Ghinst, Marc Vander; Bourguignon, Mathieu; Niesen, Maxime; Wens, Vincent; Hassid, Sergio; Choufani, Georges; Jousmäki, Veikko; Hari, Riitta; Goldman, Serge; De Tiège, Xavier

Cortical tracking of speech-in-noise develops from childhood to adulthood

Published in:
JOURNAL OF NEUROSCIENCE

DOI:
10.1523/JNEUROSCI.1732-18.2019

Published: 10/04/2019

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Ghinst, M. V., Bourguignon, M., Niesen, M., Wens, V., Hassid, S., Choufani, G., Jousmäki, V., Hari, R., Goldman, S., & De Tiège, X. (2019). Cortical tracking of speech-in-noise develops from childhood to adulthood. *JOURNAL OF NEUROSCIENCE*, *39*(15), 2938–2950. https://doi.org/10.1523/JNEUROSCI.1732-18.2019

JNeurosci
THE JOURNAL OF NEUROSCIENCE

Research Articles: Behavioral/Cognitive

Cortical tracking of speech-in-noise develops from childhood to adulthood

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

[1]*Laboratoire de Cartographie fonctionnelle du Cerveau, UNI — ULB Neuroscience Institute, Université libre de Bruxelles (ULB), 1070 Brussels, Belgium.*

[2]*Service d'ORL et de chirurgie cervico-faciale, ULB-Hôpital Erasme, Université libre de Bruxelles (ULB), 1070 Brussels, Belgium.*

[3]*Laboratoire Cognition Langage et Développement, UNI — ULB Neuroscience Institute, Université libre de Bruxelles (ULB), Brussels, Belgium.*

[4]*Basque Center on Cognition, Brain and Language (BCBL), Donostia/San Sebastian, Spain.*

[5]*Department of Functional Neuroimaging, Service of Nuclear Medicine, CUB Hôpital Erasme, Université libre de Bruxelles (ULB), Brussels, Belgium.*

[6]*Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, PO BOX 15100, FI-00076-AALTO, Espoo, Finland.*

[7]*Cognitive Neuroimaging Centre, Lee Kong Chian School of Medicine, 59 Nanyang Drive, Nanyang Technological University, Singapore 636921.*

[8]*Department of Art, Aalto University School of Arts, Design, and Architecture,, PO Box 31000, 00076 AALTO, Helsinki, Finland.*

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

Received: 10 July 2018

Revised: 8 January 2019

Accepted: 12 January 2019

Published: 11 February 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.V.G., M.B., M.N., V.W., and X.D.T. analyzed data; M.V.G. wrote the first draft of the paper; M.V.G., M.B., R.H., and X.D.T. wrote the paper; M.B., M.N., V.W., V.J., and R.H. contributed unpublished reagents/analytic tools; M.B., V.W., S.H., G.C., V.J., R.H., S.G., and X.D.T. edited the paper.

Conflict of Interest: The authors declare no competing financial interests.

Marc Vander Ghinst and Maxime Niesen were supported by a research grant from the Fonds Erasme (Brussels, Belgium). Mathieu Bourguignon 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 the Marie Sk#odowska-Curie Action of the European Commission (grant 743562). Veikko Jousmäki 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). Xavier De Tiège is Post-doctorate Clinical Master Specialist at the FRS-FNRS.

Corresponding author: Marc Vander Ghinst, Laboratoire de Cartographie fonctionnelle du Cerveau, UNI — ULB Neuroscience Institute, Université libre de Bruxelles (ULB), 808 Lennik Street, 1070 Brussels, Belgium. Telephone: +32.2.555.31.11, Fax: +32.2.555.47.01, E-mail: Marc.Vander.Ghinst@erasme.ulb.ac.be

Cite as: J. Neurosci 2019; 10.1523/JNEUROSCI.1732-18.2019

Cortical tracking of speech-in-noise develops from childhood to adulthood

Abbreviated Title: Cortical processing of speech-in-noise in children

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

[1] Laboratoire de Cartographie fonctionnelle du Cerveau, UNI – ULB Neuroscience Institute, Université libre de Bruxelles (ULB), 1070 Brussels, Belgium.

[2] Service d'ORL et de chirurgie cervico-faciale, ULB-Hôpital Erasme, Université libre de Bruxelles (ULB), 1070 Brussels, Belgium.

[3] Laboratoire Cognition Langage et Développement, UNI – ULB Neuroscience Institute, Université libre de Bruxelles (ULB), Brussels, Belgium.

[4] Basque Center on Cognition, Brain and Language (BCBL), Donostia/San Sebastian, Spain.

[5] Department of Functional Neuroimaging, Service of Nuclear Medicine, CUB Hôpital Erasme, Université libre de Bruxelles (ULB), Brussels, Belgium.

[6] Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, PO BOX 15100, FI-00076-AALTO, Espoo, Finland.

[7] Cognitive Neuroimaging Centre, Lee Kong Chian School of Medicine, 59 Nanyang Drive, Nanyang Technological University, Singapore 636921.

[8] Department of Art, Aalto University School of Arts, Design, and Architecture, PO Box 31000, 00076 AALTO, Helsinki, Finland.

*These authors equally contributed to this work.

Corresponding author: Marc Vander Ghinst, Laboratoire de Cartographie fonctionnelle du Cerveau, UNI – ULB Neuroscience Institute, Université libre de Bruxelles (ULB), 808 Lennik Street, 1070 Brussels, Belgium. Telephone: +32.2.555.31.11, Fax: +32.2.555.47.01, E-mail: Marc.Vander.Ghinst@erasme.ulb.ac.be.

Number of figures: 7, number of table: 1, number of words in *Abstract*: 250, number of words in *Introduction*: 497, number of words in *Discussion*: 1500.

Conflict of interest

The authors have no conflict of interest to declare.

Acknowledgments

Marc Vander Ghinst and Maxime Niesen were supported by a research grant from the Fonds Erasme (Brussels, Belgium). Mathieu Bourguignon 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 the Marie Skłodowska-Curie Action of the European Commission (grant 743562). Veikko Jousmäki 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). Xavier De Tiège is Post-doctorate Clinical Master Specialist at the FRS-FNRS.

This study and the MEG project at the CUB Hôpital Erasme was financially supported by the Fonds Erasme (research convention “Les Voies du Savoir”, Fonds Erasme, Brussels, Belgium).

We thank Nicolas Grimault and Fabien Perrin (Centre de Recherche en Neuroscience, Lyon, France) for providing us their entire audio material. We also thank Brice Marty for providing his help for data acquisition.

Abstract

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 twenty children (6–9 years; 8 females) and twenty adults (21–40 years; 10 females). During MEG recordings, subjects attended to 4 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 Hz, 1–4 Hz, and 4–8 Hz in all SNR conditions. Children displayed similar coupling at < 1 Hz 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.

Significance statement:

Behaviorally, children are worse than adults at understanding speech-in-noise. Here, neuromagnetic signals were recorded while healthy adults and typically developing 6–9-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 mechanisms 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 (Sanes and Woolley, 2011). However, no study has so far demonstrated such phenomena 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 below 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; Giraud and Poeppel, 2012; Ding and Simon, 2013a; 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 below 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) 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) and adults (21–40 years) 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: (i) *Attended stream* (i.e., reader's voice only), (ii) unattended *Multitalker background* only, and (iii) *Global scene* (i.e., the combination of the *Attended stream* and the unattended *Multitalker background*).

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 yrs, range 6–9 yrs; 8 females and 12 males) and twenty native French-speaking healthy adults (mean age 30 yrs, range 21–40 yrs; 10 females and 10 males) without any history of neuropsychiatric or otologic disorder participated in this study. All subjects had normal hearing according to pure tone audiometry (i.e., normal hearing thresholds (between 0–20 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 five minutes. 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 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 Hz, 3.6 Hz, and 5.54 Hz) and in adults (0.49 Hz, 3.39 Hz, 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 attended reader's voice, and noise was the multi-talker background) was randomly assigned to

each story: a noiseless condition, and 3 SiN conditions (with SNRs of +5 dB, 0 dB, and –5 dB), leading to four different SNR conditions. This randomization procedure prevented any systematic association between stories and SNRs. The noise (Fonds sonores v-1.0, Perrin & 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 configuration of cocktail-party noise was selected because it accounts for both energetic and informational masking at phonetic and lexical level (Simpson and Cooke, 2005; Hoen et al., 2007). Sound recordings were played using VLC media player (VideoLAN Project, GNU General Public License) running on a MacBook Pro (Apple Computer, Inc., Cupertino, CA). Sound signals were transmitted to a MEG-compatible 60 × 60 cm^2 high-quality flat panel loudspeaker (Panphonics SSH sound shower, Panphonics Oy, Espoo, Finland) 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, Belgium). 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 ULB-Hôpital Erasme using a whole-scalp-covering 306-channel MEG device (for 15 adults and 12 children Elekta Neuromag

Vectorview by Elekta Oy, Helsinki, Finland, and otherwise Elekta Neuromag Triux by MEGIN, Helsinki, Finland) installed in a light-weight magnetically shielded room (Maxshield, MEGIN, Helsinki, Finland), the characteristics of which being described elsewhere (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, Colchester, VT). Electro-oculogram (EOG), electrocardiogram (ECG), and the audio signals presented to the subjects were recorded simultaneously with MEG signals (passband 0.1–330 Hz for EOG and ECG, and 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 50–22000 Hz and sampled at 44.1 kHz. High-resolution 3D-T1 cerebral magnetic resonance images (MRI) were acquired on a 1.5 T MRI (Intera, Philips, The Netherlands).

