NeuroImage 265 (2023) 119770

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Cortical tracking of lexical speech units in a multi-talker background is immature in school-aged children

Maxime Niesen[a,b,*], Mathieu Bourguignon[a,c,d], Julie Bertels[a,e], Marc Vander Ghinst[a,b], Vincent Wens[a,f], Serge Goldman[a,g], Xavier De Tiège[a,f]

[a] Université libre de Bruxelles (ULB), UNI — ULB Neurosciences Institute, Laboratoire de Neuroanatomie et de Neuroimagerie translationnelles (LN2T), 1070 Brussels, Belgium

[b] Université libre de Bruxelles (ULB), Hôpital Universitaire de Bruxelles (HUB), CUB Hôpital Erasme, Department of Otorhinolaryngology, 1070 Brussels, Belgium

[c] Université libre de Bruxelles (ULB), UNI–ULB Neuroscience Institute, Laboratory of Neurophysiology and Movement Biomechanics, 1070 Brussels, Belgium.

[d] BCBL, Basque Center on Cognition, Brain and Language, 20009 San Sebastian, Spain.

[e] Université libre de Bruxelles (ULB), UNI–ULB Neuroscience Institute, Cognition and Computation group, ULBabyLab - Consciousness, Brussels, Belgium

[f] Université libre de Bruxelles (ULB), Hôpital Universitaire de Bruxelles (HUB), CUB Hôpital Erasme, Department of translational Neuroimaging, 1070 Brussels, Belgium.

[g] Université libre de Bruxelles (ULB), Hôpital Universitaire de Bruxelles (HUB), CUB Hôpital Erasme, Department of Nuclear Medicine, 1070 Brussels, Belgium.

A R T I C L E   I N F O

Keywords:
Children
Development
Speech in noise
Magnetoencephalography
Hierarchical linguistic units
Multi-talker background
Cocktail party

A B S T R A C T

Children have more difficulty perceiving speech in noise than adults. Whether this difficulty relates to an immature processing of prosodic or linguistic elements of the attended speech is still unclear. To address the impact of noise on linguistic processing *per se*, we assessed how babble noise impacts the cortical tracking of intelligible speech devoid of prosody in school-aged children and adults.

Twenty adults and twenty children (7-9 years) listened to synthesized French monosyllabic words presented at 2.5 Hz, either randomly or in 4-word hierarchical structures wherein 2 words formed a phrase at 1.25 Hz, and 2 phrases formed a sentence at 0.625 Hz, with or without babble noise. Neuromagnetic responses to words, phrases and sentences were identified and source-localized.

Children and adults displayed significant cortical tracking of words in all conditions, and of phrases and sentences only when words formed meaningful sentences. In children compared with adults, the cortical tracking was lower for all linguistic units in conditions without noise. In the presence of noise, the cortical tracking was similarly reduced for sentence units in both groups, but remained stable for phrase units. Critically, when there was noise, adults increased the cortical tracking of monosyllabic words in the inferior frontal gyri and supratemporal auditory cortices but children did not.

This study demonstrates that the difficulties of school-aged children in understanding speech in a multi-talker background might be partly due to an immature tracking of lexical but not supra-lexical linguistic units.

1. Introduction

In daily life, humans tend to gather in places often unfavorable to verbal communication due to noisy backgrounds. In such situations, successful conversation is challenging since listeners have to tune in to the speaker's voice, while tuning out the noisy auditory scene. This difficult task is especially challenging for children aged <10 years, who typically have lower speech in noise (SiN) processing abilities than adolescents or adults (Elliott, 1979; Hall et al., 2002). Whether these lower abilities relate to a higher impact of noise on the neural processing of speech linguistic or paralinguistic (e.g., prosody, pitch, fluency pauses) information in children is still unsettled.

One way to understand how the human brain processes SiN is to study how cortical activity tracks the fluctuations of natural connected speech in noisy backgrounds. Such cortical tracking of speech (CTS) typically occurs at frequencies matching with hierarchical temporal linguistic units (i.e., syllables, words, phrases/sentences) and with paralinguistic information such as prosodic cues (Giraud and Poeppel, 2012; Gross et al., 2013; Ding and Simon, 2014). It is considered to subserve the incremental neural grouping of abstract linguistic units to promote

Abbreviations: CTS, cortical tracking of speech; SiN, speech-in-noise; MEG, magnetoencephalography; SPL, sound pressure level; MRI, magnetic resonance images; SNR, signal-to-noise ratio; STAC, supra-temporal auditory cortex; IFG, inferior frontal gyrus.

* Corresponding author: Université libre de Bruxelles (ULB), UNI — ULB Neurosciences Institute, Laboratoire de Neuroanatomie et de Neuroimagerie translationnelles (LN2T), 1070 Brussels, Belgium

E-mail address: maxime.niesen@ulb.be (M. Niesen).

https://doi.org/10.1016/j.neuroimage.2022.119770.
Received 28 March 2022; Received in revised form 9 November 2022; Accepted 23 November 2022
Available online 1 December 2022.
1053-8119/

subsequent speech recognition (Ding et al., 2016). In quiet background, CTS is observed in school-aged children aged <10 years but is reduced at the syllabic level compared with adults (Vander Ghinst et al., 2019; Bertels et al., 2022). In multi-talker situations, their auditory system selectively tracks the attended speech stream rather than the global auditory scene (Vander Ghinst et al. 2019). Still, CTS in children is more compromised by increasing noise level for words and phrases/sentences compared with CTS in adults. Furthermore, syllabic CTS in children does not increase in babble noise as observed in adults (Vander Ghinst et al., 2019; Bertels et al., 2022).

Still unclear is whether the higher impact of noise on children's CTS relates to increased alterations in the cortical tracking of the attended hierarchical linguistic units, or of the paralinguistic information such as prosody. Indeed, previous studies used natural connected speech masked by babble noise, yet natural connected speech comprises both hierarchical linguistic and paralinguistic information that temporally correlate (Yamashita, 2013). Previous behavioral and electrophysiological studies demonstrated ongoing maturation of syntactic and semantic processes until late childhood, which might hinder the cortical tracking of linguistic units per se in adverse auditory scenes in school-aged children aged <10 years (Elliott 1979, Hall et al. 2002, Demanez et al. 2003, Hahne et al. 2004, Mannel et al. 2013). One way to specifically address the impact of noise on the neural processing of speech linguistic units is to eliminate prosodic cues from the attended speech stream. In quiet environments, such an approach has already evidenced successful cortical tracking of hierarchical linguistic units in adults (Yamashita, 2013; Ding et al., 2016, 2017). In this context, the adults' brain internally groups small linguistic units (words) into larger linguistic units (phrases and sentences), based on grammatical knowledge only (Ding et al., 2016, 2017).

In the present study, we used speech devoid of prosody to contrast grammar-based CTS between school-aged children aged <10 years and adults, and between quiet and noisy conditions. To this end, we quantified using magnetoencephalography (MEG) the cortical tracking of isochronous monosyllabic words either forming or not phrases/sentences in the absence or presence of babble noise. As grammar-based CTS requires increased attention (Makov et al., 2017; Ding et al., 2018) and becomes less accurate when speech intelligibility decreases (Blanco-Elorrieta et al., 2020; Meng et al., 2021), we hypothesized that babble noise would impede the grammar-based internal grouping of monosyllabic words into phrases/sentences in both children and adults. Furthermore, as syntactic and semantic processing are developing until late childhood (Nuñez et al., 2011; Skeide and Friederici, 2016) and as children have lower SiN abilities than adults, we also hypothesized that grammar-based CTS would be reduced by babble noise to a greater extent in children compared with adults. We also hypothesized that the cortical tracking of small linguistic units such as monosyllabic words would not be enhanced by babble noise in children as compared with adults, as previously shown for natural connected speech (Vander Ghinst et al., 2019; Bertels et al., 2022). Finally, as the cortical tracking of natural connected speech is dominant in the right hemisphere and is mainly driven by prosodic cues (Friederici, 2002; Bourguignon et al., 2013), we hypothesized that the CTS in the absence of prosody would become dominant in the left hemisphere.

2. Methods

2.1. Subjects

Twenty healthy adults (mean age: 24 years, age range: 20–29 years, 11 females) and twenty healthy children (mean age: 8 years, age range: 7–9 years, 9 females) took part in this study. All subjects were native French speakers and right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971) (adults score range: 50–95, mean score: 75; children score range: 50–95, mean score: 73.5; $t_{38} = 0.30$, $p = 0.77$). They all had any prior history of neurological, psychiatric or otologic disorders and had normal hearing according to pure tone audiometry (i.e., normal hearing thresholds: between 0–20 dB HL for 125 Hz to 8000 Hz). Using three separate subtests (a speech audiometry, a SiN audiometry and a dichotic test) of a validated and standardized central auditory battery in French (Demanez et al., 2003), we have ensured that participants had normal dichotic and SiN perception for their age.

This study was approved by the Ethics Committee of the CUB Hôpital Erasme. All subjects (and their legal representatives for children) gave written informed consent prior to their inclusion in the study.

2.2. Stimuli

Fig. 1 illustrates the stimuli used in the present study. They were adapted from those described in Ding et al. (2016) to the French language.

The auditory stimuli were 238 different monosyllabic French words. They were synthesized using the MacinTalk Synthesizer (male voice, Thomas, in macOs Sierra 10.12.6) and were adjusted to the same intensity and the same duration of 400 ms by truncation (without alteration of word identity) or silence padding symmetrically on both sides (original mean duration of 375 ± 70 ms, range 146–525 ms). To introduce a fade-in and a fade-out, the extremities of each word were multiplied by a 25-ms squared-sine ramp signal.

The stimuli were used to build blocks of 40 words in 4 different conditions described hereafter (*Meaningful, Meaningful*$_{noise}$*, Scrambled,* and *Scrambled*$_{noise}$). Word acoustic waveforms were concatenated without any additional acoustic gap between words. Considering the duration of each word (i.e., 400 ms), the word rate was 2.5 Hz (f_{word}). Twenty-five different blocks were built for each condition.

For the *Meaningful* condition, we constructed 250 different sentences composed of four monosyllabic words. All sentences shared the same hierarchical linguistic structures: determiner + noun + verb + adjective/adverb. In that setting, the phrase (i.e., determiner + noun and verb + adjective/adverb) rate was 1.25 Hz (f_{phrase}), and the sentence (determiner + noun + verb + adjective/adverb) rate was 0.625 Hz (f_{sentence}) (**Fig. 1a**). Critically, as in Ding et al. (2016), the linguistic units could only be extracted using grammar-based knowledge, and not prosodic cues. Indeed, the sound envelope featured fluctuations at f_{word} but not at f_{phrase} or f_{sentence} due to the absence of prosodic cues for phrase/sentence boundaries (**Fig. 1b**).

