. $\text{Coh}_{\text{bckgr}}$ was not statistically significant in any condition. At 1–4 and 4–8 Hz, Coh_{att} and $\text{Coh}_{\text{global}}$ maps displayed statistically significant (p values < 0.05) local maxima in AC bilaterally in every condition (Fig. 4). $\text{Coh}_{\text{bckgr}}$ was statistically significant at 0 and −5 dB in right AC.

In children, group-level Coh_{att} and $\text{Coh}_{\text{global}}$ maps at <1 Hz displayed statistically significant (p values < 0.05) local maxima at bilateral STS at every SNR except at −5 dB where only Coh_{att} displayed significant local maxima (Fig. 4). At 1–4 Hz, Coh_{att} and $\text{Coh}_{\text{global}}$ maps displayed statistically significant (p values < 0.05) local maxima in AC bilaterally at +5 and 0 dB, and only Coh_{att} displayed significant coherence in right AC at −5 dB (Fig. 4). At 4–8 Hz, Coh_{att} and $\text{Coh}_{\text{global}}$ maps displayed statistically significant (p values < 0.05) local maxima only at right AC at +5 and 0 dB, but not at −5 dB (Fig. 4). $\text{Coh}_{\text{bckgr}}$ was not statistically significant in any condition and frequency band of interest.

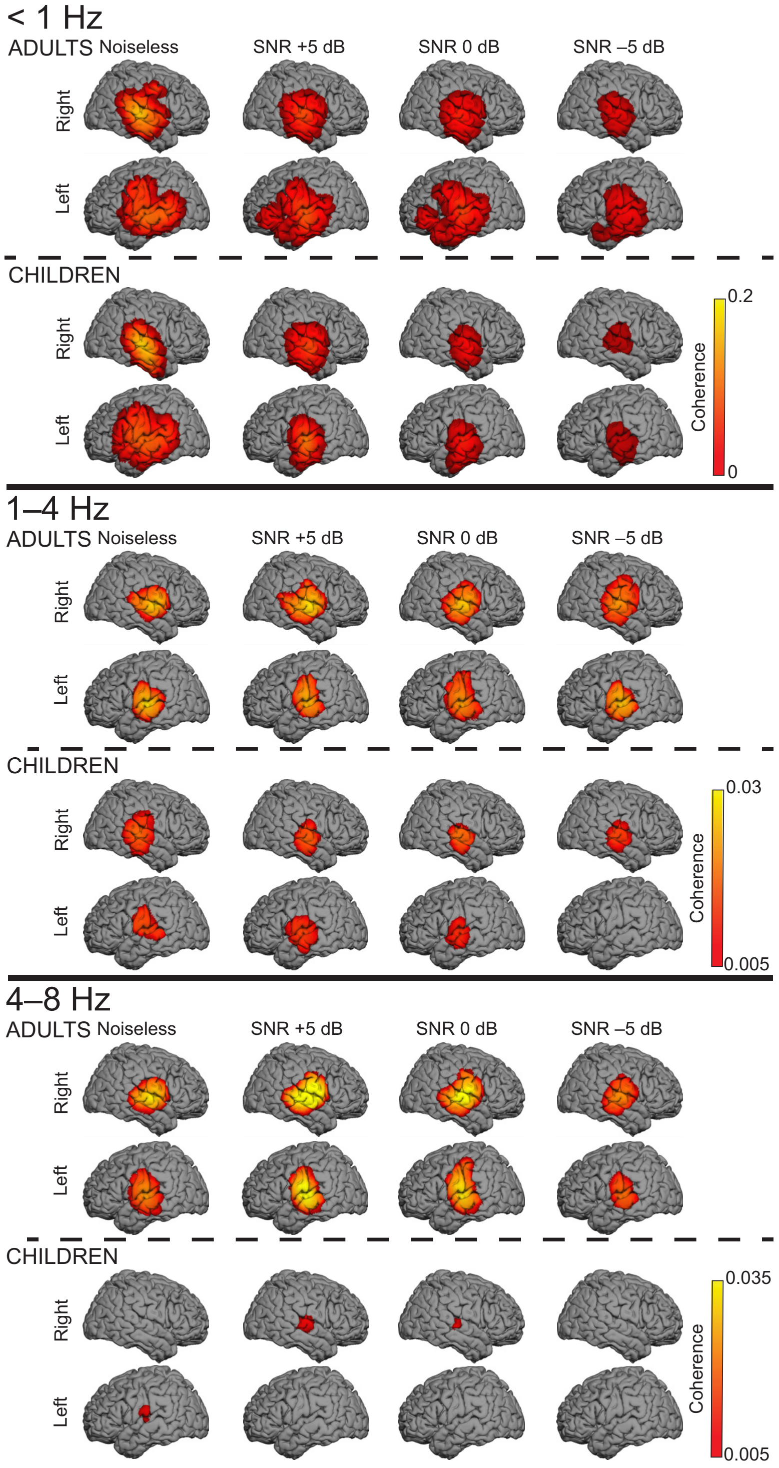

Figure 4. Cortical tracking of the attended speech stream at <1, 1–4, and 4–8 Hz. The group-level coherence maps ($n = 20$ in each map) were masked statistically above the significance level (nonparametric permutation statistics). One source distribution is displayed for each possible combination of age group (adults, top; children, bottom) and SNR condition (from left to right: noiseless, and +5, 0, and −5 dB).