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 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

177 artifacts in the MEG sensors. The number of artifact-free epochs was 695 ± 66 (mean ± SD
178 across subjects and conditions) in the adult group and 621 ± 93 in the children group.
179 A two-way repeated-measures ANOVA (factors: age group and condition; dependent
180 variable: number of epochs used to compute coherence) revealed a significant effect of group
181 on the number of epochs ($F_{1,114}$ = 4.42, p = 0.042) but no effect of SNR ($F_{3,114}$ = 0.03, p =
182 0.99) or interaction ($F_{3,114}$ = 1.46, p = 0.23). To avoid a possible methodological bias in our
183 results due to differences between age groups in the accuracy of speech tracking estimation,
184 we threw away epochs in adults' data so as to equalize the number of epochs in both groups.
185
186
187 ***Coherence analyses in sensor space***
188 For each listening condition, synchronization between the temporal envelope of wide-
189 band (50–22000 Hz) audio signals and artifact-free MEG epochs (2048-ms-long) was
190 assessed with coherence analysis in sensor space at frequencies in which speech temporal
191 envelope is critical for speech comprehension (i.e., 0.1–20 Hz) (Drullman et al., 1994).
192 Coherence is an extension of Pearson correlation coefficient to the frequency domain. It
193 quantifies the degree of coupling between two signals (say $x(t)$ and $y(t)$), providing a
194 number between 0 (no linear dependency) and 1 (perfect linear dependency) for each
195 frequency bin (Halliday et al., 1995). Coherence was computed as follows:

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

196 where $P_{xx}(f) = \sum_k |\hat{x}_k(f)^2|$, $P_{yy}(f) = \sum_k |\hat{y}_k(f)^2|$, and $P_{xy}(f) = \sum_k |\hat{x}_k(f)\hat{y}_k^*(f)|$, and
197 where $\hat{x}_k(f)$ (respectively $\hat{y}_k(f)$) is the Fourier coefficient of the k^{th} epoch of signal $x(t)$
198 (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 dB*, *0 dB*, 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 Coh_{global}, 2) the *Attended stream* only (i.e., the reader's voice), leading to Coh_{att}, and 3) the *Multitalker background* only, leading to Coh_{bckgr}. Sensor-level coherence maps were obtained using gradiometer signals only, and signals from gradiometer pairs were combined as done in 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 around 0.5 Hz, with a sensitivity profile proportional to the Fourier transform of a boxcar function: sinc(□(*f*−0.5 Hz)/0.5 Hz). The < 1 Hz, 1–4 Hz, and 4–8 Hz frequency ranges are henceforth referred to as *frequency bands of interest*.

Coherence analyses in source space

Individual MRIs 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 homogenous 5-mm-grid source space covering the whole brain, using MNE suite (Martinos Centre for Biomedical Imaging, Boston, MA, 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 dB*, *0 dB*, and *–5 dB*), audio signal (*Global scene*, *Attended stream*, *Multitalker Background*) and frequency bands of interest (< 1 Hz, 1–4 Hz, 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 non-linear transformation from individual MRIs 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 MRIs 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 above 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 $f(\cdot) = \mathrm{arctanh}(\sqrt{\cdot})$, 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 over-contribution 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 vs. children

Children' 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 dB*, *0 dB*, 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 test (Lilliefors, 1967).

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 differ also due to differences in the visual analogue scales: 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 as it takes into account the temporal auto-correlation within signals. For each subject, 1000 surrogate sensor-level coherence maps were computed as 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 non-parametric 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 prior to 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 of the difference map's absolute value 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 corresponding voxel's difference value (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 neighbouring 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 et al., 1994; Wens et al., 2015; Bourguignon et al., 2017).

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 Coh_{global} maps using the same non-parametric

permutation test described above, but with the labels *Global scene* and *Attended stream* instead of *audio* and *Rest*, leading to the $Coh_{att} - Coh_{global}$ difference maps.

Comparison of the tracking in adults vs. 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 $Coh_{att,adults}$ and $Coh_{att,children}$ instead of Coh_{att} and Coh_{global}, leading to the $Coh_{att,adults} - Coh_{att,children}$ 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 3-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 dB*, *0 dB*, 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 for 1–4 Hz and 4–8 Hz Coh_{at} values (*ps* < 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 (*ps* > 0.05). Therefore, we report on only the results of the later ANOVAs. Post-hoc comparisons were performed with pairwise t-tests.

346 Based on our results in the *Noiseless* condition, we conducted an additional analysis by
347 computing Pearson correlation between children's age and their maximum coherence values,
348 separately for both hemispheres in the three frequency bands of interest.
349
350
351
352

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.

Comparison of SiN perception in adults vs. 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).

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 3 other conditions (*+5 dB*, *0 dB*, and *–5 dB*). Doing otherwise would have violated the homoscedasticity assumption of the ANOVA.

The ANOVA performed on intelligibility rating 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 rating decreased statistically significantly

379 from *Noiseless* to *+5 dB* (adults, $t_{19} = 6.28$, $p < 0.0001$; children, $t_{19} = 5.06$, $p < 0.0001$), from
380 *+5 dB* to *0 dB* (adults, $t_{19} = 7.19$, $p < 0.0001$; children, $t_{19} = 2.11$, $p = 0.048$) and from *0 dB* to
381 *–5 dB* (adults, $t_{19} = 9.48$, $p < 0.0001$; children, $t_{19} = 5.09$, $p < 0.0001$). Comparison between
382 adults and children revealed that children gave lower intelligibility ratings than adults in
383 Noiseless ($t_{38} = 2.26$, $p = 0.030$) and *+5 dB* conditions ($t_{38} = 2.10$, $p = 0.042$) but not in the
384 two other noisiest conditions ($ps > 0.05$).
385 The ANOVA performed on comprehension scores—converted to percentage correct—
386 revealed a significant effect of SNR ($F_{3,114} = 27.6$; $p < 0.0001$), a significant effect of age
387 group ($F_{1,38} = 19.4$; $p < 0.0001$), and no significant interaction ($F_{3,114} = 0.66$; $p = 0.58$). The
388 Lilliefors test showed that the distribution of the residuals did not deviate significantly from a
389 normal distribution ($p = 0.054$). Comprehension scores were higher in adults (80.5 ± 15.2 %;
390 mean ± SD across conditions and participants) than in children (68.6 +/- 18.7 %), and
391 decreased statistically significantly from *Noiseless* to *+5 dB* ($t_{39} = 3.13$, $p = 0.0033$), from *+5*
392 *dB* to *0 dB* ($t_{39} = 2.13$, $p = 0.040$) and from *0 dB* to *–5 dB* ($t_{39} = 4.00$, $p = 0.0003$).
393
394 ---Place Figure 1 around here ---
395
396
397 ***Cortical tracking of speech in the Noiseless condition***
398 Figures 2–4 display group-averaged coherence spectra (Fig. 2), sensor distribution (Fig.
399 3) and source distribution (Fig. 4) in all conditions and in both groups.
400 Table 1 provides the number of children and adults showing significant sensor-space
401 Coh_{global}, Coh_{att}, and Coh_{bckgr} at <1 Hz, 1–4 Hz and 4–8 Hz frequencies.
402 In adults, statistically significant Coh_{att} was observed at < 1 Hz in 20/20 adults, at 1–4
403 Hz in 17/20 adults, and at 4–8 Hz in 17/20 adults in temporal-lobe MEG sensors in the
404 *Noiseless* condition (Figs. 2 and 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 [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, Coh_{att} peaked in bilateral supratemporal auditory cortices (AC) at 1–4 Hz (left [–65 –16 7], $p < 0.0001$; right [64 –5 6], $p < 0.0001$) and 4–8 Hz (left [–65 –15 9], $p < 0.0001$; right [64 –8 5], $p <$ 0.0001) (Fig. 4).

In children, statistically significant Coh_{att} was observed at < 1 Hz in 20/20 children, at 1–4 Hz in 10/20 children, and at 4–8 Hz in 11/20 children in temporal-lobe MEG sensors in the *Noiseless* condition (Figs. 2 and 3). In source space, group-level coherence at <1 Hz peaked at bilateral STS (left [–63 –10 6], $p < 0.0001$; right [63 –20 –3], $p < 0.0001$). It peaked at bilateral supratemporal AC at 1–4 Hz (left [–64 –11 13], $p = 0.0016$; right [62 –28 4], $p < 0.0001$), and in 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 Hz 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 (10/20 vs. 17/20; $p = 0.041$; Fisher exact test). A similar but statistically nonsignificant trend was observed at 4–8 Hz (11/20 vs. 17/20; $p =$ 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 [–65 –16 7], $p = 0.0013$; right [64 –5 6], $p = 0.0041$) and 4–8 Hz (left [–65 –15 9], $p < 0.0001$; right [64 –8 5], $p < 0.0001$), but not at < 1 Hz.

Correlation between Coh_{att} values and children' 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' age reached statistical significance (ps > 0.3).

- - -Place Figure 2 around here- - -

- - -Place Figure 3 around here- - -

- - -Place Figure 4 around Here- - -

- - -Place Figure 5 around here- - -

- - -Place Table 1 around Here - - -

Cortical tracking of speech in SiN conditions

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

In children, group-level Coh_{att} and Coh_{global} maps at <1 Hz displayed statistically significant (ps < 0.05) local maxima in 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 Coh_{global} maps displayed statistically significant (ps < 0.05) local maxima in AC bilaterally at *+5* and *0 dB*, and only Coh_{att} displayed significance coherence in right AC at *–5 dB* (Fig.4). At 4–8 Hz, Coh_{att} and Coh_{global} maps displayed statistically significant (ps < 0.05) local maxima only at right AC at *+5* and *0 dB*, but not at *–5 dB* (Fig.4). Coh_{bckgr} was not statistically significant in any condition and frequency band of interest.