The *Scrambled* condition was created by randomly shuffling the order of the monosyllabic words used in the *Meaningful* condition, resulting in meaningless strings of words presented at 2.5 Hz.

Two more listening conditions, *Meaningful*$_{noise}$ and *Scrambled*$_{noise}$, were created by adding a multitalker background noise to the *Meaningful* and *Scrambled* conditions at a signal-to-noise ratio (SNR) of 0 dB (**Fig. 1c**). This SNR was chosen because it is typically encountered in multi-talker situations (Bronkhorst, 2015). The background noise was the same as used in previous studies by our group (Vander Ghinst et al., 2016, 2019). Briefly, it consisted of a mix of 6 native French speakers' speech (3 females and 3 males). This configuration was chosen because it introduces interference at phonetic and lexical levels (Simpson and Cooke, 2005; Hoen et al., 2007), with good balance between these two types of interference (Simpson and Cooke, 2005; Hoen et al., 2007).

2.3. Experimental paradigm

During MEG recordings, subjects sat comfortably in the MEG chair with their arms laying on a table. They were asked to gaze at a cross on the wall in front of them and not to move. After a 5-min rest condition (i.e., stimulation-free), subjects underwent five listening sessions, each lasting ~6 min. Listening sessions consisted of 20 blocks randomly selected among the four different conditions (i.e., *Meaningful, Scrambled, Meaningful*$_{noise}$*, Scrambled*$_{noise}$) with the rule that two consecutive blocks cannot be of the same condition. The order of blocks was randomized across conditions and blocks were separated by a silent break of 3 s.

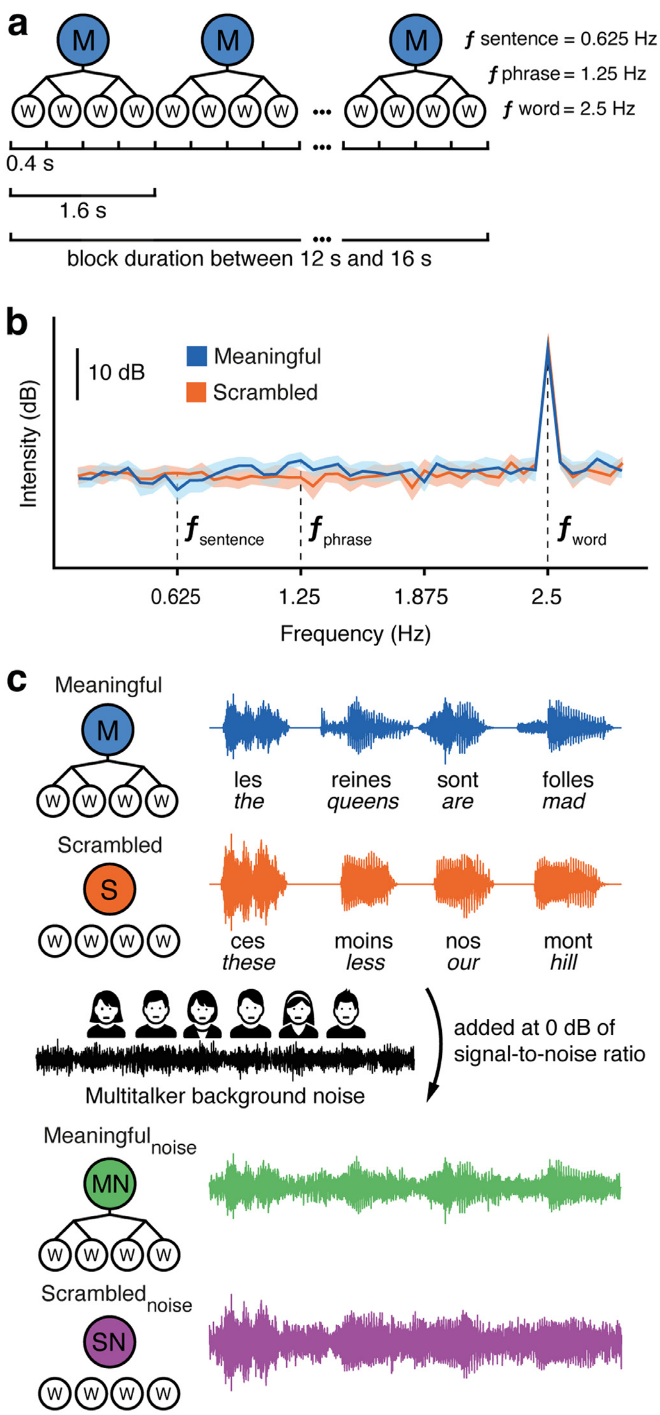

Fig. 1. Experimental stimuli. (a) Sequence of monosyllabic words (presentation rate 2.5 Hz) forming phrases (1.25 Hz, determiner + noun and verb + adjective/adverb) and sentences (0.625 Hz, determiner + noun + verb + adjective/adverb). (b) Spectrum of stimulus intensity disclosing a clear peak at the monosyllabic word rate but not at the phrase or sentence rates. (c) Time-course of stimuli for each condition.

Auditory stimuli were played using VLC Media Player (VideoLAN Project, version 2.2.6, GNU General Public License) and were delivered through a MEG-compatible 60 × 60 cm^2 high-quality flat panel loudspeaker (Panphonics SSH sound shower, Panphonics Oy) placed ~2.5 m in front of the subjects. The average sound intensity was set to 60 dB sound pressure level (SPL) as assessed by a sound level meter (Sphynx Audio System). The audio signal was fed to a miscellaneous channel of the MEG system and sampled synchronously with MEG signals. This signal was later used for precise synchronization with the audio material and to determine the onsets and offsets of word sequences.

To ensure subjects maintained their attentional focus on the auditory stimuli, they were asked to repeat the last word they heard at the end of each block during the acoustic gap (a behavioral task henceforth referred to as "word identification"). To avoid a possible prediction of the word to be repeated based on the preceding linguistic units, we randomly truncated each block by a number of words in between 0 and 8, leading to blocks of 32–40 words. The last word to be repeated could therefore be of any class (i.e., determiner, noun, verb, adjective/adverb). After the MEG recordings, subjects were asked to rate the intelligibility of a randomly chosen block of each condition on a visual analog scale ranging from 0 to 10 (0, totally unintelligible; 10, perfectly intelligible).

2.4. Data acquisition

Neuromagnetic signals were recorded at the CUB Hôpital Erasme with a whole-scalp-covering MEG system (Triux, MEGIN, Croton Healthcare, Finland) installed in a lightweight magnetically shielded room (Maxshield, MEGIN, Croton Healthcare, Finland; see De Tiège et al., 2008 for more details). The MEG system comprised 102 sensor triplets, each consisting of one magnetometer and two orthogonal planar gradiometers. The recording bandpass filter was 0.1–330 Hz and the data were sampled at 1 kHz. We used four head-tracking coils to continuously monitor subjects' head position inside the MEG helmet. We digitized with an electromagnetic tracker (Fastrak, Polhemus) the location of the coils and at least 250 head-surface points (on scalp, nose, and face) with respect to anatomical fiducials.

Subjects' high-resolution 3D-T1 weighted cerebral magnetic resonance images (MRI) were acquired on a 1.5 T MRI (Intera, Philips) with a slice thickness of 1 mm and in-plane resolution of 1 × 1 mm. When motion artifacts were visible on the MRI images at the time of acquisition, which often happens with children, images were discarded and another scan was performed after participants' informed consent until good quality images could be obtained.

2.5. Data preprocessing

Continuous MEG data were first preprocessed off-line using the temporal signal space separation method implemented in MaxFilter software (MaxFilter, MEGIN, Croton Healthcare, Finland; correlation limit 0.9, segment length 20 s) to suppress external interferences and to correct for head movements (Taulu et al., 2005; Taulu and Simola, 2006). Besides, head movement parameters, consisting of a set of 3 translations and 3 rotation angles, were saved for further analyses.

Cardiac, ocular and remaining system artifacts were further eliminated from MEG data separately, using an independent component analysis of band-passed (0.1–25 Hz) signals (Vigário et al., 2000; FastICA v2.5, http://www.cis.hut.fi/projects/ica/fastica, with dimension reduction to 30 components, symmetric approach, and cubic nonlinearity contrast). Artifactual components were identified by visual inspection and corresponding MEG signals reconstructed by means of the mixing matrix were subtracted from the full-rank and full-band data. Across subjects and conditions, the number of components rejected was 2.5 ± 0.5 (mean ± SD) in the adult group and 3.8 ± 1.0 in the children group (t_{38} = 5.06, p_{corr} < 0.0001).

The resulting data were then filtered through 0.1–40 Hz using a zero phase-lag filter implemented in the Fourier domain. In brief, the data were Fourier-transformed, then multiplied by a window raising from 0 (at 0.05 Hz) to 1 at (0.15 Hz) following a squared-sine profile and ebbing much the same way from 1 (at 37.5 Hz) to 0 (at 42.5 Hz), and then inverse Fourier-transformed. MEG epochs were extracted from the 5^{th} word onset (to avoid the transient response to the acoustic onset of each block) to the 32^{nd} word offset (because of the random truncation of blocks). Epochs were considered contaminated by artifacts and removed from further analyses when their maximum MEG amplitude exceeded 5 pT in at least one magnetometer or 1 pT/cm in at least one gradiometer. The mean ± SD number of artifact-free epochs was 24.3 ± 2.0 out of 25 (across subjects and conditions) in the adult group and 22.2 ± 2.7 out of 25 in the children group. We performed an ANOVA on the number of rejected epochs to assess the effect of condition and the effect of age. The ANOVA revealed a significant effect of age ($F_{1,38}$ = 4.64, $p = 0.038$) but no significant effect of condition ($F_{3,114} = 0.76$, $p = 0.52$) nor interaction involving this factor ($F_{3,114} = 0.82$, $p = 0.49$). To avoid a possible methodological bias in our results due to differences between age groups in the number of epochs analyzed, we randomly discarded epochs in adults' data to equalize the number of epochs in both groups.

2.6. Head movements

For each participant and listening condition, the variance of head movement parameters was estimated and summed across the three spatial directions to obtain a single value for translations and rotations. Variance values were averaged across listening conditions and converted to standard deviation values using the square root.