Cortical processing of the auditory scene in SiN conditions

In adults, Coh_{att} was stronger than $\text{Coh}_{\text{global}}$ (i.e., MEG signals tracked more the attended stream than the global scene. At <1 Hz, cortical areas showing this effect included bilateral STS at every SNR (+5, 0, and −5 dB; $p < 0.0001$) and left inferior frontal gyrus at 0 dB. Significantly stronger Coh_{att} than $\text{Coh}_{\text{global}}$ was found at bilateral AC at every SNR at 1–4 Hz (+5 dB, $p = 0.0004$; 0 dB, $p = 0.0003$; −5 dB, $p = 0.0001$) and 4–8 Hz (+5 dB, $p < 0.0001$; 0 dB, $p < 0.0001$; −5 dB, $p = 0.0005$).

In children, Coh_{att} at <1 Hz was statistically significantly stronger than $\text{Coh}_{\text{global}}$ in bilateral STS at every SNR (+5, 0, and −5 dB: p values < 0.0001) and in central opercular cortex at −5 dB. At 1–4 Hz, significantly stronger Coh_{att} than $\text{Coh}_{\text{global}}$ was found in right STS (+5 dB, $p = 0.0011$; 0 dB, $p = 0.0011$; −5 dB,

Table 1. Number of adults and children showing statistically significant coherence (surrogate data-based statistics) in at least one sensor for each audio signal, condition, and frequency band of interest

	Attended stream		Multitalker background		Global scene	
Condition	Adults	Children	Adults	Children	Adults	Children
<1 Hz						
Noiseless	20	20				
+5 dB	20	19	2	1	20	19
0 dB	20	17	1	1	18	14
−5 dB	16	6	0	4	9	1
1–4 Hz						
Noiseless	17	10				
+5 dB	16	8	2	1	14	9
0 dB	16	8	0	3	13	8
−5 dB	10	7	3	1	7	6
4–8 Hz						
Noiseless	17	11				
+5 dB	20	7	3	3	20	7
0 dB	19	5	3	4	19	5
−5 dB	15	2	7	1	13	3

p = 0.0042). At 4–8 Hz, significantly stronger Coh$_{att}$ than Coh$_{global}$ was found at −5 dB and in a nonauditory area (left superior frontal gyrus; $p = 0.034$).

Tracking in adults versus children

Tracking at frequencies <1 Hz was stronger in adults than in children in left inferior frontal gyrus at 0 dB ($p = 0.030$) and in bilateral STS at −5 dB (left, $p = 0.0003$; right, $p = 0.0056$). At 1–4 Hz, Coh$_{att}$ was significantly higher in adults than in children in bilateral AC at every SNR (p values < 0.05) except at 0 dB in the right AC, where this effect was only marginally significant ($p = 0.054$). At 4–8 Hz, Coh$_{att}$ was significantly higher in adults than in children in bilateral AC at every SNR (p values < 0.0001; Fig. 6).

In contrast, no brain area displayed significantly stronger Coh$_{att}$ values in children than in adults at any frequency band of interest.

Effect of age group, SNR, and hemispheric lateralization on cortical tracking of speech

The effect of age group, SNR, and hemispheric lateralization on cortical tracking of speech was sought for with three-way repeated-measures ANOVA for the three frequency bands of interest separately. Since both adults' and children's brains track preferentially the attended stream rather than the global scene, the ANOVA was performed on Coh$_{att}$ values only.

At <1 Hz, the ANOVA revealed a significant main effect of noise level on Coh$_{att}$ ($F_{3,114} = 66.64$, $p < 0.0001$), a significant interaction between SNR condition and age group ($F_{3,114} = 4.23$, $p = 0.0071$), and a significant interaction between SNR condition and hemispheric lateralization ($F_{3,114} = 9.35$, $p < 0.0001$), but no other significant effects (p values > 0.05). In particular, there was no significant interaction between age group and hemispheric lateralization ($F_{1,38} = 0.07$, $p = 0.79$), showing that the effect of noise on hemispheric lateralization was similar in both age groups.

Figure 7 illustrates these effects identified on cortical tracking of the attended speech stream. The main effect of noise was explained by a decrease in Coh$_{att}$ as SNR increased. The interaction between SNR condition and age group was explained by a faster decrease in children's than adults' Coh$_{att}$ with decreasing SNR (Figs. 4, 7). Supporting this interpretation, adults had higher Coh$_{att}$ values than children at −5 dB ($t_{38} = 3.88$, $p = 0.0004$; t test), but not at 0, +5 dB, and noiseless (p values > 0.17). The interaction between SNR condition and hemispheric lateralization was explained by a faster decrease in Coh$_{att}$ in the right than the left STS with decreasing SNR (Figs. 4, 6). Supporting this interpretation, *post hoc* comparisons revealed that Coh$_{att}$ at right STS decreased as soon as the multitalker background was added (noiseless vs +5 dB: $t_{19} = 4.9$, $p < 0.0001$), and it further diminished as SNR decreased (+5 vs 0 dB, $t_{19} = 4.1$, $p = 0.0002$; 0 vs −5 dB, $t_{19} = 4.23$, $p = 0.0001$). In contrast, Coh$_{att}$ in the left STS decreased significantly only in the two noisiest conditions (noiseless vs +5 dB, $t_{19} = 0.49$, $p = 0.62$; +5 vs 0 dB, $t_{19} = 3.11$, $p = 0.0035$; 0 vs −5 dB, $t_{19} = 5.3$, $p < 0.0001$).