Cortical processing of the auditory scene in SiN conditions

In adults, Coh_{att} was stronger than Coh_{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 dB*, *0 dB* and *–5 dB*: $p < 0.0001$) and left inferior frontal gyrus at *0 dB*. Significantly stronger Coh_{att} than Coh_{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 Coh_{global} in bilateral STS at every SNR (*+5 dB*, *0 dB* and *–5 dB*: *ps* < 0.0001) and in central opercular cortex at *–5 dB*. At 1–4 Hz, significantly stronger Coh_{att} than Coh_{global} was found in right STS (*+5 dB*, $p = 0.0011$; *0 dB*, $p = 0.0011$; *–5 dB*, $p = 0.0042$). At 4–8 Hz, significantly stronger Coh_{att} than Coh_{global} was found at *–5 dB* and in a non-auditory area (left superior frontal gyrus; $p = 0.034$).

Comparison of the tracking in adults vs. children

The comparison of coherence values between adults and children revealed that tracking at frequencies < 1 Hz was stronger in adults than 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 (*ps* < 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 (*ps* < 0.0001) (Fig. 6).

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

- - -Place Figure 6 around here- - -

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 3-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 ($ps > 0.05$). 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 *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} as a function of decreasing SNR (see Figs. 4 and 7). Supporting this interpretation, adults had higher Coh_{att} than children at *–5 dB* (t_{38} = 3.88, $p = 0.0004$; *t*-test) but not at *0*, *+5 dB* and *Noiseless* ($ps > 0.17$). The interaction between SNR condition and hemispheric lateralization was explained by a faster decrease in Coh_{att} in the right than the left STS as a function of decreasing SNR (see Figs. 4 and 6). Supporting this interpretation, post-hoc comparisons revealed that Coh_{att} at right STS decreased as soon as the

504 *Multitalker background* was added (*Noiseless* vs. *+5 dB*: $t_{19} = 4.9$, $p < 0.0001$) and it further
505 diminished as SNR decreased (*+5 dB* vs. *0 dB*, $t_{19} = 4.1$, $p = 0.0002$; *0 dB* vs. *–5 dB*, $t_{19} =$
506 4.23, $p = 0.0001$). In contrast, Coh_{att} in the left STS decreased significantly only in the two
507 noisiest conditions (*Noiseless* vs. *+5 dB*, $t_{19} = 0.49$, $p = 0.62$; *+5 dB* vs. *0 dB*, $t_{19} = 3.11$, $p =$
508 0.0035; *0 dB* vs. *–5 dB*, $t_{19} = 5.3$, $p < 0.0001$).

509 At 1–4 Hz, the ANOVA revealed a significant main effect of age group on Coh_{att} ($F_{3,114}$
510 = 10.04, $p = 0.003$), a significant main effect of hemispheric lateralization ($F_{3,114} = 10.04$, $p =$
511 0.003), no significant main effect of the SNR condition ($F_{3,114} = 1.53$, $p = 0.21$), and no
512 significant interactions (*ps* > 0.05). The main effect of age group was due to higher Coh_{att} in
513 adults than in children. The main effect of hemispheric lateralization was explained by higher
514 Coh_{att} values in the right than the left AC.

515 At 4–8 Hz, the ANOVA revealed a significant main effect of age group on Coh_{att} ($F_{3,114}$
516 = 54.64, $p < 0.0001$), a significant main effect of noise level ($F_{3,114} = 13.96$, $p < 0.0001$), a
517 significant effect of hemispheric lateralization ($F_{3,114} = 7.37$, $p = 0.0099$), a significant
518 interaction between SNR condition and age group ($F_{3,114} = 10.08$, $p < 0.0001$), and no other
519 significant interactions (*ps* > 0.05). The main effect of age group was explained by higher
520 Coh_{att} in adults than in children at all SNRs (*ps* < 0.001). The interaction between SNR
521 condition and age group was explained by the presence of SNR-related modulation in adults'
522 Coh_{att} and the absence of such modulation in children's Coh_{att} (see Fig. 7). Indeed, adults'
523 Coh_{att} (mean across hemispheres) was significantly higher at intermediate SNRs (*5 dB* and *0*
524 *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*
525 *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*
526 *dB* ($t_{19} = 2.30$, $p = 0.033$), and marginally higher in *Noiseless* than at *–5 dB* ($t_{19} = 1.96$, $p =$
527 0.065). The main effect of hemispheric lateralization was explained by higher or marginally

higher Coh_{att} in right than 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*, $t_{39} = 1.71$, $p = 0.095$).

- - -Place Figure 7 around Here - - -

Discussion

This study discloses commonalities between children and adult's cortical tracking of SiN. First, both children and adults' auditory systems similarly tracked the attended speaker's voice more than the global auditory scene at < 1 Hz 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. Compared with adults, children displayed (i) reduced cortical tracking of speech at 1–4 Hz, and particularly at 4–8 Hz (even in noiseless conditions); (ii) increasing multitalker background level compromised children more than adults in cortical tracking of speech at < 1 Hz, and (iii) children did not exhibit selective cortical representation of SiN at 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 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., 2017).

552 We demonstrated that children's auditory system is less proficient than adults' auditory
553 system at tracking the attended speech envelope at 1–4 Hz and 4–8 Hz. Still, without noise,
554 about half of children showed significant cortical tracking of the attended speech envelope at
555 1–4 Hz and 4–8 Hz. Crucially, cortical tracking of speech envelope at syllable rate correlated
556 in the right AC positively with the children's age.

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

561 These results are at odds with at least one of the views claiming that 1) the sensitivity to
562 syllables is acquired early in age and that 2) tracking at 4–8 Hz relates to processing of
563 syllabic units. Several studies highlighted the early-developed ability of infants and children
564 to discriminate syllables. Neurophysiological evidence of syllable discrimination has been
565 found already in preterm infants (Mahmoudzadeh et al., 2013) and 8-month-old infants are
566 able to segregate newly learned words from syllable strings (Saffran et al., 1996).
567 Furthermore, speech slow amplitude modulations are one of the first cues that children
568 identify when they listen to speech (Nittrouer, 2006), allowing them to detect syllable
569 boundaries, which is mandatory for an accurate perception of speech (Goswami, 2011;
570 Poelmans et al., 2011). However, recent findings indicate that children under 10 years are less
571 accurate than adults at identifying syllable boundaries when these are only defined by
572 amplitude modulations in temporal envelope (Cameron et al., 2018), and that theta band
573 cortical tracking is not speech-specific (Molinaro and Lizarazu, 2017). Also questioning the
574 link between syllable processing and the 4–8-Hz tracking is the consistent finding that such
575 coupling is right-hemisphere dominant (Luo and Poeppel, 2007; Giraud and Poeppel, 2012;
576 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*), 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 behavioural 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 elsewhere (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 are in line with our results and 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 system tracks the attended speech stream in SiN conditions at < 1 Hz and at 1–4 Hz.

At < 1 Hz and 1–4 Hz, coupling between envelopes of the listened sounds and the activity of non-primary 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 performed in adults (Ding and Simon, 2012a; Mesgarani and Chang, 2012; Ding and Simon, 2013b; 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, like 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, 2014; 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 SiN difficulties in children.

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

As demonstrated here and elsewhere (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 more 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 . As noise impairs children more strongly than adults in addition to auditory- and speech-related tasks also in non-auditory cognitive processes, such as reading and writing (for a 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 environment.

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. Similarly as 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 a function of decreasing SNR 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.

652

653

654 **References**
655

656 Alain C, Reinke K, McDonald KL, Chau W, Tam F, Pacurar A, Graham S (2005) Left
657 thalamo-cortical network implicated in successful speech separation and identification.
658 NeuroImage 26:592-599.
659 Ashburner J, Friston KJ (1999) Nonlinear spatial normalization using basis functions. Hum
660 Brain Mapp 7:254-266.
661 Ashburner J, Neelin P, Collins DL, Evans A, Friston K (1997) Incorporating prior knowledge
662 into image registration. NeuroImage 6:344-352.
663 Berman S, Friedman D (1995) The development of selective attention as reflected by event-
664 related brain potentials. J Exp Child Psychol 59:1-31.
665 Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic system? A
666 critical review and meta-analysis of 120 functional neuroimaging studies. Cereb
667 Cortex 19: 2767-2796.
668 Bishop CW, Miller LM (2009) A multisensory cortical network for understanding speech in
669 noise. J Cogn Neurosci 21:1790-1805.
670 Bortel R, Sovka P (2007) Approximation of statistical distribution of magnitude squared
671 coherence estimated with segment overlapping. Signal Processing 87:1100-1117.
672 Bourguignon M, Piitulainen H, De Tiege X, Jousmäki V, Hari R (2015) Corticokinematic
673 coherence mainly reflects movement-induced proprioceptive feedback. NeuroImage
674 106:382-390.
675 Bourguignon M, Jousmäki V, Op de Beeck M, Van Bogaert P, Goldman S, De Tiege X
676 (2012) Neuronal network coherent with hand kinematics during fast repetitive hand
677 movements. NeuroImage 59:1684-1691.