2.7. Sensor-space data analyses

Retained epochs were Fourier-transformed (frequency resolution 0.089 Hz). For each subject, listening condition and sensor, amplitude spectra were obtained as the modulus (i.e., absolute value) of the averaged Fourier-transformed epochs. Note that because the modulus was taken after averaging Fourier coefficients, our derivation of amplitude spectra allowed for phase cancellation of activity not phase-locked with audio sequences. At each sensor triplet, we retained only the Euclidean norm of the amplitude across pairs of planar gradiometers. This approach allowed us to draw strong conclusions on the potential underlying cortical sources. For each subject, condition and sensor, SNR responses were computed as the ratio between the amplitude at each frequency bin and the average amplitude at the 10 surrounding frequency bins (5 on each side, excluding the immediately adjacent bins) (Lins et al. 1996; John and Picton 2000; Aiken and Picton 2008; Peykarjou et al. 2017; Barry-Anwar et al. 2018; Bertels et al. 2020). SNR values significantly above 1 at f_{word}, f_{phrase} or f_{sentence} would indicate specific cortical tracking of the corresponding linguistic units.

2.8. Source-space data analyses

Source reconstruction was used to estimate brain maps of SNR. For that, MEG and MRI coordinate systems were co-registered using the 3 anatomical fiducial points for initial estimation and the head-surface points for further manual refinement. Then, the individual MRIs were segmented using Freesurfer software (Martinos Center for Biomedical Imaging, Boston, MA, RRID:SCR_001847; (Reuter et al., 2012)), and a non-linear transformation from individual MRIs to the MNI brain was 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)). This transformation was used to map a homogeneous 5-mm grid sampling the MNI brain volume onto individual brain volumes. For each subject and grid point, the MEG forward model corresponding to three orthogonal current dipoles was computed using the one-layer Boundary Element Method implemented in the MNE software suite (Martinos Center 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. A Minimum-Norm Estimates inverse solution (Dale and Sereno, 1993) was then used to project sensor-level Fourier coefficients (averaged across epochs) into the source space. We followed the same approach as that used at the sensor level to estimate source-level SNR (source pairs taking the place of gradiometer pairs).

We further identified the coordinates of local maxima in group-averaged SNR maps. Such local maxima of SNR are sets of contiguous voxels displaying higher SNR values than all neighboring voxels. We only report statistically significant local maxima of SNR, disregarding the extent 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).

The significant local maxima were visualized on the MNI glass brain using the BrainNet viewer (Xia et al., 2013) (see **Fig. 5**).

2.9. Statistical analysis

2.9.1. Behavioral results

Because results of speech and SiN audiometry were not normally distributed as indicated by Shapiro-Wilk tests (both $p < 0.05$ for children and adults' results in silence), we performed a *Mann-Whitney U* test to compare and identify statistical differences between adults and children.

A three-way repeated-measures ANOVA was used to assess the effects of listening condition (within-subject factor; *Meaningful, Scrambled,* $Meaningful_{noise}$, $Scrambled_{noise}$), word position (within-subject factor; determiner, noun, verb, adjective/adverb) and of age group (between-subjects factor; children, adults) on word identification.

A two-way repeated-measures ANOVA was used to assess the effects of listening condition (within-subject factor; *Meaningful, Scrambled,* $Meaningful_{noise}$, $Scrambled_{noise}$) and of age group (between-subjects factor, children, adults) on the intelligibility rating.

Post hoc comparisons were performed with pairwise *t*-tests with Bonferroni adjustment for multiple comparisons.

2.9.2. Head movements

Head movement values, consisting of one standard deviation for translations and one for rotations, were compared between groups using an independent sample *t*-test.

2.9.3. Individual levels of SNR

A nonparametric permutation-like test, first described in Bertels et al. (2020), was used to estimate the statistical significance of the SNR for each participant and at f_{word}, f_{phrase} and f_{sentence} separately. The test sought for significant responses in all gradiometer pairs, with correction for multiple comparisons across them. Such a statistical test was chosen because it can support claims of statistical significance at each gradiometer pair separately, in contrast with, e.g., cluster-based permutation tests (Sassenhagen and Draschkow, 2019). In a nutshell, the statistical procedure trims the epochs to randomize the position of the linguistic units under assessment and hence to destroy the phase locking across epochs of possible responses specific to these units. The parameters of the test for phrase and sentence SNR were different from those for word SNR. We therefore present the procedure for phrase and sentence SNR, and highlight the modifications for word SNR.

To test the significance of the SNR at f_{phrase} and f_{sentence}, we first built a permutation distribution for that SNR, or more specifically, for the maximum SNR across gradiometer pairs. Elements of the permutation distribution were computed as the maximum SNR derived from epochs randomly trimmed by a duration corresponding to the $n = 0, 1, 2$ or 3 first words ($n \times 400$ ms) and 4–n last words. This procedure was repeated 1000 times. To match epoch length across permuted and genuine data, genuine SNR in each gradiometer pair was re-computed based on epochs in which either the first or last 1.6 s of data (corresponding to 4 words) was removed. The significance of the genuine response at each gradiometer pair was computed as the proportion of values in the permutation distribution that were above the observed genuine value. This test, being akin to a permutation test (Nichols and Holmes, 2002), is exact, and because the permutation distribution was built on maximum values across gradiometer pairs, it intrinsically deals with the multiple comparison issue.

To test the significance of SNR at f_{word}, a different trimming scheme was used. Word SNR was recomputed based on epochs in which either the first or last 400 ms of data (corresponding to 1 word) was removed; and to estimate the permutation distribution, epochs were randomly trimmed by a duration corresponding to the $n = 0$ or 1 first half-words ($n \times 200$ ms) and 2–n last half-words. The trimming scheme randomized the position of words within epochs, so that epochs either started at word onset or in the middle of a word.

The total number of instances in which significant SNR at f_{phrase} or $f_{sentence}$ was uncovered in scrambled conditions (40 subjects × 2 frequencies × 2 conditions) was compared with the number expected by chance. Under the null hypothesis of no tendency to display significant SNR, this number follows a binomial distribution with n = 160 experiments and success probability $p = 0.05$).

2.9.4. Local maxima of group-level SNR

The statistical significance of the local maxima of SNR observed in group-averaged maps for each age group, listening condition and frequency of interest was assessed with a non-parametric permutation test that intrinsically corrects for multiple spatial comparisons (Nichols and Holmes, 2002). This test was conducted to support claims of generalizability of our findings to the general population. First, participant and group-averaged *rest* maps of SNR were computed with MEG epochs randomly extracted from the rest condition. Group-averaged difference maps were obtained by subtracting *genuine* and *rest* group-averaged SNR maps. Under the null hypothesis that SNR maps are the same whatever the experimental condition, the labeling *genuine* or *rest* are exchangeable prior to difference map computation (Nichols and Holmes, 2002). To reject this hypothesis and to compute a significance level for the correctly labeled difference map, the sample distribution of the maximum of the difference map's absolute value within the entire brain was computed from a subset of 1000 permutations. The threshold at $p < 0.05$ was computed as the 95 percentile of the sample distribution (Nichols and Holmes, 2002). All supra-threshold local maxima of SNR were interpreted as indicative of brain regions showing statistically significant CTS and will be referred to as sources of CTS.

Permutation tests can be too conservative for voxels other than the one with the maximum observed statistic (Nichols and Holmes, 2002). For example, dominant SNR values in the right hemisphere could bias the permutation distribution and overshadow weaker SNR values in the left hemisphere, even if these were highly consistent across subjects. Therefore, the permutation test described above was conducted separately for left- and right-hemisphere voxels.

2.9.5. Effect of age and multi-talker background noise on SNR values

We used a linear mixed model (LMM) to compare the SNR of sources of CTS between adults and children with an additional factor of noise (*Meaningful* and *Meaningful$_{noise}$* conditions). The dependent variable was the maximum SNR value within a sphere of 10-mm radius around the maxima of the group-level SNR map averaged across age groups and conditions in order to limit potential bias coming from differences in source location. However, since all sources of CTS were bilateral, the factor hemisphere (left and right) was added to the analysis. The factors noise, age group and hemisphere were defined as fixed effects, and the variability between subjects was accounted for by modeling the participant ID as a random effect. We used an unstructured covariance matrix to account for the repeated measures. Residuals were tested and confirmed graphically for normality. p-values were calculated from F statistics of type III by using between-within approximation of degrees of freedom. Results at $p < 0.05$ were considered to be statistically significant for all tests. They were not corrected for multiple comparisons. Statistical tests were performed with SAS software version 9.04 (SAS Institute Inc, Cary, NC, USA). Separate LMM were run for word, phrase, and sentence SNR, and for the different local maxima. Post hoc comparisons were performed with pairwise t-tests with Bonferroni adjustment for multiple comparisons. When a significant effect was uncovered for at least one but not all local maxima, we used a t-test to assess to what extent the contrast underlying the effect was significantly different between local maxima.

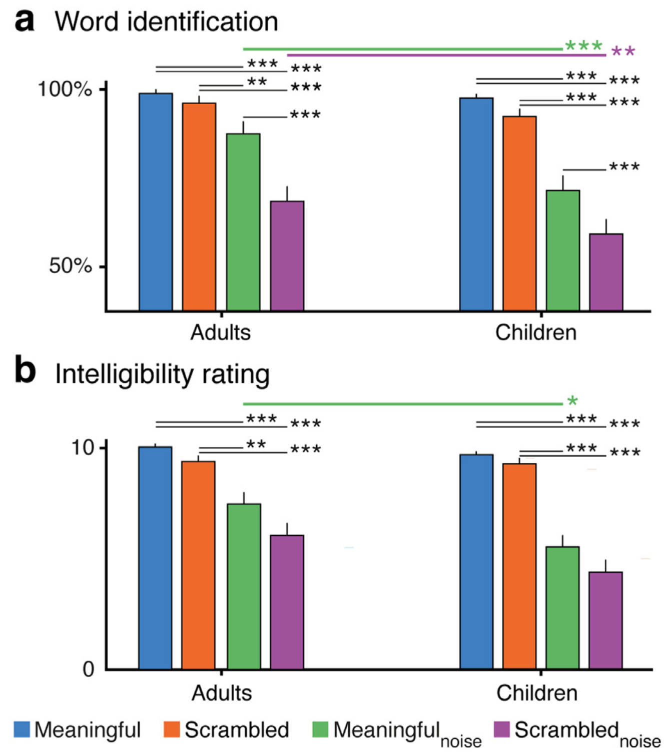

Fig. 2. Mean ± SD scores for the word identification (a) and the intelligibility rating (b) in both age groups. Asterisks indicate significant differences (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$).