At 1–4 Hz, the ANOVA revealed a significant main effect of age group on Coh$_{att}$ ($F_{3,114} = 10.04$, $p = 0.003$), a significant main effect of hemispheric lateralization ($F_{3,114} = 10.04$, $p = 0.003$), no significant main effect of the SNR condition ($F_{3,114} = 1.53$, $p = 0.21$), and no significant interactions (p values > 0.05). The main effect of age group was due to higher Coh$_{att}$ values in adults than in children. The main effect of hemispheric lateralization was explained by higher Coh$_{att}$ values in the right AC than in the left AC.

At 4–8 Hz, the ANOVA revealed a significant main effect of age group on Coh$_{att}$ ($F_{3,114} = 54.64$, $p < 0.0001$), a significant main effect of noise level ($F_{3,114} = 13.96$, $p < 0.0001$), a significant effect of hemispheric lateralization ($F_{3,114} = 7.37$, $p = 0.0099$), a significant interaction between SNR condition and age group ($F_{3,114} = 10.08$, $p < 0.0001$), and no other significant interactions (p values > 0.05). The main effect of age group was explained by higher Coh$_{att}$ values in adults than in children at all SNRs (p values < 0.001). The interaction between SNR condition and age group was explained by the presence of SNR-related modulation in adults' Coh$_{att}$ values and the absence of such modulation in children's Coh$_{att}$ values (Fig. 7). Indeed, adults' Coh$_{att}$ values (mean across hemispheres) was significantly higher at intermediate SNRs (5 and 0 dB) than in noiseless (5 dB, $t_{19} = 4.29$, $p = 0.0004$; 0 dB, $t_{19} = 2.85$, $p = 0.010$) and at −5 dB (5 dB, $t_{19} = 5.26$, $p < 0.0001$; 0 dB, $t_{19} = 5.20$, $p < 0.0001$), significantly higher at 5 dB than at 0 dB ($t_{19} = 2.30$, $p = 0.033$), and marginally higher in noiseless than at −5 dB ($t_{19} = 1.96$, $p = 0.065$). The main effect of hemispheric lateralization was explained by higher or marginally higher Coh$_{att}$ values in the right AC than in the left AC in every SNR condition (noiseless, $t_{39} = 2.06$, $p = 0.046$; 5 dB, $t_{39} = 2.37$, $p = 0.023$; 0 dB, $t_{39} = 1.93$, $p = 0.061$; −5 dB, $t_{39} = 1.71$, $p = 0.095$).

Discussion

This study discloses commonalities between children's and adults' cortical tracking of SiN. First, both children's and adults' auditory systems similarly tracked the attended speaker's voice more than the global auditory scene at <1 and 1–4 Hz. Second, cortical tracking of the attended stream in SiN conditions was at <1 Hz similarly left hemisphere dominant in children and adults. Furthermore, in both groups, the STS was the main brain area underpinning this tracking at <1 Hz. There were also marked differences between children and adults. (1) Compared with adults, children displayed reduced cortical tracking of speech at 1–4 Hz, and particularly at 4–8 Hz (even in noiseless conditions). (2) Increasing multitalker background level compromised children more than adults in cortical tracking of speech at <1 Hz. (3) Children did not exhibit selective cortical representation of SiN at 4–8 Hz.

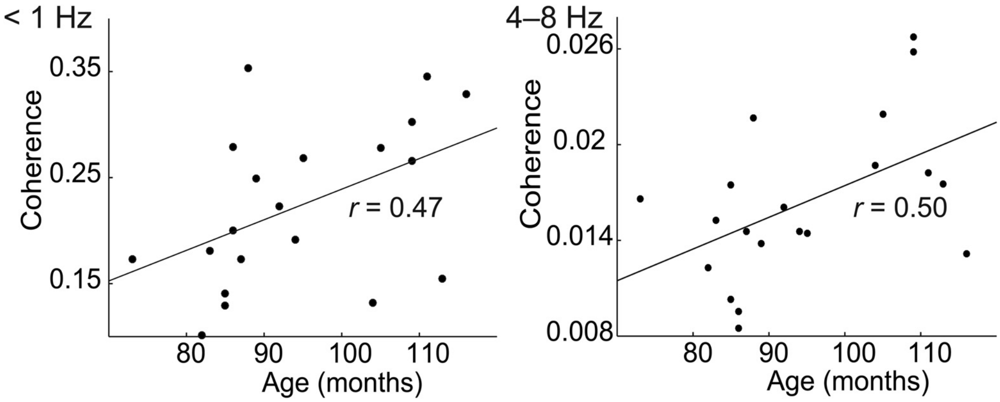

Figure 5. Correlation between children's age and their speech-tracking values, depicted by the peak group-level coherence, in the noiseless condition at the right superior temporal sulcus at <1 Hz and the supratemporal auditory cortex at 4 – 8 Hz.