678 Bourguignon M, De Tiege X, Op de Beeck M, Ligot N, Paquier P, Van Bogaert P, Goldman
679 S, Hari R, Jousmäki (2013) The pace of prosodic phrasing couples the listener's cortex
680 to the reader's voice. Hum Brain Mapp 34:314-326.
681 Bourguignon M, Molinaro N, Wens V (2017) Contrasting functional imaging parametric
682 maps: The mislocation problem and alternative solutions. NeuroImage 169:200–211.
683 Cameron S, Chong-White N, Mealings K, Beechey T, Dillon H, Young T (2018) The parsing
684 syllable envelopes test for assessment of amplitude modulation discrimination skills in
685 children: development, normative data, and test-retest reliability studies. J Am Acad
686 Audiol 29:151-163.
687 Carrette E, Op de Beeck M, Bourguignon M, Boon P, Vonck K, Legros B, Goldman S, Van
688 Bogaert P, De Tiege X (2011) Recording temporal lobe epileptic activity with MEG in
689 a light-weight magnetic shield. Seizure 20:414-418.
690 Clumeck C, Suarez Garcia S, Bourguignon M, Wens V, Op de Beeck M, Marty B, Deconinck
691 N, Soncarrieu MV, Goldman S, Jousmäki V, Van Bogaert P, De Tiege X (2014)
692 Preserved coupling between the reader's voice and the listener's cortical activity in
693 autism spectrum disorders. PloS one 9:e92329.
694 Dale AM, Sereno MI (1993) Improved localizadon of cortical activity by combining EEG and
695 MEG with MRI cortical surface reconstruction: a linear approach. J Cogn Neurosci
696 5:162-176.
697 De Tiege X, Op de Beeck M, Funke M, Legros B, Parkkonen L, Goldman S, Van Bogaert P
698 (2008) Recording epileptic activity with MEG in a light-weight magnetic shield.
699 Epilepsy Res 82:227-231.
700 Demanez L, Dony-Closon B, Lhonneux-Ledoux E, Demanez JP (2003) Central auditory
701 processing assessment: a French-speaking battery. Acta Otorhinolaryngol Belg
702 57:275-290.

703 Destoky F, Philippe M, Bertels J, Verhasselt M, Coquelet N, Vander Ghinst M, Wens V, De
704 Tiège X, Bourguignon M (2019) Comparing the potential of MEG and EEG to
705 uncover brain tracking of speech temporal envelope. NeuroImage 184:201-213.
706 Ding N, Simon JZ (2012a) Emergence of neural encoding of auditory objects while listening
707 to competing speakers. Proc Natl Acad Sci U S A 109:11854-11859.
708 Ding N, Simon JZ (2012b) Neural coding of continuous speech in auditory cortex during
709 monaural and dichotic listening. J Neurophysiol 107:78-89.
710 Ding N, Simon JZ (2013a) Robust cortical encoding of slow temporal modulations of speech.
711 Adv Exp Med Biol 787:373-381.
712 Ding N, Simon JZ (2013b) Adaptive temporal encoding leads to a background-insensitive
713 cortical representation of speech. J Neurosci 33:5728-5735.
714 Ding N, Melloni L, Zhang H, Tian X, Poeppel D (2016) Cortical tracking of hierarchical
715 linguistic structures in connected speech. Nat Neurosci 19:158-164.
716 Doelling KB, Arnal LH, Ghitza O, Poeppel D (2014) Acoustic landmarks drive delta-theta
717 oscillations to enable speech comprehension by facilitating perceptual parsing.
718 NeuroImage 85 Pt 2:761-768.
719 Drullman R, Festen JM, Plomp R (1994) Effect of temporal envelope smearing on speech
720 reception. J Acoust Soc Am 95:1053-1064.
721 Elliott LL (1979) Performance of children aged 9 to 17 years on a test of speech intelligibility
722 in noise using sentence material with controlled word predictability. J Acoust Soc Am
723 66:651-653.
724 Faes L, Pinna GD, Porta A, Maestri R, Nollo G (2004) Surrogate data analysis for assessing
725 the significance of the coherence function. IEEE Trans Biomed Eng 51:1156-1166.
726 Ferstl EC, Walther K, Guthke T, von Cramon DY (2005) Assessment of story comprehension
727 deficits after brain damage. J Clin Exp Neuropsychol 27:367-384.

728 Giraud AL, Poeppel D (2012) Cortical oscillations and speech processing: emerging
729 computational principles and operations. Nat Neurosci 15:511-517.
730 Goswami U (2011) A temporal sampling framework for developmental dyslexia. Trends
731 Cogn Sci 15:3-10.
732 Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Parkkonen L,
733 Hamalainen MS (2014) MNE software for processing MEG and EEG data.
734 NeuroImage 86:446-460.
735 Gross J, Kujala J, Hamalainen M, Timmermann L, Schnitzler A, Salmelin R (2001) Dynamic
736 imaging of coherent sources: Studying neural interactions in the human brain. Proc
737 Natl Acad Sci U S A 98:694-699.
738 Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, Garrod S (2013) Speech
739 rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS Biol
740 11:e1001752.
741 Hall JW, 3rd, Buss E, Grose JH (2005) Informational masking release in children and adults.
742 J Acoust Soc Am 118:1605-1613.
743 Halliday DM, Rosenberg JR, Amjad AM, Breeze P, Conway BA, Farmer SF (1995) A
744 framework for the analysis of mixed time series/point process data--theory and
745 application to the study of physiological tremor, single motor unit discharges and
746 electromyograms. Prog Biophys Mol Biol 64:237-278.
747 Hämäläinen ML, Lin F-H, Mosher J (2010) Anatomically and functionally constrained
748 minimum-norm estimates. In: MEG-an introduction to methods (Press OU, ed), pp
749 186-215.
750 Hämäläinen MS, Ilmoniemi RJ (1994) Interpreting magnetic fields of the brain: minimum
751 norm estimates. Med. Biol. Eng. Comput. 32:35–42.

752 Hoen M, Meunier F, Grataloup C-L, Pellegrino F, Grimault N, Perrin F, Perrot X, Collet L
753 (2007) Phonetic and lexical interferences in informational masking during speech-in-
754 speech comprehension. Speech Commun 49:905-916.
755 Johnson CE (2000) Children's phoneme identification in reverberation and noise. J Speech
756 Lang Hear Res 43:144-157.
757 Karns CM, Isbell E, Giuliano RJ, Neville HJ (2015) Auditory attention in childhood and
758 adolescence: An event-related potential study of spatial selective attention to one of
759 two simultaneous stories. Dev Cogn Neurosci 13:53-67.
760 Keitel A, Gross J, Kayser C (2018) Perceptually relevant speech tracking in auditory and
761 motor cortex reflects distinct linguistic features. Plos Biol 16(3):e2004473.
762 Klatte M, Bergstrom K, Lachmann T (2013) Does noise affect learning? A short review on
763 noise effects on cognitive performance in children. Front Psychol 4:578.
764 Koskinen M, Seppa M (2014) Uncovering cortical MEG responses to listened audiobook
765 stories. NeuroImage 100:263-270.
766 Lakatos P, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE (2013) The
767 spectrotemporal filter mechanism of auditory selective attention. Neuron 77:750-761.
768 Leibold LJ, Neff DL (2007) Effects of masker-spectral variability and masker fringes in
769 children and adults. J Acoust Soc Am 121:3666-3676.
770 Liebenthal E, Desai RH, Humphries C, Sabri, Desai A (2014) The functional organization of
771 the left STS: a large scale meta-analysis of PET and fMRI studies of healthy adults.
772 Front Neurosci 8:289.
773 Lilliefors HW (1967) On the Kolmogorov-Smirnov test for normality with mean and variance
774 unknown. J Am Stat Assoc 62:399–402.
775 Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in
776 human auditory cortex. Neuron 54:1001-1010.

777 Mahmoudzadeh M, Dehaene-Lambertz G, Fournier M, Kongolo G, Goudjil S, Dubois J,
778 Grebe R, Wallois F (2013) Syllabic discrimination in premature human infants prior to
779 complete formation of cortical layers. Proc Natl Acad Sci U S A 110:4846-4851.
780 Massaro DW (2017) Reading aloud to children: Benefits and implications for acquiring
781 literacy before schooling begins. Am J Psychol 130:63-72.
782 McCoy SL, Tun PA, Cox LC, Colangelo M, Stewart RA, Wingfield A (2005) Hearing loss
783 and perceptual effort: downstream effects on older adults' memory for speech. Q J Exp
784 Psychol A 58:22-33.
785 Mesgarani N, Chang EF (2012) Selective cortical representation of attended speaker in multi-
786 talker speech perception. Nature 485:233-236.
787 Molinaro N, Lizarazu M (2017) Delta(but not theta)-band cortical entrainment involves
788 speech-specific processing. Eur J Neurosci.
789 Molinaro N, Lizarazu M, Lallier M, Bourguignon M, Carreiras M (2016) Out-of-synchrony
790 speech entrainment in developmental dyslexia. Hum Brain Mapp 37:2767-2783.
791 Moore DR, Ferguson MA, Edmondson-Jones AM, Ratib S, Riley A (2010) Nature of auditory
792 processing disorder in children. Pediatrics 126:e382-390.
793 Muzik O, Chugani DC, Juhasz C, Shen C, Chugani HT (2000) Statistical parametric mapping:
794 assessment of application in children. NeuroImage 12:538-549.
795 Neuman AC, Wroblewski M, Hajicek J, Rubinstein A (2010) Combined effects of noise and
796 reverberation on speech recognition performance of normal-hearing children and
797 adults. Ear Hear 31:336-344.
798 Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional
799 neuroimaging: a primer with examples. Hum Brain Mapp 15:1-25.
800 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.

Poelmans H, Luts H, Vandermosten M, Boets B, Ghesquiere 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 (5):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.