2.9.6. Link between SNR values and behavioral results, children age and head movements

Based on our results, we assessed the Pearson correlation between (i) SNR values in both sources locations and behavioral results (identification score and speech audiometry in noise) for both groups and both hemispheres at each frequency of interest (ii) age in months in children and the difference of SNR (mean over both hemispheres) between *Meaningful$_{noise}$* and *Meaningful* conditions, (iii) age in months and head movement parameters, and (iv) head movement parameters (for translation and rotations) and SNR values in both sources locations for both groups and both hemisphere at each frequency of interest.

3. Results

3.1. Behavioral results

Speech audiometry in silence did not differ between adults and children (mean score ± SD: adults = 28.3 ± 0.9, children = 28.1 ± 1.1, $U = 232.5$, $p = 0.356$), but differed in noise (adults = 27.4 ± 1.2, children = 25.9 ± 2.0, $U = 284.5$, $p = 0.020$).

Similar results were observed for word identification (Fig. 2**a**). Indeed, the ANOVA performed on these scores revealed a significant effect of age group ($F_{1,38} = 39.9$, $p < 0.0001$), a significant effect of listening condition ($F_{3,114} = 200.8$, $p < 0.0001$) and a significant interaction between age group and listening condition ($F_{3,114} = 8.7$, $p < 0.0001$), but no effect of word class (determiner, noun, verb, adjective/adverb)

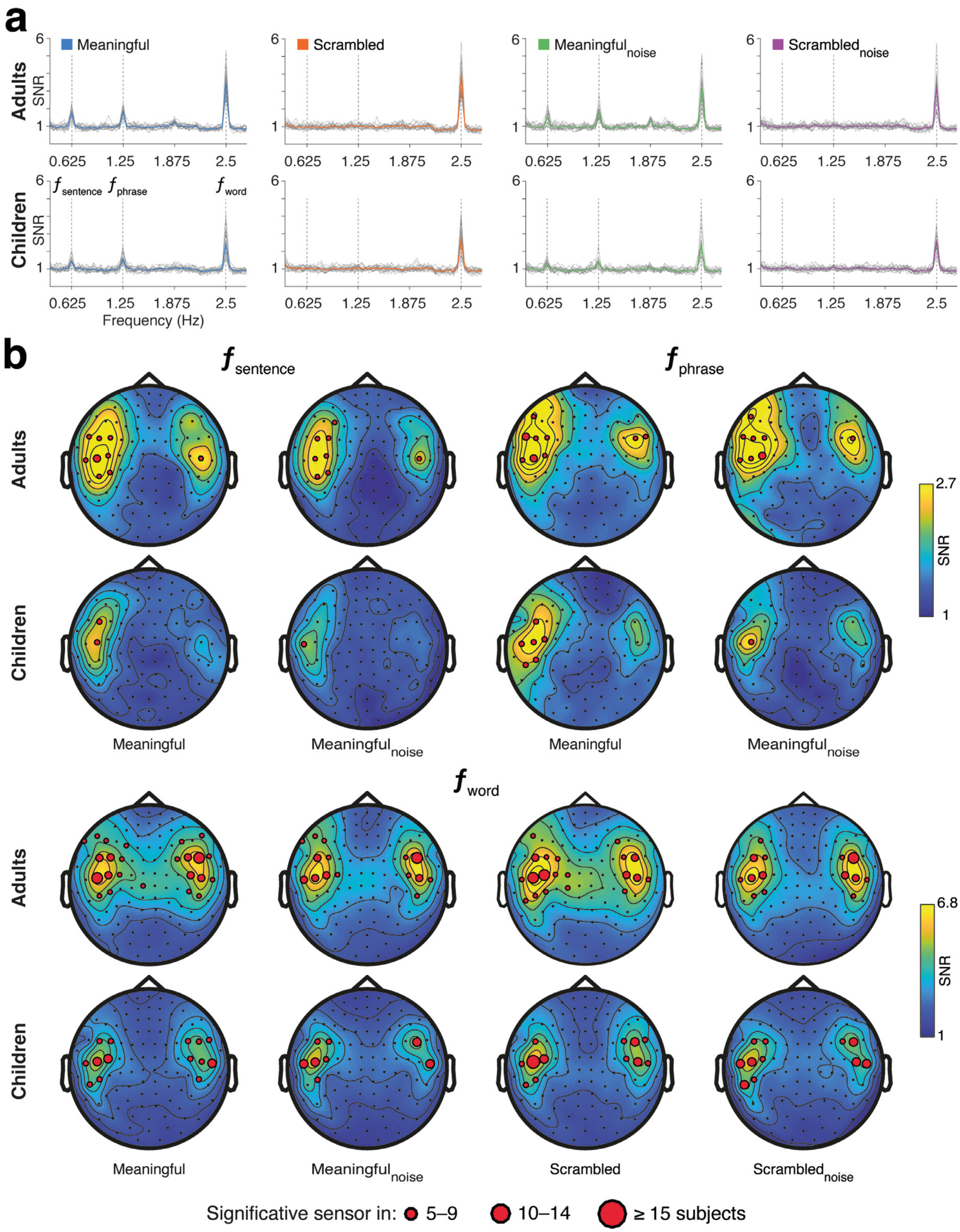

Fig. 3. Group-level SNR spectra (a) and the corresponding topographical maps (b) for the three frequencies of interest and both groups. In SNR spectra, individual traces are provided in thin gray lines.

in the sentence ($F_{3,114} < 1$) nor interaction involving this latter factor (all $p > 0.05$). *Post-hoc* comparisons between listening conditions demonstrated that scores were (1) not significantly different between *Meaningful* and *Scrambled* conditions (adults, $t_{19} = 1.26$, $p_{corr} = 1$; children, $t_{19} = 2.23$, $p_{corr} = 0.67$) but significantly higher in noiseless conditions compared to noisy conditions (all $p_{corr} < 0.0001$), (2) significantly higher in $Meaningful_{noise}$ compared with $Scrambled_{noise}$ condition (adults, $t_{19} = 8.64$, $p_{corr} < 0.0001$; children, $t_{19} = 5.57$, $p_{corr} < 0.0001$), and (3) lower in children compared with adults only in noisy conditions ($Meaningful_{noise}$, $t_{38} = 7.11$, $p_{corr} < 0.0001$; $Scrambled_{noise}$, $t_{38} = 4.11$, $p_{corr} = 0.002$).

Table 1
Number of adults and children showing statistically significant peaks in at least 1 sensor pair in each frequency and condition.

Condition	$f_{sentence}$ Adults	Children	$f_{phrasal}$ Adults	Children	f_{word} Adults	Children
Meaningful	18	9	20	17	20	20
Meaningful_{noise}	18	12	19	15	20	20
Scrambled	2	1	0	3	20	20
Scrambled_{noise}	2	1	3	0	20	20

The ANOVA performed on the intelligibility ratings (Fig. 2**b**) revealed the same significant effect of age group ($F_{1,38}$ = 5.99, p_{corr} = 0.019) and listening condition ($F_{3,114}$ = 84.5, p_{corr} < 0.0001), and a significant interaction between age group and listening condition ($F_{3,114}$ = 9.52, p_{corr} = 0.009). Post hoc comparisons demonstrated that intelligibility ratings were (1) significantly higher in noiseless compared to noisy conditions (all p_{corr} < 0.01), (2) not significantly different between *Meaningful* and *Scrambled* conditions nor between *Meaningful*$_{noise}$ and *Scrambled*$_{noise}$ conditions (all p_{corr} > 0.05), and (3) lower in children compared with adults only in the *Meaningful*$_{noise}$ condition (t_{38} = 3.5, p_{corr} = 0.018).

3.2. Cortical tracking of linguistic units

Fig. 3 displays group-averaged SNR spectra and sensor distributions for each listening condition and age group. There was a clear peak of SNR at 2.5 Hz (i.e., word rate, f_{word}) in all listening conditions and age groups, demonstrating excellent tracking of monosyllabic words. Other peaks were also noticeable at 1.25 Hz (i.e., phrase rate, f_{phrase}) and at 0.625 Hz (i.e., sentence rate, f_{sentence}), but only in the *Meaningful* and *Meaningful*$_{noise}$ conditions where sentential units were present. In all cases, SNR peaked at MEG sensors covering bilateral fronto-temporal areas. The cortical activity appears to be bilateral at f_{word} but predominantly left-lateralized at f_{sentence} and f_{phrase} in both groups. -**Supplementary Fig. 1a** illustrates the individual variability of SNR.

Table 1 provides the number of adults and children showing significant SNR at f_{word}, f_{phrase} and f_{sentence} for each condition. Tracking at f_{word} was significant in all participants and conditions. Furthermore, tracking at f_{phrase} and f_{sentence} were significant in most of the subjects in *Meaningful* and *Meaningful*$_{noise}$ conditions and in a non-significant proportion of subjects overall in the Scrambled and Scramblednoise conditions (12 in 160 instances, binomial cumulative density B(12,160,0.05) = 0.94, p = 0.12). Interestingly, sentence tracking (f_{sentence}) in children was significant only in about half of the children.

3.3. Cortical source localization

Fig. 4 displays the SNR source distribution in all conditions and in both age groups for each frequency of interest. Peaks of local maxima were consistently identified at the supra-temporal auditory cortex (STAC) and at the inferior frontal gyrus (IFG) bilaterally for all frequencies of interest, age groups and conditions. Indeed, across the listening conditions and age groups, local maxima of SNR were less than 7.0 mm apart at f_{word}, less than 13.4 mm at f_{sentence}, and less than 5.4 mm at f_{phrase}.

Table 2 presents the coordinates of the significant local maxima of SNR and the standard deviation for each coordinate.

Table 2
Local maxima of group-level SNR map for all frequencies of interest: MNI coordinates ± standard deviation (SD) and SNR. STAC = Supra-temporal auditory cortex; IFG = Inferior frontal gyrus;

Frequency	Regions	Location (Mean ± SD)			SNR
f_{sentence}	STAC	[−50 −25 8]	±	[4 7 5]	2.69
		[56 −22 12]	±	[5 3 7]	2.25
	IFG	[−42 9 14]	±	[3 4 4]	2.62
		[43 12 8]	±	[13 2 8]	1.93
f_{phrase}	STAC	[−43 −13 6]	±	[5 1 5]	3.51
		[57 0 12]	±	[2 1 2]	2.92
	IFG	[−35 30 4]	±	[1 2 4]	2.96
		[51 37 6]	±	[2 3 3]	2.35
f_{word}	STAC	[−42 −12 12]	±	[4 2 3]	6.93
		[51 −6 10]	±	[5 3 2]	6.52
	IFG	[−40 29 1]	±	[3 3 3]	4.38
		[45 32 −4]	±	[6 4 7]	4.19

3.4. Effect of age group, noise and hemisphere on the cortical tracking of hierarchical linguistic units

Fig. 5 displays the SNR values in *Meaningful* and *Meaningful*$_{noise}$ conditions, in both age groups and both hemispheres, for each frequency, and for the two identified sources of CTS (i.e., STAC, Fig. 5**a**, and IFG, Fig. 5**b**). **Supplementary Fig. 1b** illustrates the individual variability of SNR.