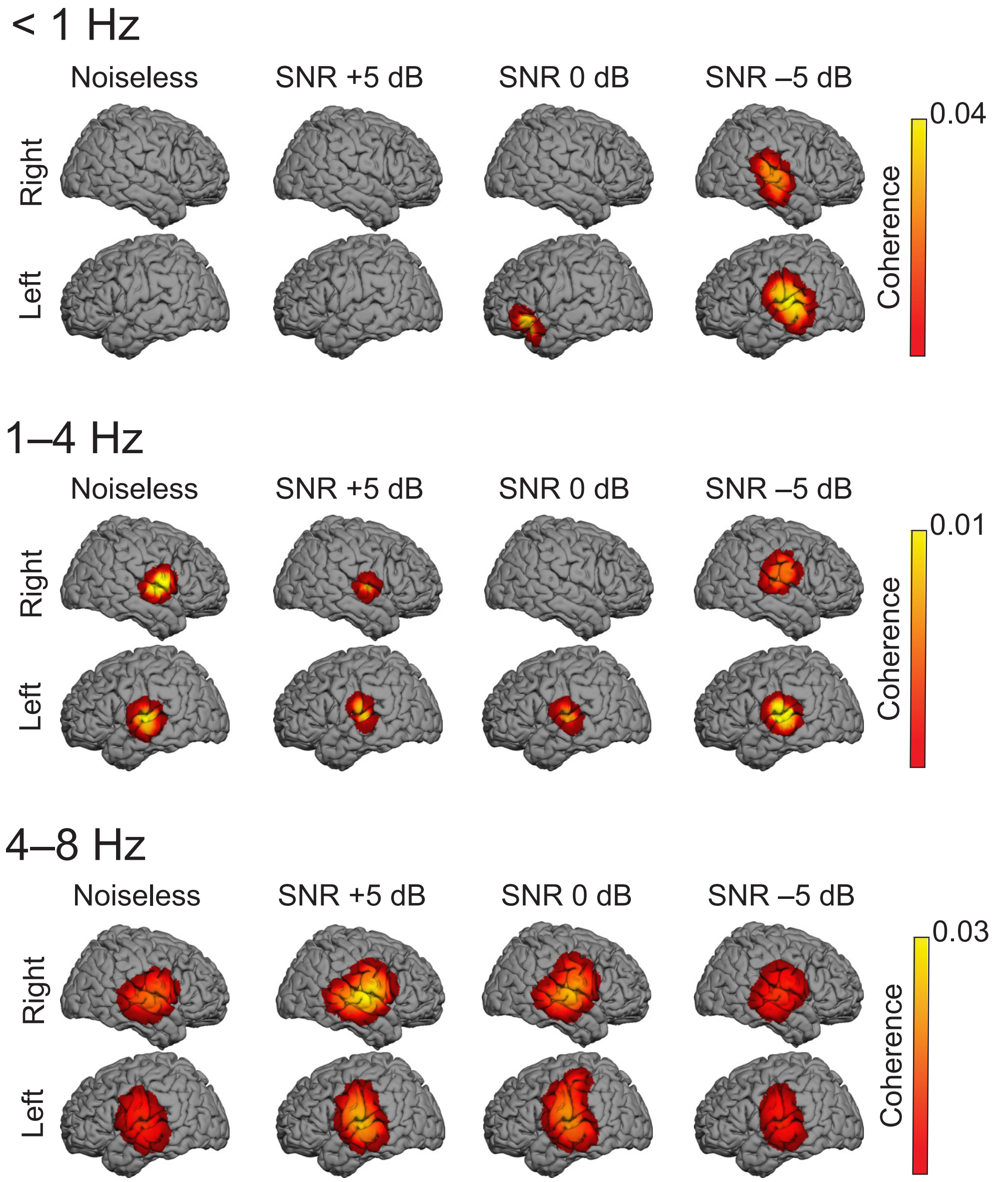

Figure 6. Contrasts between cortical tracking of the attended speech (Coh_{att}) in adults vs children ($\mathrm{Coh}_{att,adults} - \mathrm{Coh}_{att,children}$) at <1, 1– 4, and 4 – 8 Hz in all SNR conditions (noiseless, and +5, 0, and −5 dB). The group-level difference coherence maps ($n = 20$ in each map) were masked statistically above the significance level.