825 Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of
826 sensory selection. Trends Neurosci 32:9-18.
827 Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillations and
828 visual amplification of speech. Trends Cogn Sci 12:106-113.
829 Simon JZ (2014) The encoding of auditory objects in auditory cortex: Insights from
830 magnetoencephalography. Int J Psychophysiol 95(2):184-90. .
831 Simpson SA, Cooke M (2005) Consonant identification in N-talker babble is a nonmonotonic
832 function of N. J Acoust Soc Am 118:2775-2778.
833 Sussman E, Steinschneider M (2009) Attention effects on auditory scene analysis in children.
834 Neuropsychologia 47:771-785.
835 Talarico M, Abdilla G, Aliferis M, Balazic I, Giaprakis I, Stefanakis T, Foenander K,
836 Grayden DB, Paolini AG (2007) Effect of age and cognition on childhood speech in
837 noise perception abilities. Audiol Neurootol 12:13-19.
838 Taulu S, Simola J (2006) Spatiotemporal signal space separation method for rejecting nearby
839 interference in MEG measurements. Phys Med Biol 51:1759-1768.
840 Taulu S, Simola J, Kajola M (2005) Applications of the signal space separation method. IEEE
841 Trans Signal Process 53:3359-3372.
842 Teng X, Tian X, Doelling K, Poeppel D (2017) Theta band oscillations reflect more than
843 entrainment: behavioral and neural evidence demonstrates an active chunking process.
844 Eur J Neurosci
845 Thompson EC, Woodruff Carr K, White-Schwoch T, Otto-Meyer S, Kraus N (2017)
846 Individual differences in speech-in-noise perception parallel neural speech processing
847 and attention in preschoolers. Hear Res 344:148-157.
848 Vander Ghinst M, Bourguignon M, Op de Beeck M, Wens V, Marty B, Hassid S, Choufani
849 G, Jousmäki V, Hari R, Van Bogaert P, Goldman S, De Tiege 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.

Wightman FL, Callahan MR, Lutfi RA, Kistler DJ, Oh E (2003) Children's detection of pure-tone signals: informational masking with contralateral maskers. J Acoust Soc Am 113:3297-3305.

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.

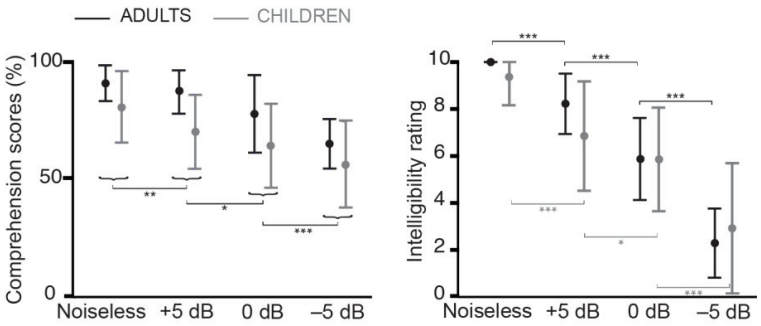


Figure 1: Comprehension scores (left graph) and intelligibility ratings (right graph) in adults (black) and children (grey). Bars indicate the mean and range. Comprehension scores are reported in percentage of questions (16 for adults and 8 for children) answered correctly, and intelligibility ratings 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 (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

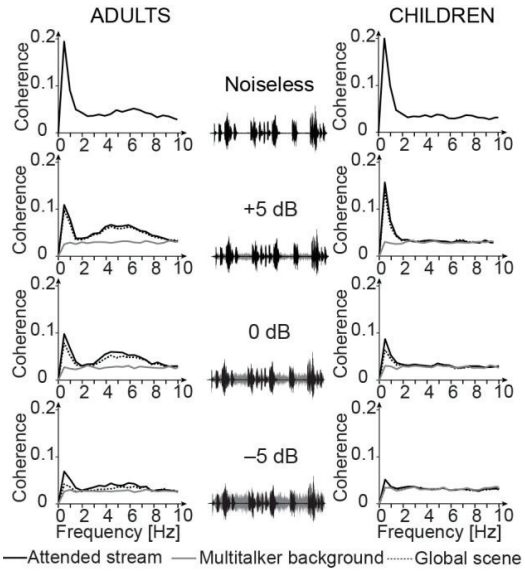
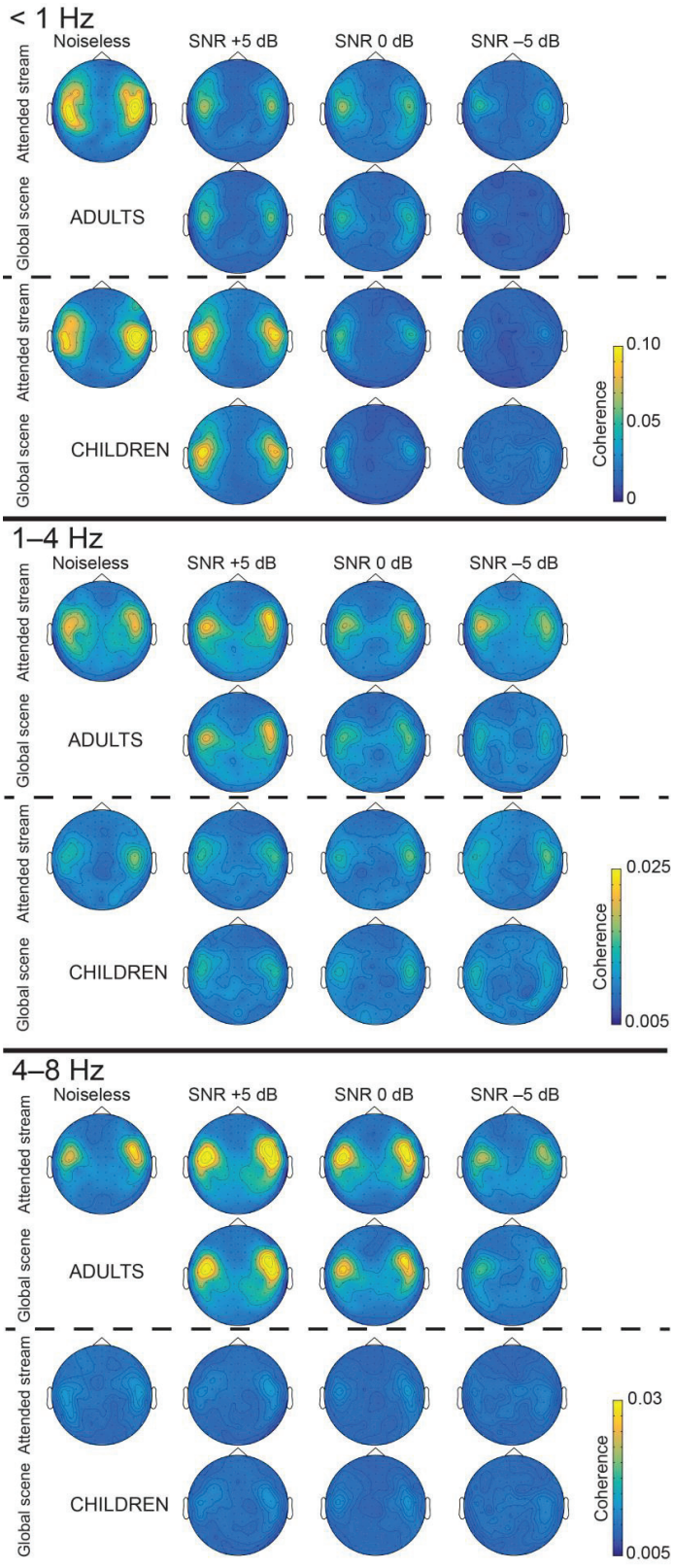
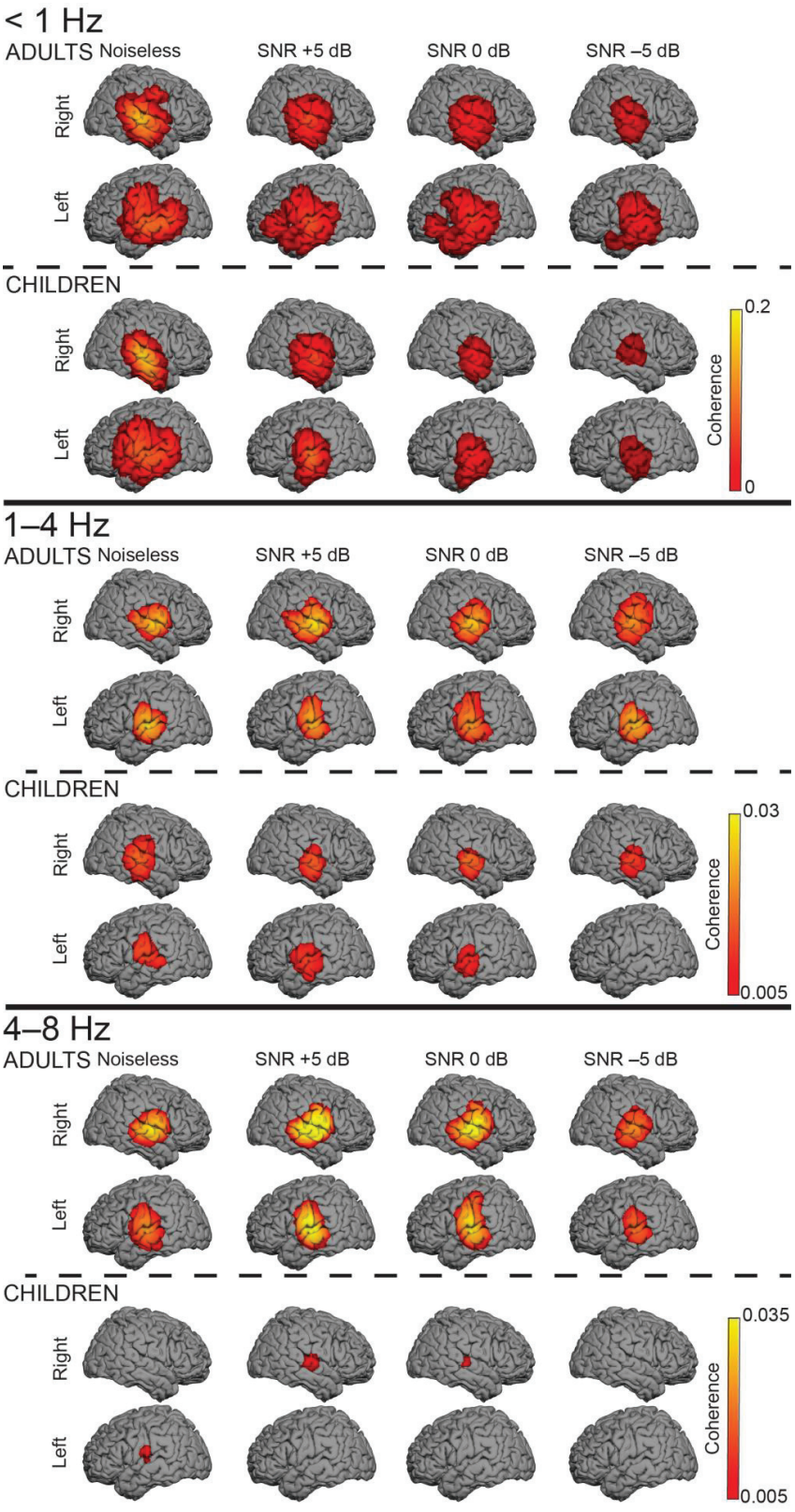