Table 3 summarizes the results of the separate LMM performed on SNR values in STAC and IFG, at $f_{sentence}$, f_{phrase} and f_{word} with factors noise, age group and hemisphere. The LMM revealed a significant effect of age group for both cortical areas, explained by lower SNR values in children compared with adults. They also revealed a significant effect of the hemisphere at f_{sentence} and f_{phrase} but not at f_{word} in both cortical areas, reflecting higher SNR in the left hemisphere compared with the right. No significant interactions were found at f_{sentence} and f_{phrase}. Analyses also revealed a main effect of noise at f_{sentence}, reflecting higher SNR in *Meaningful* compared with *Meaningful*$_{noise}$ condition. A main effect of noise was also uncovered at f_{word} but only in the inferior frontal gyri, with a significant interaction between age and noise. This interaction was explained by significantly higher SNR in *Meaningful*$_{noise}$ compared with *Meaningful* condition in adults (t_{19} = 3.39, p_{corr} = 0.0098) but not in children (t_{39} = 0.36, p_{corr} = 1).

In Supplementary Material S1, we present the results of an analysis of the SNR at f_{word} where the factor "meaning" (i.e., *Meaningful* and *Scrambled*) is included.

We further assessed to what extent the contrasts underlying the significant effects uncovered at the IFG were significantly different between IFG and STAC. The difference between conditions of the mean SNR across hemispheres did not differ significantly between sources (t_{39} = 0.30, p = 0.76; assessment for main effect of noise), and the contrast thereof between sources did not differ significantly between age groups (t_{38} = 0.57, p = 0.57; assessment for the interaction between age and noise). Overall, these findings demonstrate that SNR at f_{word} is significantly increased in adults compared with children across the considered sources (i.e., bilateral IFG and STAC).

3.5. Link between the cortical tracking of hierarchical linguistic units and behavior, children's age and head movements

3.5.1. Behavioral relevance

The SNR value at $f_{sentence}$, f_{phrase} and f_{word} in *Meaningful* and *Meaningful*$_{noise}$ did not correlate significantly with word identification score or SiN audiometry results (all p > 0.05). Of note, there was a significant positive correlation between the result of the speech audiometry in noise and the identification score in the *Meaningful*$_{noise}$ condition (r = 0.67, p = 0.001) in children but not in adults (r = 0.36, p = 0.12).

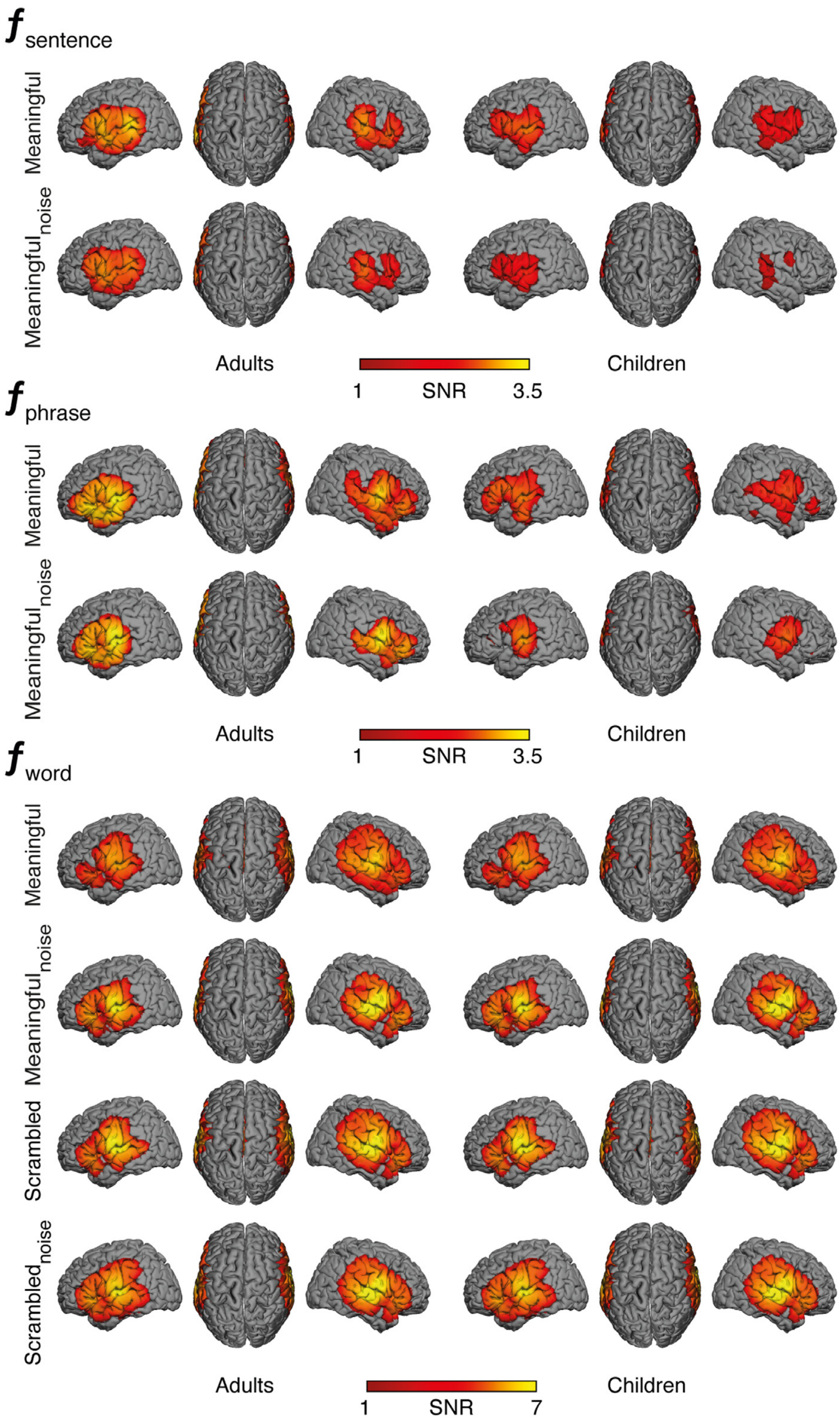

Fig. 4. Source distributions of the SNR for each group, condition, and frequency of interest.

3.5.2. *Effect of age*

We computed the difference of SNR between *Meaningful*$_{noise}$ and *Meaningful* conditions (mean over both hemispheres) to quantify the increase in SNR. Then, we correlated this measure with the age of children in months. A significant positive correlation ($r = 0.51$, $p = 0.02$) was observed (See **Supplementary Fig. 2**). The ability to boost the tracking of monosyllabic words in noise therefore develops throughout childhood. Moreover, the identification score in *Meaningful*$_{noise}$ positively correlates with the age in months ($r = 0.46$, $p = 0.044$).

3.5.3. *Head movements*

MEG recordings can be challenging with children, especially with their limited ability to remain still during acquisitions. Therefore, in Supplementary Material S2, we compare head movements between

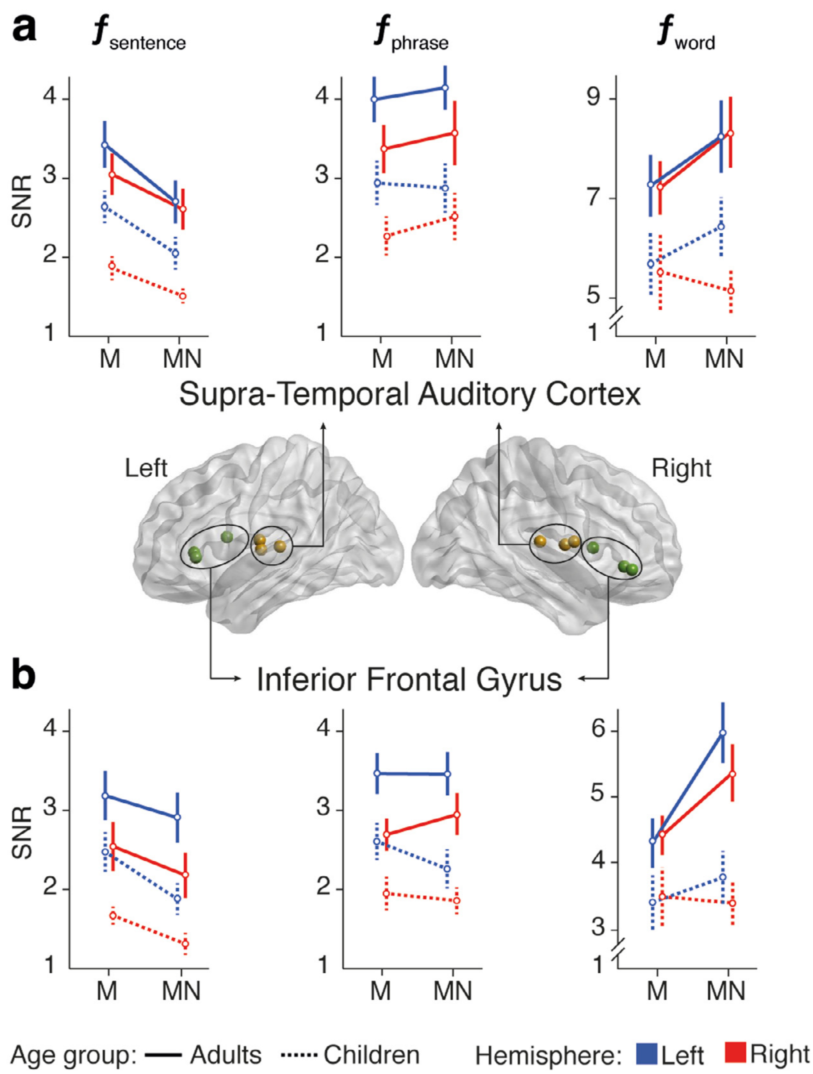

Fig. 5. SNR of CTS sources (mean ± 2 SEM) in *Meaningful* (M) and *Meaningful*$_{noise}$ (MN) conditions, in both age groups and in hemispheres, for each frequency at the two regions of interest: (a) Supra-Temporal Auditory Cortex and (b) Inferior Frontal Gyrus. Significant local maxima are illustrated on the MNI glass brain.