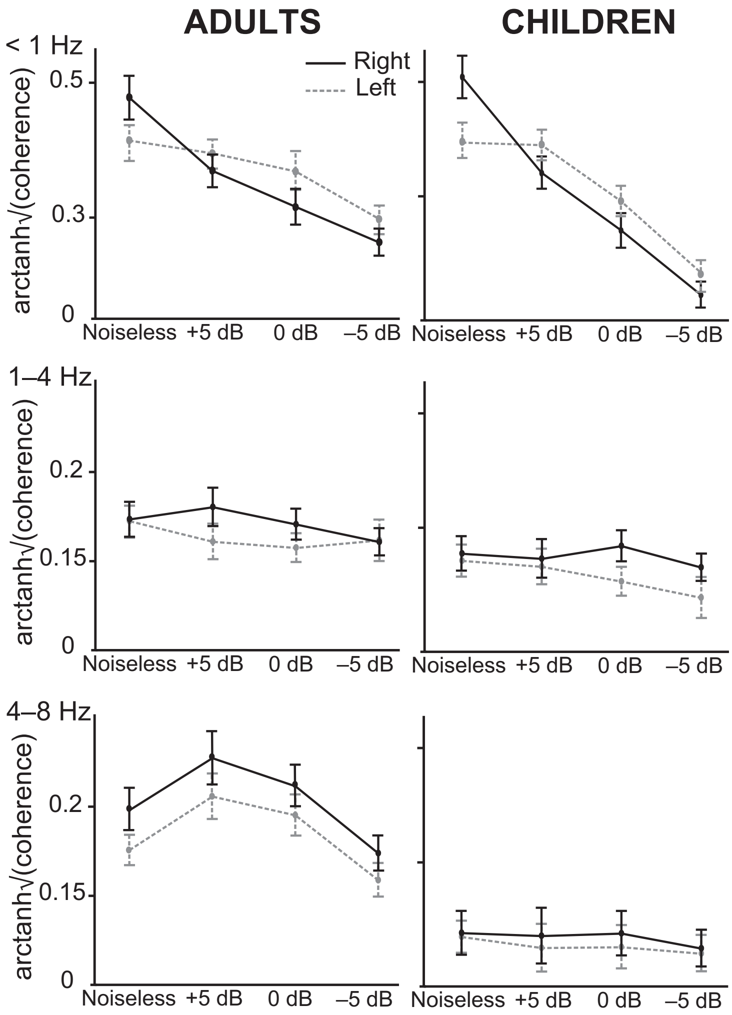

Figure 7. Illustration of the interaction effects identified on speech brain tracking of the attend speech stream in adults (left column) and children (right column), at <1, 1–4, and 4–8 Hz. Bars indicate the mean and SEM of *f*-transformed (with $f(\cdot) = \text{arctanh}(\sqrt{\cdot})$) coherence values at the right (solid lines) and left (dashed lines) cortical area of peak group-level coherence (superior temporal sulcus at <1 Hz, and supratemporal auditory cortex at 1–4 and 4–8 Hz).

Inaccurate cortical tracking of speech at 4–8 Hz in children

Studies previously reported that ongoing cortical oscillations track speech regularities, especially at the syllable timescale, which corresponds to 4–8 Hz frequencies (Ding and Simon, 2012b; Gross et al., 2013; Koskinen and Seppä, 2014; Ding et al., 2016). Since the strength of this 4–8 Hz cortical tracking is related to speech intelligibility (Luo and Poeppel, 2007; Peelle et al., 2013; Doelling et al., 2014), it has been hypothesized to reflect active speech perception mechanisms, very likely involved in parsing incoming connected speech into discrete syllabic units (Giraud and Poeppel, 2012; Teng et al., 2018).

We demonstrated that children's auditory system is less proficient than adults' auditory system at tracking the attended speech envelope at 1–4 and 4–8 Hz. Still, without noise, about half of children showed significant cortical tracking of the attended speech envelope at 1–4 and 4–8 Hz. Crucially, cortical tracking of the speech envelope at syllable rate correlated in the right AC positively with children's age.

Furthermore, 4–8 Hz tracking was sensitive to noise intensity so that the number of children with significant tracking decreased with increasing noise level, from 11 subjects in noiseless to only 2 subjects at −5 dB. Finally, at 4–8 Hz, children did not exhibit the selective tracking of speech in auditory areas seen in adults.

These results are at odds with at least one of the views claiming that (1) the sensitivity to syllables is acquired early in age and that (2) tracking at 4–8 Hz relates to the processing of syllabic units. Several studies highlighted the early-developed ability of infants and children to discriminate syllables. Neurophysiological evidence of syllable discrimination has been found already in preterm infants (Mahmoudzadeh et al., 2013), and 8-month-old infants are able to segregate newly learned words from syllable strings (Saffran et al., 1996). Furthermore, speech slow-amplitude modulations are one of the first cues that children identify when they listen to speech (Nittrouer, 2006), allowing them to detect syllable boundaries, which is mandatory for an accurate perception of speech (Goswami, 2011; Poelmans et al., 2011). However, recent findings indicate that children under 10 years of age are less accurate than adults at identifying syllable boundaries when these are defined only by amplitude modulations in temporal envelope (Cameron et al., 2018), and that theta band cortical tracking is not speech specific (Molinaro and Lizarazu, 2018). Also questioning the link between syllable processing and the 4–8 Hz tracking is the consistent finding that such coupling is right hemisphere dominant (Luo and Poeppel, 2007; Giraud and Poeppel, 2012; Gross et al., 2013; Peelle et al., 2013). In this context, our results argue for progressive evolution from childhood to adulthood of abilities to track the acoustic envelope of speech at 4–8 Hz.

Noise easily corrupts cortical tracking of speech in children

In adverse listening conditions (i.e., −5 dB), the auditory system lost, in a substantial proportion of children (70%), the capability to track the attended stream at <1 Hz, whereas 80% of adults exhibited significant tracking in this condition. Since children's SiN behavioral performances are typically lower than those of adults (Elliott, 1979), we can postulate that the poor performance was at least partially related to a limited central auditory capacity to segregate the attended stream from the multitalker background at <1 Hz when SNR decreased. This ability of the auditory system likely improves during adolescence, given the outcome in young adults reported here and previously (Ding and Simon, 2012a, 2013b; Vander Ghinst et al., 2016). Hence, our study argues for a developmental origin of the selective cortical tracking of the attended stream at <1 Hz. These data and our results are in line with previous psychoacoustic studies that demonstrated children's poor speech comprehension in adverse listening conditions (Johnson, 2000; Talarico et al., 2007; Neuman et al., 2010; Klatte et al., 2013).