Figure 2: Spectra of cortical tracking of speech in the 4 speech-in-noise conditions and corresponding sound excerpts. Group-averaged coherence spectra are shown separately for adults (left column) and children (right column), and when estimated with the temporal envelope of the different components 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 (12 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.



899

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 Hz, 1–4 Hz, 4–8 Hz), age group (adults, children) component of the auditory scene (*Attended stream*, *Global scene*), and SNR condition (*Noiseless*, *+5 dB*, *0 dB*,*–5 dB*).



909

910 **Figure 4:** Cortical tracking of the attended speech stream at < 1 Hz, 1–4 Hz, and 4–8 Hz. The
911 group-level coherence maps (n = 20 in each map) were masked statistically above the
912 significance level (non-parametric permutation statistics). One source distribution is displayed
913 for each possible combination of age group (adults, top panel; children, bottom panel) and
914 SNR condition (from left to right, Noiseless, +5 dB, 0 dB, and –5 dB).
915
916
917

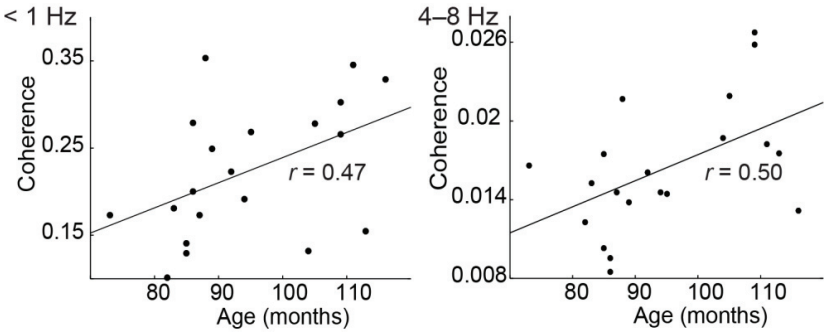


Figure 5: Correlation between children's age and their speech tracking values, depicted by the peak group-level coherence, in the noiseless condition at the right superior temporal sulcus at < 1 Hz and supratemporal auditory cortex at 4–8 Hz.

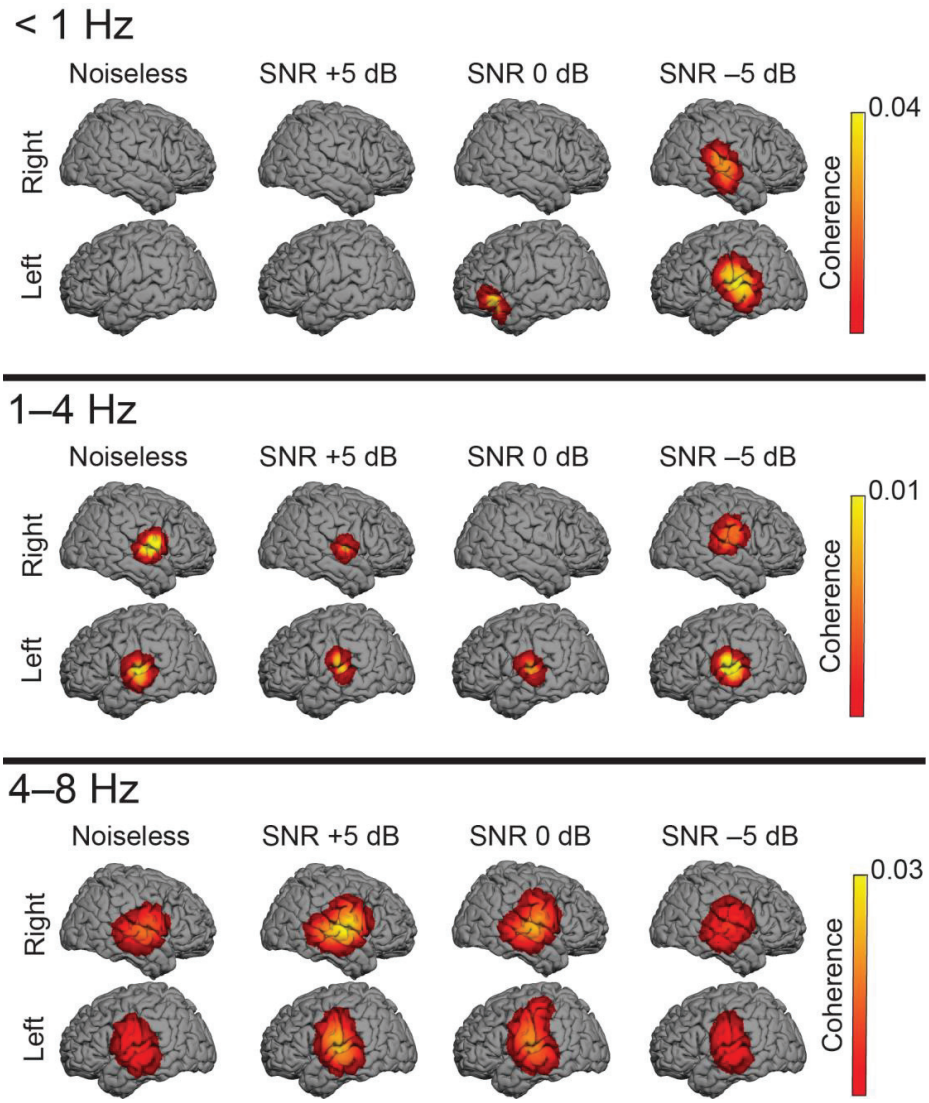
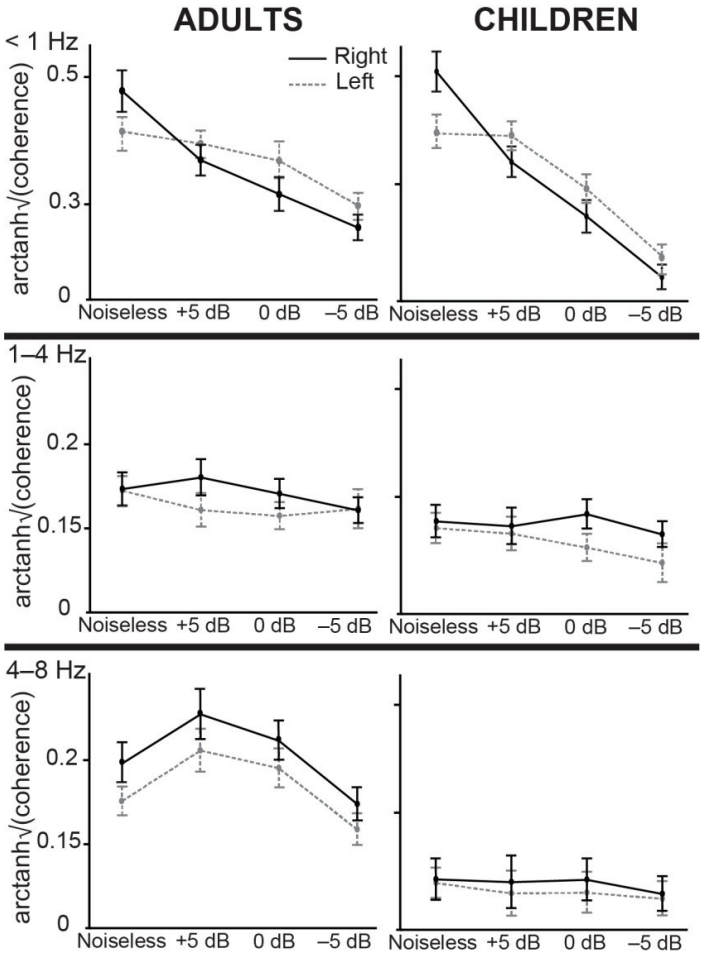


Figure 6: Contrasts between cortical tracking of the attended speech (Coh_{att}) in adults versus children ($Coh_{att,adults} - Coh_{att,children}$) at < 1 Hz, 1–4 Hz, and 4–8 Hz in all SNR conditions (*Noiseless*, *+5 dB*, *0 dB*, and *–5 dB*). The group-level difference coherence maps (n = 20 in each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