Table 3
Factors affecting the cortical tracking of hierarchical linguistic units. All significant results are in boldface.

Region	Factor	$f_{sentence}$ $F_{1,152}$	p	$f_{phrasal}$ $F_{1,152}$	p	f_{word} $F_{1,152}$	p
Supra-temporal auditory cortex	Age	**31.73**	**< 0.0001**	**27.28**	**< 0.0001**	**20.98**	**< 0.0001**
	Noise	**9.92**	**0.002**	0.19	0.66	1.9	0.17
	Hemisphere	**7.87**	**0.006**	**7.31**	**0.008**	0.61	0.44
	Age x Hemisphere	2	0.16	0	0.96	0.68	0.41
	Age x Noise	0.3	0.59	0.08	0.78	0.94	0.33
	Noise x Hemisphere	0.94	0.34	0.1	0.75	0.3	0.58
	Age x Hemisphere x Noise	0.02	0.89	0.07	0.8	0.49	0.48
Inferior frontal gyrus	Age	**20.45**	**< 0.0001**	**35.16**	**< 0.0001**	**28.46**	**< 0.0001**
	Noise	**5.44**	**0.02**	0.05	0.82	**6.67**	**0.01**
	Hemisphere	**11.92**	**0.0007**	**11.94**	**0.0007**	0.5	0.48
	Age x Hemisphere	0.03	0.87	0.03	0.88	0.03	0.86
	Age x Noise	0.31	0.58	0.7	0.4	**4.35**	**0.04**
	Noise x Hemisphere	0.01	0.94	0.78	0.38	1.14	0.29
	Age x Hemisphere x Noise	0.21	0.65	0	0.99	0.05	0.82

groups. As head movements were significantly related to SNR values mainly at f_{word} in children and as one of our main findings was an increase of CTS in adults in the presence of noise at f_{word}, we corrected SNR values for head movements ($SNR_{corrected}$) and computed the same LMM used before to measure if this facilitation effect is independent from head motions (i.e., separate LMM on $SNR_{corrected}$ values in STAC and IFG at f_{word} with factors noise, age group and hemisphere). We retrieved some effects already described in the previous LMM (i) significant effect of noise in the IFG ($F_{1,152} = 7.79$, $p = 0.006$) and (ii) no effect of hemisphere for both areas (all $p > 0.3$). Interestingly, no significant effects of age group for both cortical areas were found, explained by similar $SNR_{corrected}$ in children and adults. Crucially, we also retrieved the interaction between age and noise ($F_{1,152} = 4.94$, $p = 0.028$) reflected by higher $SNR_{corrected}$ in *Meaningful*$_{noise}$ compared with *Meaningful* condition in adults (t_{19} = 3.39, p_{corr} = 0.01) but not in children (t_{19} = 0.39, p_{corr} = 1). No other significant interactions were uncovered. Overall, these findings demonstrate that $SNR_{corrected}$ at f_{word} is significantly increased in adults compared with children and the absence of this facilitation effect in children is unlikely to be related to their increased head movement. Crucially, the difference of SNR between *Meaningful$_{noise}$* and *Meaningful* conditions (see above, section 3.5.2) was not correlated with head translations (r = –0.005, p = 0.98) or angles (r = 0.069, p = 0.77).

4. Discussion

This study shows that the child brain tracks the hierarchical linguistic units of clear speech devoid of prosody, with a left-hemisphere dominance as the adult brain, but with reduced accuracy. This study also demonstrates that a multi-talker background noise similarly reduces grammar-based cortical tracking of sentences in children and adults. Critically, when such noise is present, the adult brain increases the tracking of monosyllabic words while there is no evidence for such a mechanism in children. This cortical tracking enhancement process appears to develop through childhood.

4.1. Reduced grammar-based cortical tracking of speech hierarchical linguistic units in children

In a quiet environment, adults and children exhibited a clear peak of SNR, indicative of the presence of CTS, at word frequency in all conditions, and at phrase and sentence frequencies only when words formed meaningful sentences. In both groups, grammar-based cortical tracking of words, phrases and sentences originated from bilateral STAC and IFG. These brain areas are key nodes of the speech processing network (Hickok and Poeppel, 2007; Friederici, 2011). They have already been highlighted using non-invasive (Sohoglu et al. 2012; Peelle et al. 2013; Park et al. 2015; Di Liberto et al. 2018; Vander Ghinst et al. 2019; Bertels et al. 2022) and intracranial (Kubanek et al. 2013; Ding et al. 2016) electrophysiological recordings as being involved in the cortical tracking of natural connected speech both in adults and children. Those intracranial electrophysiological studies demonstrated that the CTS predominates over temporal auditory areas but also involved non-auditory high-order speech-related brain areas such as the IFG, and that this CTS was speech-specific (Kubanek et al. 2013; Ding et al. 2016).

CTS was lower in children compared with adults for all linguistic units (i.e., words, phrases, sentences). Several non-exclusive hypotheses can be raised to explain this finding. First, these results might illustrate the key role that prosodic cues play in childhood in supporting the cortical tracking of linguistic units in synergy with grammar-based knowledge (Mehler et al., 1988; Kalashnikova et al., 2018; Teixidó et al., 2018; Myers et al., 2019). The use of a similar experimental paradigm, but with the inclusion of prosodic features, might be of interest to confirm this hypothesis. Still, the reduced cortical tracking of monosyllabic words in *Meaningful* conditions does not really support this hypothesis, as the absence of prosody should affect the monosyllabic tracking to a lesser extent than for phrases and sentences. Second, connected speech processing also involves attentional processes (Sanes and Woolley, 2011; Jones et al., 2015; Shinn-Cunningham et al., 2017; Thompson et al., 2017, 2019). A certain level of attention is indeed required for combining syllables into words (Ding et al., 2018), or words into phrases/sentences (Makov et al., 2017). Thus, the reduced CTS observed in children might also be driven by reduced attention that is known to develop through childhood (Leibold, 2012). Third, this decrease can be the sign of cortical maturation, which allows better neural synchronization when it is fully developed until early adulthood (Ulhaas et al., 2010). Finally, this might be explained by reduced SNR of MEG signals in children due to smaller head size or increased head movements (Wehner et al., 2008; Witton et al., 2014). In fact, children's head moved more than that of adults in the MEG helmet and the mean displacement was negatively correlated with SNR at word frequency in the Meaningful conditions. This can explain why we found smaller SNR responses at the word frequency compared with adults, while the cortical tracking of natural connected speech in the theta range (corresponding to the syllabic rate) has been shown to be pretty stable and similar from childhood (>5 years) to adulthood (<27 years) in the left hemisphere (Bertels et al. 2022). Still, after correcting SNR for head movements, the cortical tracking of words appeared similar in children and adults. Crucially, even if the amount of head movements was greater in children, it was not correlated with the SNR at the sentence frequency, arguing that head movements can not explain on its own the lower SNR observed in children at the sentence frequency. The future use of on-scalp MEG based on optically pumped magnetometers that is more resilient to head motion should clarify those issues (Hill et al., 2019; de Lange et al., 2021).

The interpretation of this reduced CTS for all linguistic units (i.e., words, phrases, sentences) in children compared with adults is also complicated by the difficulty to assess its behavioral relevance for speech processing/understanding. Our experimental paradigm was designed to keep participants focused on the auditory stream by asking them to identify a randomly selected monosyllabic word at the end of each block of 12-16 words. This task was therefore not a direct measure of internal neural sentence building or comprehension. This can explain why we did not find any correlation between CTS values and word identification scores. One of the main interests of the paradigm used in this study is that SNR at sentence frequency provides a quantitative estimation of the participant's ability to group monosyllabic words into meaningful sentences, but it's still unclear how the inter-individual variability in SNR amplitudes actually relates to the variability in comprehension abilities (see Ding et al. 2017 for correlation with behavioral measurement). Further studies are needed to address this issue by designing a behavioral task directly assessing sentence comprehension. This could bring novel insights into the difference in CTS observed between children and adults, and also between individuals in similar age groups.

4.2 Impact of a multi-talker background on the cortical tracking of hierarchical linguistic units

Behavioral scores (SiN audiometry, word identification and intelligibility ratings) were significantly lower in children compared with adults in noise, while they were similar in a quiet environment. These results are perfectly in line with the well-described reduced SiN processing abilities of children <10 years (Elliott, 1979; Hall et al., 2002).

In a multi-talker background, adults and children's cortical activity tracked phrases and sentences, reflecting ongoing grammar-based neural building of the hierarchical linguistic units in such adverse auditory scenes. These results are in line with previous studies that used competing speakers as masker and were conducted in adults (Mesgarani and Chang 2012; Zion Golumbic et al. 2013; Ding and Simon 2013; O'Sullivan et al. 2015; Rimmele et al. 2015; Vander Ghinst et al. 2016; Decruy et al. 2019; Destoky et al. 2019; Fuglsang et al. 2017; Ding et al. 2018; Kulasingham et al. 2021) or children (Destoky et al., 2020; Vander Ghinst et al., 2019; Bertels et al., 2022). Those studies showed significant cortical tracking of the attended speech in a multi-talker background at similar SNR (i.e., +3 to 0 dB) than in our study. The amplitudes of STAC and IFG responses at sentence frequency were significantly lower in noisy conditions, which is comparable to the dampening of the cortical tracking of natural connected speech previously reported at ~0.5-Hz in children and adults when speech is polluted by a multi-talker background (Vander Ghinst et al., 2016, 2019; Giordano et al., 2017; Destoky et al., 2019, 2020; Bertels et al., 2022). This noise-related reduction in the cortical tracking of the attended speech at the sentence level plausibly accounts for the lower behavioral scores observed both in children and adults. Unfortunately, this was not verified by correlations between scores and SNR values, which might be related to the fact that these scores poorly reflected the process of internal grammar-based grouping of monosyllabic words into meaningful sentences. Still, no difference was observed in the effect of noise on the tracking of phrases/sentences between children and adults. But considering that children have a weaker tracking of phrases/sentences in the *Meaningful* condition compared with adults, the similar effect of noise between children and adults might actually have a higher functional impact in children that would partly explain why school-aged children have lower SiN processing abilities than adults.