Children's and adults' auditory systems track the attended speech stream in SiN conditions at <1 and at 1–4 Hz

At <1 and 1–4 Hz, coupling between envelopes of the listened sounds and the activity of nonprimary auditory cortex was stronger with the attended stream than with the global scene, both in children and adults. This finding is in line with former studies conducted in adults (Ding and Simon, 2012a, 2013b; Mesgarani and Chang, 2012; Zion Golumbic et al., 2013; Vander Ghinst et al., 2016). Yet, and in contradiction with our initial hypothesis, the current study is the first to demonstrate that this selective tracking already exists in children, at least up to a certain noise level. The preferential tracking of the slowest (<1 Hz) speech modulations took place, as in adults, at bilateral STS, demonstrating that this brain area extracts or has a preferential access to the attended stream in reasonable SNR conditions. Still, the preferential tracking of the attended stream at left IFG at 0 dB was higher in adults than in children. Moreover, the absence of specific tracking of the multitalker background argues for an object-based neural coding of the attended speaker's voice in children's higher-order auditory cortical areas up to a certain SNR level (Simon, 2015; Puvvada and Simon, 2017). Interestingly, at 1–4 Hz, this selective tracking occurred in bilateral AC in adults but only in right AC in children. Since recent findings have shown that perceptually relevant word segmentation takes place in left temporal cortex (Keitel et al., 2018), the lack of selective cortical tracking of speech at word repetition rate (1–4 Hz) in left temporal cortex could partly explain the SiN difficulties in children.

Effects of the multitalker background on hemispheric lateralization of cortical tracking of speech at <1 Hz

As demonstrated here and previously (Power et al., 2012; Vander Ghinst et al., 2016; Destoky et al., 2019), the left hemisphere cortical tracking of speech at <1 Hz is essentially preserved in a multitalker background up to a SNR of 0 dB, while it is compromised in the right hemisphere as soon as a background noise is added. Ours is the first study to demonstrate that this hemispheric asymmetry in cortical tracking occurred similarly in children and in adults, but with the noticeable difference that children lost the tracking in the most challenging conditions.

Left hemisphere-dominant noise-resistant cortical tracking of speech at STS (and IFG in adults) could imply an active cognitive process that promotes speech recognition (Schroeder et al., 2008; Schroeder and Lakatos, 2009; Power et al., 2012) through increased access to lexical and semantic representations (Binder et al., 2009; Liebenthal et al., 2014). This left-lateralized process is likely related to correct identification and comprehension of the targeted auditory stream (Alain et al., 2005; Bishop and Miller, 2009). Since the coupling between cortical oscillations and the low-frequency rhythmic structure of an attended acoustic stream seems to be under attentional control (Lakatos et al., 2013), the differential hemispheric effect of noise on cortical tracking of speech could be related to mechanisms of selective attention. Because noise impairs children more strongly than adults not only in auditory- and speech-related tasks, but also in nonauditory cognitive processes, such as reading and writing (for review, see Klatte et al., 2013), we can hypothesize that the detrimental effect of acute and chronic noise exposure on different cognitive functions is at least partially related to the crucial attentional load needed to understand speech in adverse hearing environments.

Conclusion

The ability of children's brains to track speech temporal envelopes at syllable rate (4–8 Hz) was drastically reduced in comparison with adults, regardless of the SNR. Similar to adults, children displayed stronger tracking of the attended stream than of the global scene in SiN conditions at phrasal and word rates, but their tracking ability was more easily corrupted by increasing noise. Children's poor SiN comprehension performances were therefore likely related to a limited central auditory capacity to segregate the attended stream from the multitalker background at phrasal and word rates as the SNR decreased and at 4–8 Hz regardless of the SNR. These results further elucidate the neurophysiological mechanisms accounting for children's difficulties to adequately segregate speech in informational masking conditions.

References

Alain C, Reinke K, McDonald KL, Chau W, Tam F, Pacurar A, Graham S (2005) Left thalamo-cortical network implicated in successful speech separation and identification. Neuroimage 26:592–599.

Ashburner J, Friston KJ (1999) Nonlinear spatial normalization using basis functions. Hum Brain Mapp 7:254–266.

Ashburner J, Neelin P, Collins DL, Evans A, Friston K (1997) Incorporating prior knowledge into image registration. Neuroimage 6:344–352.

Berman S, Friedman D (1995) The development of selective attention as reflected by event-related brain potentials. J Exp Child Psychol 59:1–31.

Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb Cortex 19:2767–2796.

Bishop CW, Miller LM (2009) A multisensory cortical network for understanding speech in noise. J Cogn Neurosci 21:1790–1805.

Bortel R, Sovka P (2007) Approximation of statistical distribution of magnitude squared coherence estimated with segment overlapping. Signal Processing 87:1100–1117.

Bourguignon M, Jousmäki V, Op de Beeck M, Van Bogaert P, Goldman S, De Tiège X (2012) Neuronal network coherent with hand kinematics during fast repetitive hand movements. Neuroimage 59:1684–1691.

Bourguignon M, De Tiège X, Op de Beeck M, Ligot N, Paquier P, Van Bogaert P, Goldman S, Hari R, Jousmäki V (2013) The pace of prosodic phrasing couples the listener's cortex to the reader's voice. Hum Brain Mapp 34: 314–326.

Bourguignon M, Piitulainen H, De Tiège X, Jousmäki V, Hari R (2015) Corticokinematic coherence mainly reflects movement-induced proprioceptive feedback. Neuroimage 106:382–390.

Bourguignon M, Molinaro N, Wens V (2018) Contrasting functional imaging parametric maps: the mislocation problem and alternative solutions. Neuroimage 169:200–211.

Cameron S, Chong-White N, Mealings K, Beechey T, Dillon H, Young T (2018) The parsing syllable envelopes test for assessment of amplitude modulation discrimination skills in children: development, normative data, and test-retest reliability studies. J Am Acad Audiol 29:151–163.