934



935

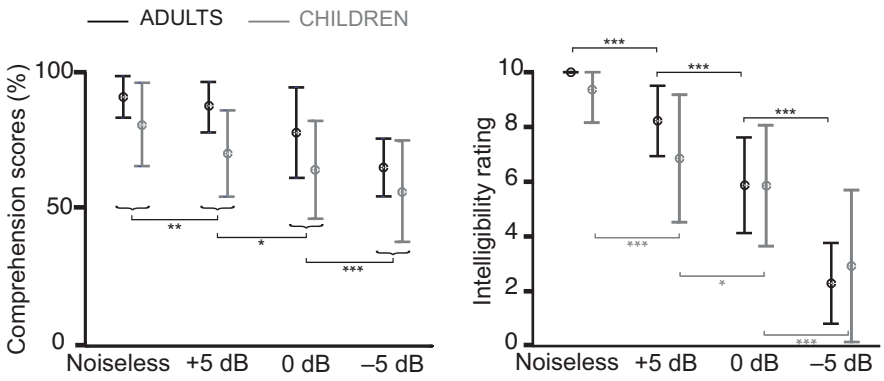
936

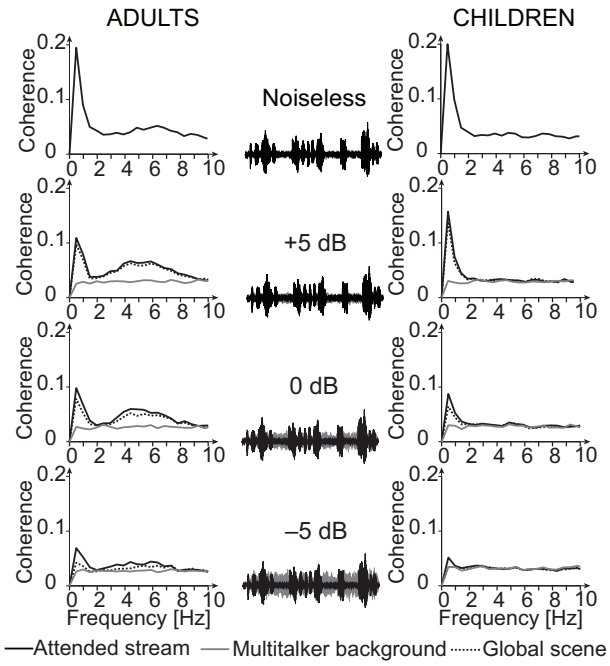
937 **Figure 7:** Illustration of the interaction effects identified on speech brain tracking of the
938 attend speech stream in adults (left column) and children (right column), at < 1 Hz, 1–4 Hz
939 and 4–8 Hz. Bars indicate the mean and standard error on the mean of *f*-transformed (with
940 $f(\cdot) = \text{arctanh}(\sqrt{\cdot})$) coherence values at the right (solid lines) and left (dashed lines) cortical
941 area of peak group-level coherence (superior temporal sulcus at <1 Hz, and supratemporal
942 auditory cortex at 1–4 Hz and 4–8 Hz).

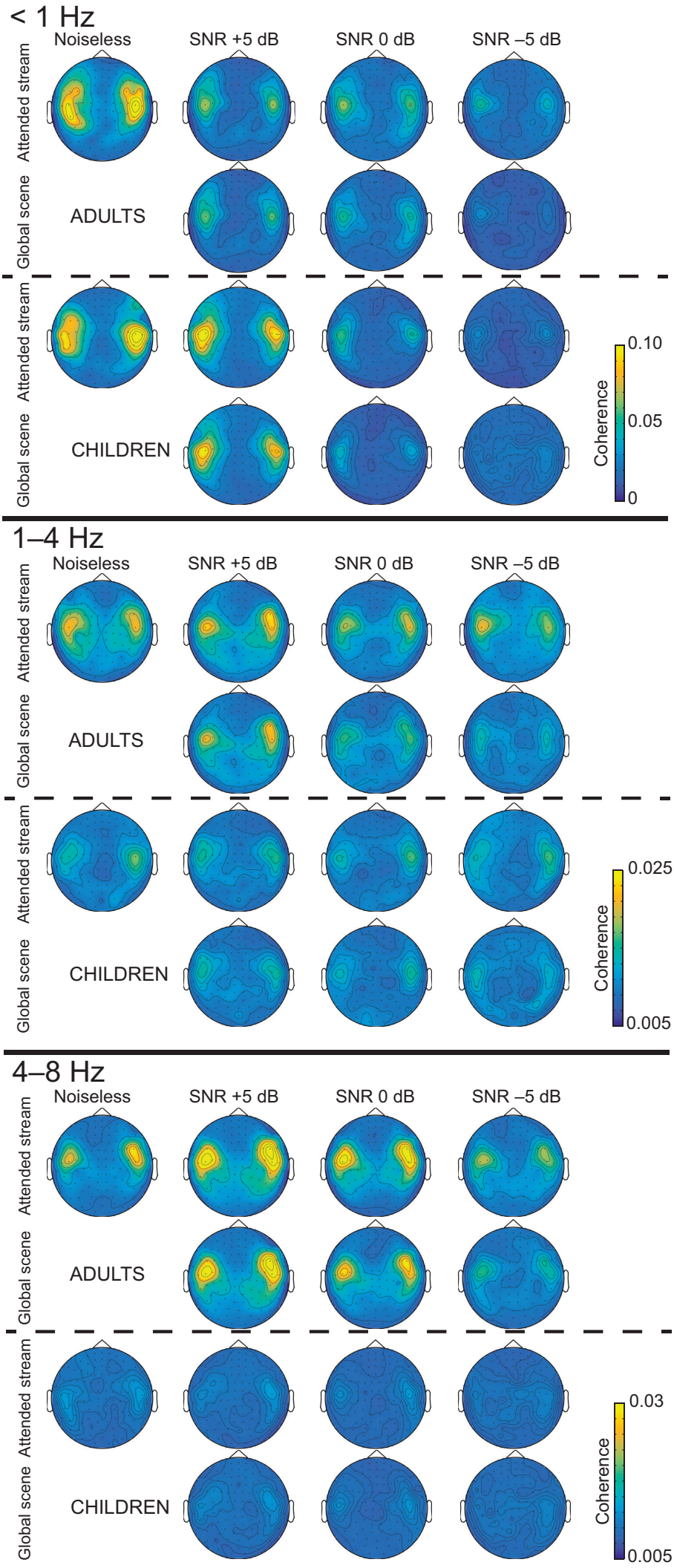
943

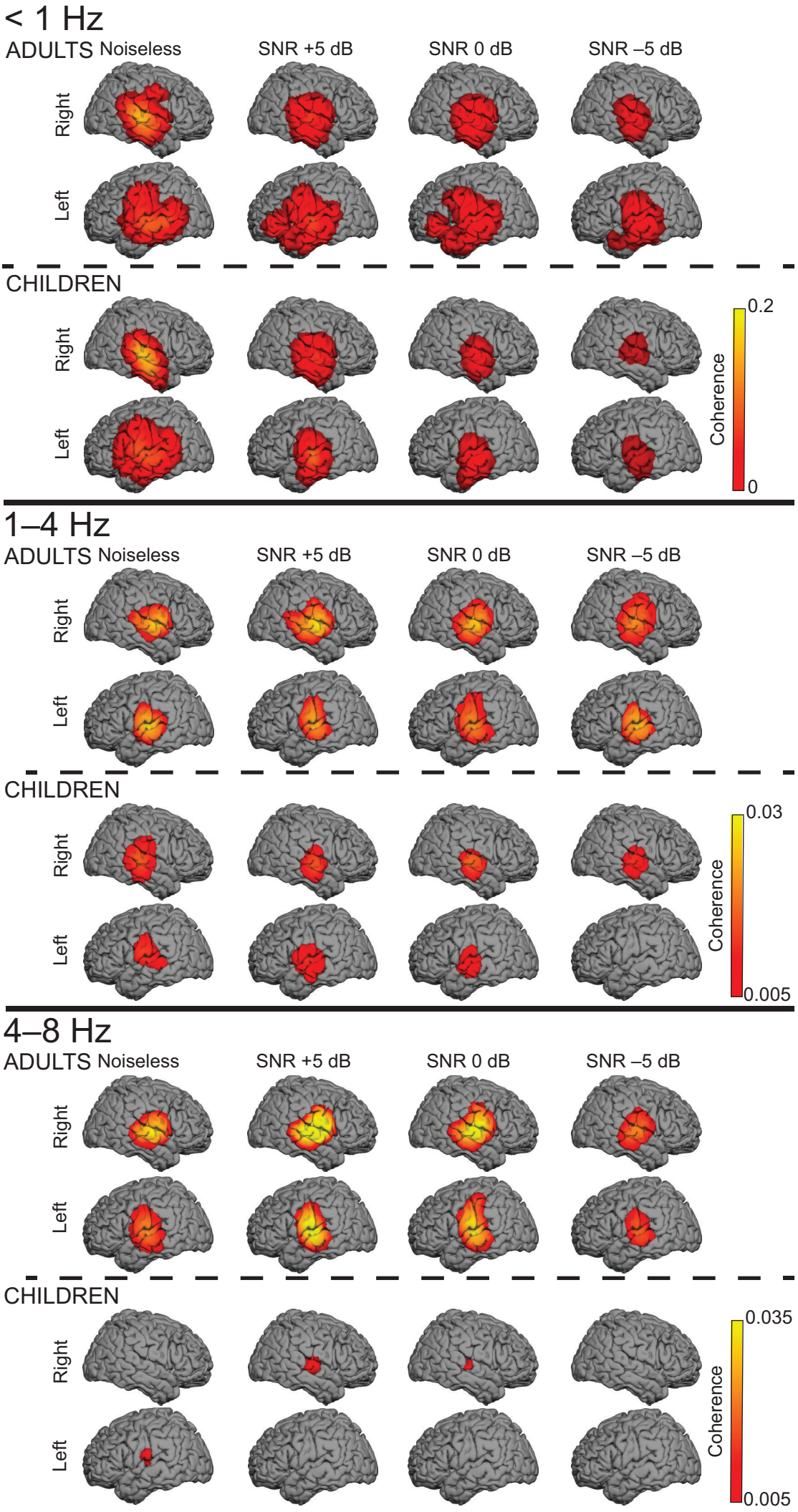
Condition	*Attended stream*		*Multitalker background*		*Global scene*	
	Adult	Children	Adult	Children	Adult	Children
	<1 Hz					
Noiseless	20	20				
+5 dB	20	19	2	1	20	19
0 dB	20	17	1	1	18	14
–5 dB	16	6	0	4	9	1
	1–4 Hz					
Noiseless	17	10				
+5 dB	16	8	2	1	14	9
0 dB	16	8	0	3	13	8
–5 dB	10	7	3	1	7	6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