The cortical tracking of phrases was not affected by noise. Binding two elements into a syntactic hierarchy is considered as the most basic operation of the hierarchic syntactic building (Mueller et al., 2012; Friederici, 2020) and has been shown to be already operational in prelinguistic infants (Mueller et al., 2012; Friederici, 2020). This might explain the robustness of the cortical tracking of phrases at this sound level of babble noise even in children.

The multi-talker background noise induced an increase in the cortical tracking of monosyllabic words, in adults but not in children. This increase was specific to meaningful connected speech, as it was observed between *Meaningful* conditions but not between *Scrambled* ones. The amplification of CTS (as assessed based on temporal response function) in noisy auditory scenes has already been described in elderly subjects (Presacco et al. 2016; Decruy et al. 2019; Mesik et al. 2021) and in adults with impaired hearing (Decruy et al. 2020; Fuglsang et al. 2020). Unlike our participants, these populations suffer from reduced hearing leading to less accurate peripheral neural representations. Therefore, in such populations, the increase in CTS may correspond to a central mechanism aiming at compensating for the peripheral deficit to improve SiN processing. On the other hand, such finding has also been observed for the cortical tracking of natural connected speech at the syllabic rate (i.e, 4-8 Hz) in healthy adults compared with school-aged children and in adults with impaired speech perception in noise (ISPiN); those populations being characterized by normal peripheral hearing. While the CTS was significantly increased in a multi-talker background (up to 0 dB) in healthy adults, such increase was not observed in children (<10 years) (Vander Ghinst et al., 2019), nor in adults with ISPiN (Vander Ghinst et al., 2021). These findings therefore argue for a central origin of the reduced SiN perception abilities in those two latter populations. As the strength of the CTS has been related to speech intelligibility (Ding and Simon 2013; Peelle et al. 2013; Ding et al. 2014; Doelling et al. 2014; O'Sullivan et al. 2015; Vanthornhout et al. 2018; Decruy et al. 2019; Bednar and Lalor 2020; Teoh et al. 2022), these findings highlight key neural correlates of the reduced behavioral abilities of school-aged children and adults with ISPiN to understand SiN (Vander Ghinst et al., 2019, 2021). Indeed, these data highly suggest that the ability to increase the neural tracking of (sub-)lexical linguistic units in adverse auditory scenes plays a key functional role in the human capacity to properly perceive and understand SiN. Impairments in this ability would represent a common neural correlate to physiological or pathological conditions characterized by a reduced SiN understanding at the behavioral level.

In the multi-talker background noise, the increased CTS of monosyllabic words observed in adults but not in children similarly involved bilateral STAC and IFG, though it predominated over frontal areas that are known to be involved in high-order speech processing. The similar modulations of syllabic CTS in these brain areas might be related to the limited ability of MEG to distinguish activity from nearby brain sources. Still, they appeared as clear independent local maxima in the reconstructed source volume. Their common involvement is also in line with the fact that they both tracked monosyllabic words (leading to correlated CTS values) and with previous intracranial recordings that demonstrated their speech-specific involvement in syllabic neural tracking (Kubanek et al. 2013; Ding et al. 2016). The fact that this enhanced CTS was only observed between the Meaningful conditions also supports the idea that it involved or contributed to high-order speech processing rather than being limited to (non-speech related) acoustic processing. Such increased CTS in the *Meaningful*$_{noise}$ condition might play a key role to support the internal neural grouping of small linguistic units into meaningful phrases/sentences in adverse auditory scenes. Finally, the ability to increase the syllabic CTS in noise positively correlated with age (in months) in children, arguing that this ability develops throughout childhood. Overall, these data suggest an immature cortical tracking of (sub-)lexical speech units in noise in children compared with adults. Further research is needed to confirm our hypothesis by comparing this specific speech processing to non-speech stimuli (using non-words or music rhythms).

4.3. Hemispheric dominance of the cortical tracking of speech hierarchical linguistic units

In the absence of any prosodic cue, the CTS in a quiet environment was clearly left-hemisphere dominant at the phrase and sentence frequencies, both in children and adults. This was already observed in studies using a similar paradigm (Sheng et al. 2019; Kulasingham et al. 2021). Contrastingly, previous studies using natural connected speech revealed that the auditory system tracks the slow fluctuations of speech temporal envelope (< 2Hz) preferentially in the right hemisphere (Bourguignon et al., 2013, 2018; Gross et al., 2013; Molinaro et al., 2016; Vander Ghinst et al., 2016, 2019; Destoky et al., 2019). These opposing results therefore provide additional empirical evidence supporting the hypothesis that the right-dominant CTS is mainly driven by the neural tracking of prosodic cues (Friederici, 2002; Bourguignon et al., 2013).

The present study identified a similar impact of noise on the grammar-based cortical tracking of sentences in children and adults. Yet, previous studies performed in school-aged children and adults consistently demonstrated that the cortical tracking of attended natural connected speech at phrase/sentence levels is more robust to noise in the left hemisphere compared with the right (Peelle et al., 2013; Rimmele et al., 2015; Vander Ghinst et al., 2016, 2019; Destoky et al., 2019). The right-hemisphere CTS is also more easily corrupted by noise in children than in adults (Vander Ghinst et al., 2019; Bertels et al., 2022). Considering that right-hemisphere CTS appears to be mainly driven by the neural tracking of prosodic cues, these data might thus provide indirect evidence suggesting that children's behavioral SiN processing difficulties might also be rooted in suboptimal non-verbal (i.e., prosodic) rather than grammar-based neural processing of the attended speech stream. Further studies are needed to confirm that hypothesis.

5. Conclusion

As compared to adults, school-aged children appear to be unable to enhance the cortical tracking of monosyllabic words in a multi-talker background noise, which might impair their ability to group them into meaningful phrases/sentences in such adverse auditory scenes. This might partly contribute to their lower behavioral ability to understand speech in noise. This effect comes in addition to a restricted cortical tracking of prosodic elements that has been previously shown in a multi-talker background.

6. Data and code availability statement

MEG data used in this study will be made available upon reasonable request to the corresponding author and after approval by institutional authorities (Hôpital Universitaire de Bruxelles and Université libre de Bruxelles)..

7. Contributions

M.N., M.B.., M.V.G, X.D.T. designed the study; M.N., M.B., J.B. collected the data; M.N., M.B., V.W, X.D.T. analyzed the data; M.N. wrote the initial version of the manuscript; and all authors discussed the results, their interpretation and commented on the manuscript.

Declaration of Competing Interest

The authors declare no competing financial interests.

Data availability

Data will be made available on request after approval by institutional authorities (Hôpital Universitaire de Bruxelles and Université libre de Bruxelles).

Acknowledgments

Maxime Niesen and Marc Vander Ghinst were supported by the Fonds Erasme (Brussels, Belgium). Mathieu Bourguignon and Julie Bertels have been supported by the program Attract of Innoviris (grants 2015-BB2B-10 and 2019-BFB-110). Julie Bertels has been supported by a research grant from the Fonds de Soutien Marguerite-Marie Delacroix (Brussels, Belgium). Xavier De Tiège is Clinical Researcher at the Fonds de la Recherche Scientifique (FRS-FNRS, Brussels, Belgium). We warmly thank Mélina Houinsou Hans for her statistical support during the review process.

This study and the MEG project at CUB Hôpital Erasme were financially supported by the Fonds Erasme (Research Convention: "Les Voies du Savoir", Fonds Erasme, Brussels, Belgium).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119770.