Carrette E, Op de Beeck M, Bourguignon M, Boon P, Vonck K, Legros B, Goldman S, Van Bogaert P, De Tiège X (2011) Recording temporal lobe epileptic activity with MEG in a light-weight magnetic shield. Seizure 20:414–418.

Clumeck C, Suarez Garcia S, Bourguignon M, Wens V, Op de Beeck M, Marty B, Deconinck N, Soncarrieu MV, Goldman S, Jousmäki V, Van Bogaert P, De Tiège X (2014) Preserved coupling between the reader's voice and the listener's cortical activity in autism spectrum disorders. PLoS One 9:e92329.

Dale AM, Sereno MI (1993) Improved localizadon of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. J Cogn Neurosci 5:162–176.

Demanez L, Dony-Closon B, Lhonneux-Ledoux E, Demanez JP (2003) Central auditory processing assessment: a French-speaking battery. Acta Otorhinolaryngol Belg 57:275–290.

Destoky F, Philippe M, Bertels J, Verhasselt M, Coquelet N, Vander Ghinst M, Wens V, De Tiège X, Bourguignon M (2019) Comparing the potential of MEG and EEG to uncover brain tracking of speech temporal envelope. Neuroimage 184:201–213.

De Tiège X, Op de Beeck M, Funke M, Legros B, Parkkonen L, Goldman S, Van Bogaert P (2008) Recording epileptic activity with MEG in a light-weight magnetic shield. Epilepsy Res 82:227–231.

Ding N, Simon JZ (2012a) Emergence of neural encoding of auditory objects while listening to competing speakers. Proc Natl Acad Sci U S A 109:11854–11859.

Ding N, Simon JZ (2012b) Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. J Neurophysiol 107:78–89.

Ding N, Simon JZ (2013a) Robust cortical encoding of slow temporal modulations of speech. Adv Exp Med Biol 787:373–381.

Ding N, Simon JZ (2013b) Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. J Neurosci 33:5728–5735.

Ding N, Melloni L, Zhang H, Tian X, Poeppel D (2016) Cortical tracking of hierarchical linguistic structures in connected speech. Nat Neurosci 19:158–164.

Doelling KB, Arnal LH, Ghitza O, Poeppel D (2014) Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage 85:761–768.

Drullman R, Festen JM, Plomp R (1994) Effect of temporal envelope smearing on speech reception. J Acoust Soc Am 95:1053–1064.

Elliott LL (1979) Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. J Acoust Soc Am 66:651–653.

Faes L, Pinna GD, Porta A, Maestri R, Nollo G (2004) Surrogate data analysis for assessing the significance of the coherence function. IEEE Trans Biomed Eng 51:1156–1166.

Ferstl EC, Walther K, Guthke T, von Cramon DY (2005) Assessment of story comprehension deficits after brain damage. J Clin Exp Neuropsychol 27:367–384.

Giraud AL, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci 15: 511–517.

Goswami U (2011) A temporal sampling framework for developmental dyslexia. Trends Cogn Sci 15:3–10.

Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Parkkonen L, Hämäläinen MS (2014) MNE software for processing MEG and EEG data. Neuroimage 86:446–460.

Gross J, Kujala J, Hämäläinen M, Timmermann L, Schnitzler A, Salmelin R (2001) Dynamic imaging of coherent sources: studying neural interactions in the human brain. Proc Natl Acad Sci U S A 98:694–699.

Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, Garrod S (2013) Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS Biol 11:e1001752.

Halliday DM, Rosenberg JR, Amjad AM, Breeze P, Conway BA, Farmer SF (1995) A framework for the analysis of mixed time series/point process data–theory and application to the study of physiological tremor, single motor unit discharges and electromyograms. Prog Biophys Mol Biol 64: 237–278.

Hämäläinen MS, Ilmoniemi RJ (1994) Interpreting magnetic fields of the brain: minimum norm estimates. Med Biol Eng Comput 32:35–42.

Hämäläinen ML, Lin F-H, Mosher J (2010) Anatomically and functionally constrained minimum-norm estimates. In: MEG: an introduction to methods (Hansen P, Kringelbach M, Salmelin R, eds), pp 186–215. New York: Oxford UP.

Hoen M, Meunier F, Grataloup C-L, Pellegrino F, Grimault N, Perrin F, Perrot X, Collet L (2007) Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. Speech Commun 49:905–916.

Johnson CE (2000) Children's phoneme identification in reverberation and noise. J Speech Lang Hear Res 43:144–157.

Keitel A, Gross J, Kayser C (2018) Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. PLoS Biol 16:e2004473.

Klatte M, Bergström K, Lachmann T (2013) Does noise affect learning? A short review on noise effects on cognitive performance in children. Front Psychol 4:578.

Koskinen M, Seppä M (2014) Uncovering cortical MEG responses to listened audiobook stories. Neuroimage 100:263–270.

Lakatos P, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE (2013) The spectrotemporal filter mechanism of auditory selective attention. Neuron 77:750–761.

Liebenthal E, Desai RH, Humphries C, Sabri M, Desai A (2014) The functional organization of the left STS: a large scale meta-analysis of PET and fMRI studies of healthy adults. Front Neurosci 8:289.

Lilliefors HW (1967) On the Kolmogorov-Smirnov test for normality with mean and variance unknown. J Am Stat Assoc 62:399–402.

Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54:1001–1010.

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 U S A 110:4846–4851.

Massaro DW (2017) Reading aloud to children: benefits and implications for acquiring literacy before schooling begins. Am J Psychol 130:63–72.

Mesgarani N, Chang EF (2012) Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485:233–236.

Molinaro N, Lizarazu M (2018) Delta(but not theta)-band cortical entrainment involves speech-specific processing. Eur J Neurosci 48:2642–2650.

Molinaro N, Lizarazu M, Lallier M, Bourguignon M, Carreiras M (2016) Out-of-synchrony speech entrainment in developmental dyslexia. Hum Brain Mapp 37:2767–2783.

Moore DR, Ferguson MA, Edmondson-Jones AM, Ratib S, Riley A (2010) Nature of auditory processing disorder in children. Pediatrics 126:e382–390.

Muzik O, Chugani DC, Juhász C, Shen C, Chugani HT (2000) Statistical parametric mapping: assessment of application in children. Neuroimage 12:538–549.

Neuman AC, Wroblewski M, Hajicek J, Rubinstein A (2010) Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults. Ear Hear 31:336–344.

Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp 15:1–25.

Nittrouer S (2006) Children hear the forest. J Acoust Soc Am 120:1799–1802.

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97–113.

Peelle JE, Gross J, Davis MH (2013) Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb Cortex 23:1378–1387.

Perrin F, Grimault N (2005) Fonds Sonores (Version 1.0) [Sound samples]. Available at: http://olfac.univ-lyon1.fr/unite/equipe-02/FondsSonores.html

Poelmans H, Luts H, Vandermosten M, Boets B, Ghesquière P, Wouters J (2011) Reduced sensitivity to slow-rate dynamic auditory information in children with dyslexia. Res Dev Disabil 32:2810–2819.

Power AJ, Foxe JJ, Forde EJ, Reilly RB, Lalor EC (2012) At what time is the cocktail party? A late locus of selective attention to natural speech. Eur J Neurosci 35:1497–1503.

Power AJ, Colling LJ, Mead N, Barnes L, Goswami U (2016) Neural encoding of the speech envelope by children with developmental dyslexia. Brain Lang 160:1–10.

Puvvada KC, Simon JZ (2017) Cortical representations of speech in a multitalker auditory scene. J Neurosci 37:9189–9196.

Reiss AL, Abrams MT, Singer HS, Ross JL, Denckla MB (1996) Brain development, gender and IQ in children. A volumetric imaging study. Brain 119:1763–1774.

Reuter M, Schmansky NJ, Rosas HD, Fischl B (2012) Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 61:1402–1418.

Rosenberg JR, Amjad AM, Breeze P, Brillinger DR, Halliday DM (1989) The fourier approach to the identification of functional coupling between neuronal spike trains. Prog Biophys Mol Biol 53:1–31.

Saffran JR, Aslin RN, Newport EL (1996) Statistical learning by 8-month-old infants. Science 274:1926–1928.

Sanes DH, Woolley SM (2011) A behavioral framework to guide research on central auditory development and plasticity. Neuron 72:912–929.

Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci 32:9–18.

Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillations and visual amplification of speech. Trends Cogn Sci 12:106–113.

Simon JZ (2015) The encoding of auditory objects in auditory cortex: insights from magnetoencephalography. Int J Psychophysiol 95:184–190.

Simpson SA, Cooke M (2005) Consonant identification in N-talker babble is a nonmonotonic function of N. J Acoust Soc Am 118:2775–2778.

Sussman E, Steinschneider M (2009) Attention effects on auditory scene analysis in children. Neuropsychologia 47:771–785.

Talarico M, Abdilla G, Aliferis M, Balazic I, Giaprakis I, Stefanakis T, Foenander K, Grayden DB, Paolini AG (2007) Effect of age and cognition on childhood speech in noise perception abilities. Audiol Neurootol 12:13–19.

Taulu S, Simola J (2006) Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys Med Biol 51:1759–1768.

Taulu S, Simola J, Kajola M (2005) Applications of the signal space separation method. IEEE Trans Signal Process 53:3359–3372.

Teng X, Tian X, Doelling K, Poeppel D (2018) Theta band oscillations reflect more than entrainment: behavioral and neural evidence demonstrates an active chunking process. Eur J Neurosci 48:2770–2782.

Thompson EC, Woodruff Carr K, White-Schwoch T, Otto-Meyer S, Kraus N (2017) Individual differences in speech-in-noise perception parallel neural speech processing and attention in preschoolers. Hear Res 344:148–157.

Vander Ghinst M, Bourguignon M, Op de Beeck M, Wens V, Marty B, Hassid S, Choufani G, Jousmäki V, Hari R, Van Bogaert P, Goldman S, De Tiège X (2016) Left superior temporal gyrus is coupled to attended speech in a cocktail-party auditory scene. J Neurosci 36:1596–1606.

Wens V, Marty B, Mary A, Bourguignon M, Op de Beeck M, Goldman S, Van Bogaert P, Peigneux P, De Tiège X (2015) A geometric correction scheme for spatial leakage effects in MEG/EEG seed-based functional connectivity mapping. Hum Brain Mapp 36:4604–4621.

Wightman FL, Kistler DJ (2005) Informational masking of speech in children: effects of ipsilateral and contralateral distracters. J Acoust Soc Am 118:3164–3176.

Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, Poeppel D, Schroeder CE (2013) Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party." Neuron 77:980–991.