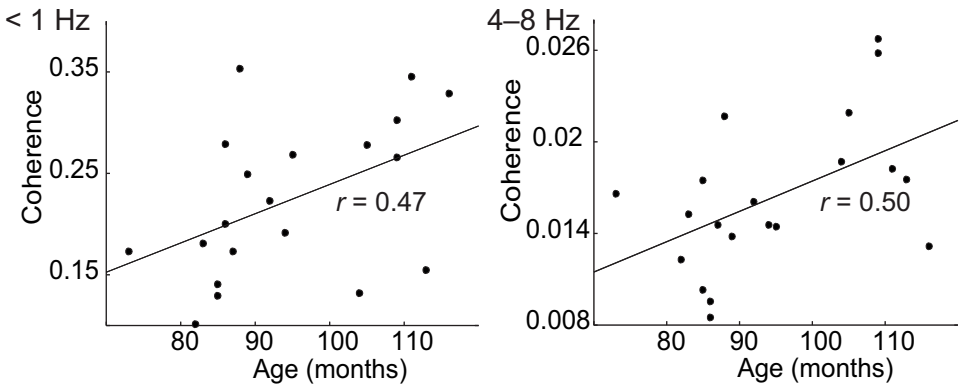
Table 1: Number of adults and children showing statistically significant coherence (surrogate-data-based statistics) in at least one sensor for each audio signal, condition, and frequency band of interest.

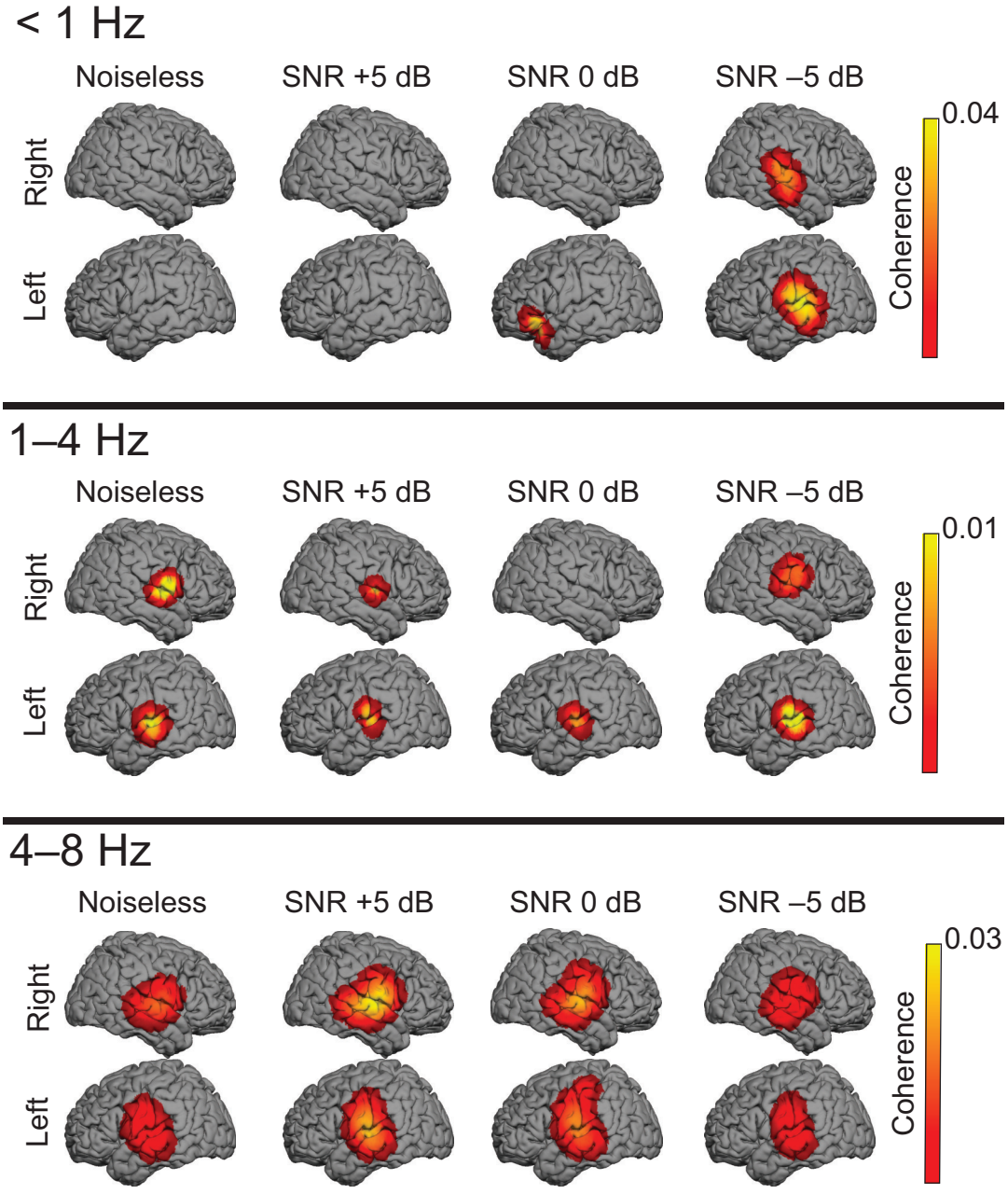


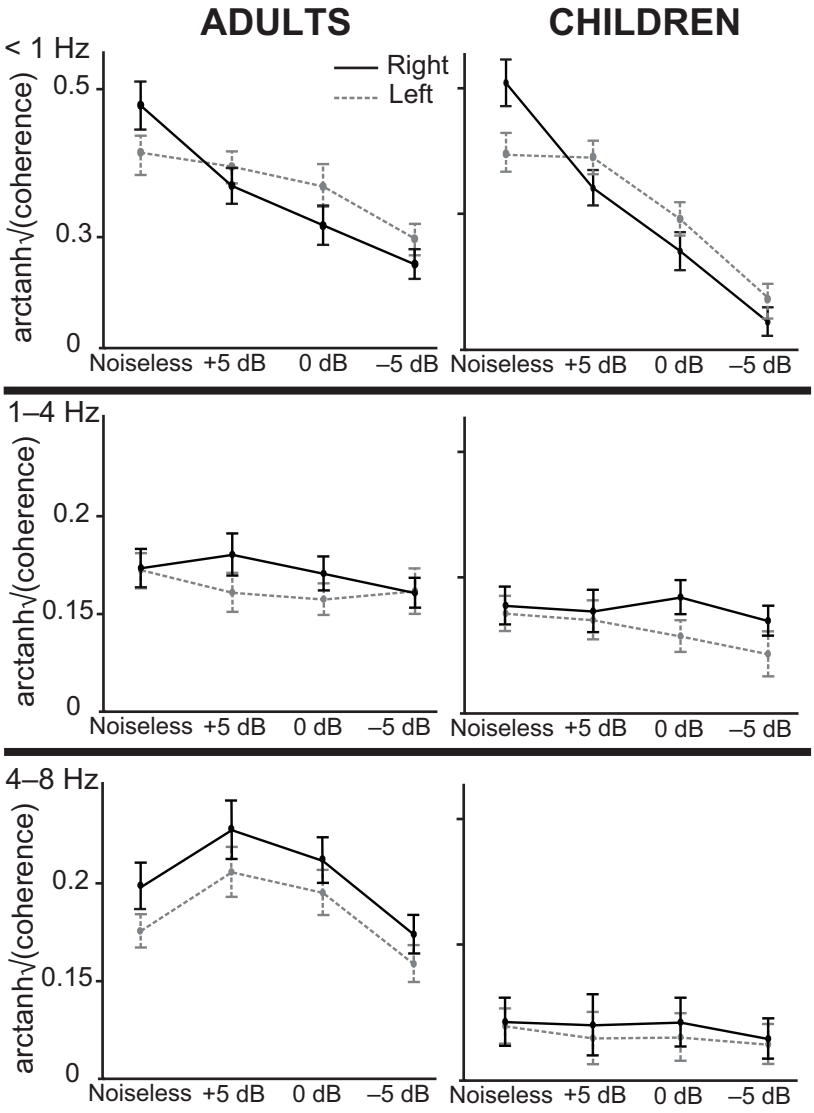












Condition	*Attended stream*		*Multitalker background*		*Global scene*	
	Adult	Children	Adult	Children	Adult	Children
	<1 Hz					
Noiseless	20	20				
+5 dB	20	19	2	1	20	19
0 dB	20	17	1	1	18	14
–5 dB	16	6	0	4	9	1
	1–4 Hz					
Noiseless	17	10				
+5 dB	16	8	2	1	14	9
0 dB	16	8	0	3	13	8
–5 dB	10	7	3	1	7	6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