References

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.
Aiken, SJ, Picton, TW, 2008. Envelope and spectral frequency-following responses to vowel sounds. Hear. Res. 245 (1-2), 35–47.
Barry-Anwar, R, Hadley, H, Conte, S, Keil, A, Scott, LS, 2018. The developmental time course and topographic distribution of individual-level monkey face discrimination in the infant brain. Neuropsychologia 108, 25–31.
Bednar, A, Lalor, EC, 2020. Where is the cocktail party? Decoding locations of attended and unattended moving sound sources using EEG. Neuroimage 205, 116283.
Bertels, J, Bourguignon, M, de Heering, A, Chetail, F, De Tiège, X, Cleeremans, A, Destrebecqz, A, 2020. Snakes elicit specific neural responses in the human infant brain. Sci. Rep. 10, 7443.
Bertels, J, Niesen, M, Destoky, F, Coolen, T, Vander Ghinst, M, Rovai, A, Trotta, N, Baart, M, Molinaro, N, De Tiege, X, 2022. Neurodevelopmental oscillatory basis of speech processing in noise. Dev Cogn Neurosci 101181. https://doi.org/10.1016/j.dcn.2022.101181.
Blanco-Elorrieta, E, Ding, N, Pylkkänen, L, Poeppel, D, 2020. Understanding requires tracking: noise and knowledge interact in bilingual comprehension. J. Cogn. Neurosci. 32, 1975–1983.
Bourguignon, M, De Tiège, X, de Beeck, MO, 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, Molinaro, N, Wens, V, 2018. Contrasting functional imaging parametric maps: the mislocation problem and alternative solutions. Neuroimage 169, 200–211. doi:10.1016/j.neuroimage.2017.12.033, Available at.
Bronkhorst, AW, 2015. The cocktail-party problem revisited: early processing and selection of multi-talker speech. Atten. Percept. Psychophys. 77, 1465–1487.
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.
Decruy, L, Vanthornhout, J, Francart, T, 2019. Evidence for enhanced neural tracking of the speech envelope underlying age-related speech-in-noise difficulties. J. Neurophysiol. 122 (2), 601–615.
Decruy, L, Vanthornhout, J, Francart, T, 2020. Hearing impairment is associated with enhanced neural tracking of the speech envelope. Hear. Res. 393, 107961.
de Lange, P, Boto, E, Holmes, N, Hill, RM, Bowtell, R, Wens, V, De Tiège, X, Brookes, MJ, Bourguignon, M, 2021. Measuring the cortical tracking of speech with optically-pumped magnetometers. Neuroimage 233, 117969.
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, Bertels, J, Niesen, M, Wens, V, Ghinst, MV, Leybaert, J, Lallier, M, Ince, RAA, Gross, J, De Tiège, X, Bourguignon, M, 2020. Cortical tracking of speech in noise accounts for reading strategies in children. PLoS Biol. 18, e3000840.
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.
Di Liberto, GM, Lalor, EC, Millman, RE, 2018. Causal cortical dynamics of a predictive enhancement of speech intelligibility. Neuroimage 166, 247–258.
Ding, N, Chatterjee, M, Simon, JZ, 2014. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. Neuroimage 88, 41–46.
Ding, N, Melloni, L, Yang, A, Wang, Y, Zhang, W, Poeppel, D, 2017. Characterizing neural entrainment to hierarchical linguistic units using Electroencephalography (EEG). Front. Hum. Neurosci. 11, 481.
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.
Ding, N, Pan, X, Luo, C, Su, N, Zhang, W, Zhang, J, 2018. Attention is required for knowledge-based sequential grouping: insights from the integration of syllables into words. J. Neurosci. 38, 1178–1188.
Ding, N, Simon, JZ, 2013. Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. J. Neurosci. 33 (13), 5728–5735.
Ding, N, Simon, JZ, 2014. Cortical entrainment to continuous speech: functional roles and interpretations. Front. Hum. Neurosci. 8, 311.
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 Pt 2.
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.
Friederici, AD, 2002. Towards a neural basis of auditory sentence processing. Trends Cogn. Sci. 6, 78–84.
Friederici, AD, 2011. The brain basis of language processing: from structure to function. Physiol. Rev. 91, 1357–1392.
Friederici, AD, 2020. Hierarchy processing in human neurobiology: how specific is it? Philos. Trans. R. Soc. Lond. B Biol. Sci. 375, 20180391.
Fuglsang, SA, Dau, T, Hjortkjær, J, 2017. Noise-robust cortical tracking of attended speech in real-world acoustic scenes. Neuroimage 156, 435–444.
Fuglsang, SA, Märcher-Rørsted, J, Dau, T, Hjortkjær, J, 2020. Effects of sensorineural hearing loss on cortical synchronization to competing speech during selective attention. J. Neurosci. 40 (12), 2562–2572.
Giordano BL, Ince RAA, Gross J, Schyns PG, Panzeri S, Kayser C (2017) Contributions of local speech encoding and functional connectivity to audio-visual speech perception. eLife 6.
Giraud, A-L, Poeppel, D, 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15, 511–517.
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, 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.
Hall, JW, Grose, JH, Buss, E, Dev, MB, 2002. Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. Ear Hear. 23, 159–165. doi:10.1097/00003446-200204000-00008, Available at.
Hämäläinen, MS, Ilmoniemi, RJ, 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Med. Biol. Eng. Comput. 32, 35–42.
Hahne, A, Eckstein, K, Friederici, AD, 2004. Brain signatures of syntactic and semantic processes during children's language development. J. Cogn. Neurosci. 16 (7), 1302–1318.
Hickok, G, Poeppel, D, 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402.
Hill, RM, Boto, E, Holmes, N, Hartley, C, Seedat, ZA, Leggett, J, Roberts, G, Shah, V, Tierney, TM, Woolrich, MW, Stagg, CJ, Barnes, GR, Bowtell, R, Slater, R, Brookes, MJ, 2019. A tool for functional brain imaging with lifespan compliance. Nat. Commun. doi:10.1038/s41467-019-12486-x, Available at.
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.
Jones, PR, Moore, DR, Amitay, S, 2015. Development of auditory selective attention: why children struggle to hear in noisy environments. Dev. Psychol. 51, 353–369.
Kalashnikova, M, Peter, V, Di Liberto, GM, Lalor, EC, Burnham, D, 2018. Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. Sci. Rep. 8, 13745.
Kubanek, J, Brunner, P, Gunduz, A, Poeppel, D, Schalk, G, 2013. The tracking of speech envelope in the human cortex. PLoS One 8 (1), e53398.
Kulasingham, JP, Joshi, NH, Rezaeizadeh, M, Simon, JZ, 2021. Cortical processing of arithmetic and simple sentences in an auditory attention task. J. Neurosci. 41 (38), 8023–8039.
Leibold, LJ, 2012. Development of auditory scene analysis and auditory attention. Human Auditory Dev. 137–161. doi:10.1007/978-1-4614-1421-6_5, Available at.
Lins, OG, Picton, TW, Boucher, BL, Durieux-Smith, A, Champagne, SC, Moran, LM, Perez-Abalo, MC, Martin, V, Savio, G, 1996. Frequency-specific audiometry using steady-state responses. Ear Hear. 17 (2), 81–96.
Makov, S, Sharon, O, Ding, N, Ben-Shachar, M, Nir, Y, Zion Golumbic, E, 2017. Sleep disrupts high-level speech parsing despite significant basic auditory processing. J. Neurosci. 37, 7772–7781.
Männel, C, Schipke, CS, Friederici, AD, 2013. The role of pause as a prosodic boundary

marker: Language ERP studies in German 3-and 6-year-olds. Dev. Cogn. Neurosci. 5, 86–94.

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

Meng, Q, Hegner, YL, Giblin, I, McMahon, C, Johnson, BW, 2021. Lateralized cerebral processing of abstract linguistic structure in clear and degraded speech. Cereb. Cortex 31, 591–602. doi:10.1093/cercor/bhaa245, Available at.

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

Mesik, J, Ray, L, Wojtczak, M, 2021. Effects of age on cortical tracking of word-level features of continuous competing speech. Front. Neurosci. 15, 635126.

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.

Mueller, JL, Friederici, AD, Männel, C, 2012. Auditory perception at the root of language learning. Proc. Natl. Acad. Sci. U. S. A. 109, 15953–15958.

John, MS, Picton, TW, 2000. Human auditory steady-state responses to amplitude-modulated tones: phase and latency measurements. Hear. Res. 141 (1-2), 57–79.

Myers, BR, Lense, MD, Gordon, RL, 2019. Pushing the envelope: developments in neural entrainment to speech and the biological underpinnings of prosody perception. Brain Sci. 9. doi:10.3390/brainsci9030070, Available at.

Nichols, TE, Holmes, AP, 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15, 1–25. doi:10.1002/hbm.1058, Available at.

Nuñez, SC, Dapretto, M, Katzir, T, Starr, A, Bramen, J, Kan, E, Bookheimer, S, Sowell, ER, 2011. fMRI of syntactic processing in typically developing children: structural correlates in the inferior frontal gyrus. Dev. Cogn. Neurosci. 1, 313–323.

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

O'Sullivan, JA, Power, AJ, Mesgarani, N, Rajaram, S, Foxe, JJ, Shinn-Cunningham, BG, Slaney, M, Shamma, SA, Lalor, EC, 2015. Attentional selection in a cocktail party environment can be decoded from single-trial EEG. Cereb. Cortex 25, 1697–1706. doi:10.1093/cercor/bht355, Available at.

Park, H, Ince, RA, Schyns, PG, Thut, G, Gross, J, 2015. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. Curr. Biol. 25 (12), 1649–1653.

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.

Peykarjou, S, Hoehl, S, Pauen, S, Rossion, B, 2017. Rapid categorization of human and ape faces in 9-month-old infants revealed by fast periodic visual stimulation. Sci. Rep. 7, 12526.

Presacco, A, Simon, JZ, Anderson, S, 2016. Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. J. Neurophysiol. 116 (5), 2346–2355.

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

Rimmele, JM, Zion Golumbic, E, Schröger, E, Poeppel, D, 2015. The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. Cortex 68, 144–154.

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

Sassenhagen, J, Draschkow, D, 2019. Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. Psychophysiology 56, e13335.

Sheng, J, Zheng, L, Lyu, B, Cen, Z, Qin, L, Tan, LH, Huang, MX, Ding, N, Gao, JH, 2019. The cortical maps of hierarchical linguistic structures during speech perception. Cereb. Cortex 29 (8), 3232–3240.

Shinn-Cunningham, B, Best, V, Lee, AKC, 2017. Auditory Object Formation and Selection. Springer Handbook of Auditory Research:7–40 Available at doi:10.1007/978-3-319-51662-2_2.

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

Skeide, MA, Friederici, AD, 2016. The ontogeny of the cortical language network. Nat. Rev. Neurosci. 17, 323–332.

Sohoglu, E, Peelle, JE, Carlyon, RP, Davis, MH, 2012. Predictive top-down integration of prior knowledge during speech perception. J. Neurosci. 32 (25), 8443–8453.

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. doi:10.1109/tsp.2005.853302, Available at.

Teixidó, M, François, C, Bosch, L, Männel, C, 2018. The role of prosody in early speech segmentation and word-referent mapping. In: Trends in Language Acquisition Research. John Benjamins Publishing Company, Amsterdam, pp. 79–100.

Teoh, ES, Ahmed, F, Lalor, EC, 2022. Attention differentially affects acoustic and phonetic feature encoding in a multispeaker environment. J. Neurosci. 42 (4), 682–691.

Thompson, EC, Krizman, J, White-Schwoch, T, Nicol, T, Estabrook, R, Kraus, N, 2019. Neurophysiological, linguistic, and cognitive predictors of children's ability to perceive speech in noise. Dev. Cogn. Neurosci. 39, 100672.

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.

Uhlhaas, PJ, Roux, F, Rodriguez, E, Rotarska-Jagiela, A, Singer, W, 2010. Neural synchrony and the development of cortical networks. Trends Cogn. Sci. 14 (2), 72–80.

Vander Ghinst, M, 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. J. Neurosci. doi:10.1523/JNEUROSCI.1732-18.2019, at.

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.

Vander Ghinst, M, Bourguignon, M, Wens, V, Naeije, G, Ducène, C, Niesen, M, Hassid, S, Choufani, G, Goldman, S, De Tiège, X, 2021. Inaccurate cortical tracking of speech in adults with impaired speech perception in noise. Brain Commun. 3, fcab186.

Vanthornhout, J, Decruy, L, Wouters, J, Simon, JZ, Francart, T, 2018. Speech intelligibility predicted from neural entrainment of the speech envelope. J. Assoc. Res. Otolaryngol. 19 (2), 181–191.

Vigario, R, Sarela, J, Jousmäki, V, Hämäläinen, M, Oja, E, 2000. Independent component approach to the analysis of EEG and MEG recordings. IEEE Trans. Biomed. Eng. 47, 589–593.

Wehner, DT, Hämäläinen, MS, Mody, M, Ahlfors, SP, 2008. Head movements of children in MEG: quantification, effects on source estimation, and compensation. Neuroimage 40, 541–550.

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.

Witton, C, Furlong, PL, Seri, S, 2014. Technological challenges of pediatric MEG and potential solutions: the aston experience. Magnetoencephalography 645–655. doi:10.1007/978-3-642-33045-2_30, Available at.

Xia, M, Wang, J, He, Y, 2013. BrainNet Viewer: a network visualization tool for human brain connectomics. PLoS One 8, e68910.

Yamashita, Y, 2013. A review of paralinguistic information processing for natural speech communication. Acoust. Sci. Technol. 34, 73–79. doi:10.1250/ast.34.73, Available at.

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.