Neurodevelopmental oscillatory basis of speech processing in noise

Bertels, Julie; Niesen, Maxime; Destoky, Florian; Coolen, Tim; Vander Ghinst, Marc; Wens, Vincent; Rovai, Antonin; Trotta, Nicola; Baart, Martijn; Molinaro, Nicola; De Tiège, Xavier; Bourguignon, Mathieu

Published in:
Developmental Cognitive Neuroscience

DOI:
10.1016/j.dcn.2022.101181

Publication date:
2023

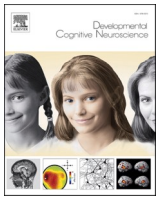
Document Version
Publisher's PDF, also known as Version of record

Developmental Cognitive Neuroscience 59 (2023) 101181

Contents lists available at ScienceDirect

Developmental Cognitive Neuroscience

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



Neurodevelopmental oscillatory basis of speech processing in noise

Julie Bertels [a,b,*,1], Maxime Niesen [a,c,1], Florian Destoky [a], Tim Coolen [a,d], Marc Vander Ghinst [a,c], Vincent Wens [a,e], Antonin Rovai [a,e], Nicola Trotta [a,e], Martijn Baart [f,g], Nicola Molinaro [f,h], Xavier De Tiège [a,e], Mathieu Bourguignon [a,f,i]

[a] Laboratoire de Neuroanatomie et Neuroimagerie translationnelles, UNI – ULB Neuroscience Institute, Université libre de Bruxelles (ULB), Brussels, Belgium
[b] ULBabyLab – Consciousness, Cognition and Computation group, UNI – ULB Neuroscience Institute, Université libre de Bruxelles (ULB), Brussels, Belgium
[c] Service d'ORL et de chirurgie cervico-faciale, CUB Hôpital Erasme, Université libre de Bruxelles (ULB), Brussels, Belgium
[d] Department of Radiology, CUB Hôpital Erasme, Université libre de Bruxelles (ULB), Brussels, Belgium
[e] Department of Functional Neuroimaging, Service of Nuclear Medicine, CUB Hôpital Erasme, Université libre de Bruxelles (ULB), Brussels, Belgium
[f] BCBL, Basque Center on Cognition, Brain and Language, San Sebastian, Spain
[g] Department of Cognitive Neuropsychology, Tilburg University, Tilburg, the Netherlands
[h] Ikerbasque, Basque Foundation for Science, Bilbao, Spain
[i] Laboratory of Neurophysiology and Movement Biomechanics, UNI – ULB Neuroscience Institute, Université libre de Bruxelles (ULB), Brussels, Belgium

ARTICLE INFO

Keywords:
Speech-in-noise (SiN) perception
Development
Cortical tracking of speech (CTS)
Magnetoencephalography (MEG)
Audiovisual speech integration

ABSTRACT

Humans' extraordinary ability to understand speech in noise relies on multiple processes that develop with age. Using magnetoencephalography (MEG), we characterize the underlying neuromaturational basis by quantifying how cortical oscillations in 144 participants (aged 5–27 years) track phrasal and syllabic structures in connected speech mixed with different types of noise. While the extraction of prosodic cues from clear speech was stable during development, its maintenance in a multi-talker background matured rapidly up to age 9 and was associated with speech comprehension. Furthermore, while the extraction of subtler information provided by syllables matured at age 9, its maintenance in noisy backgrounds progressively matured until adulthood. Altogether, these results highlight distinct behaviorally relevant maturational trajectories for the neuronal signatures of speech perception. In accordance with grain-size proposals, neuromaturational milestones are reached increasingly late for linguistic units of decreasing size, with further delays incurred by noise.

1. Introduction

Understanding speech in noise (SiN) is a challenging task, especially for children (Sanes and Woolley, 2011; Elliott, 1979). Paradoxically, noise is ubiquitous in children's lives (e.g., in classrooms, school cafeterias and playgrounds) and has deleterious effects on learning and academic performances (Shield and Dockrell, 2008). Still, how the neural mechanisms involved in SiN comprehension mature across development is poorly understood. Characterizing these developmental phenomena appears critical to devise strategies to help children cope with ambient noise in their daily life and to better understand the etiology of learning disorders.

A large body of literature has examined the neurophysiological correlates of SiN processing through investigations of the cortical tracking of speech (CTS) (Destoky et al., 2019; Ding and Simon, 2012, 2013; Fuglsang et al., 2017; Horton et al., 2013; Mesgarani and Chang, 2012; O'Sullivan, 2014; Puschmann et al., 2017; Rimmele et al., 2015; Simon, 2015; Vander Ghinst et al., 2016, 2019; Zion-Golumbic and Schroeder, 2012). CTS is the synchronization between human cortical activity and the fluctuations of speech temporal envelope at frequencies that match the hierarchical temporal structure of linguistic units such as phrases/sentences (0.2–1.5 Hz) and syllables/words (2–8 Hz) (Ahissar et al., 2001; Bourguignon et al., 2013; Donhauser and Baillet, 2020; Gross et al., 2013; Luo and Poeppel, 2007; Meyer et al., 2017; Meyer and Gumbert, 2018; Molinaro et al., 2016; Peelle et al., 2013). Functionally, CTS would subserve the segmentation of these units in connected speech to promote subsequent speech recognition (Ahissar et al., 2001; Gross et al., 2013; Meyer et al., 2017; Meyer, 2018; Ding et al., 2016; Ding and

* Corresponding author at: Laboratoire de Neuroanatomie et Neuroimagerie translationnelles, UNI – ULB Neuroscience Institute, Université libre de Bruxelles (ULB), Brussels, Belgium.
E-mail address: julie.bertels@ulb.be (J. Bertels).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.dcn.2022.101181
Received 28 March 2022; Received in revised form 31 October 2022; Accepted 25 November 2022
Available online 26 November 2022

Simon, 2014). Importantly, school-age children show reliable CTS (Molinaro et al., 2016; Power et al., 2016; Ríos-López et al., 2020) that is however lower at the syllabic level compared to adults (Vander Ghinst et al., 2019). In SiN conditions, children's and adults' cortical activity preferentially tracks the attended speech rather than the global sound (Destoky et al., 2019; Ding and Simon, 2012; Vander Ghinst et al., 2016; Zion-Golumbic et al., 2013), suggesting that CTS is modulated by endogenous attentional components and plays a role in segregating the attended linguistic signal (Ding and Simon, 2012; Fuglsang et al., 2017; Horton et al., 2013; Mesgarani and Chang, 2012; Puschmann et al., 2017; Vander Ghinst et al., 2016, 2019). However, the fidelity of the tracking decreases with increasing noise intensity in adults and more so in children (Destoky et al., 2019; Vander Ghinst et al., 2016, 2019), especially when the noise is concurrent speech babble (Destoky et al., 2020) as opposed to non-speech noise such as white or spectrally-matched noise. Of note, the visual speech signal (comprising the articulatory movements of a talker) boosts CTS in adults (Crosse et al., 2015; Luo et al., 2010; Park et al., 2018), especially in noise conditions (Zion-Golumbic et al., 2013; Crosse et al., 2016a; Giordano et al., 2017), and in children, at least in babble noise conditions (Destoky et al., 2020). Importantly, the modulations of CTS we have just outlined mirror tangible behavioral effects. That is, SiN perception and comprehension *(i)* decline with increasing noise level (Destoky et al., 2019; Vander Ghinst et al., 2016, 2019), *(ii)* are more affected by babble than non-speech noise in adults and especially in children (Corbin et al., 2016; Hall et al., 2002; Leibold, 2017), *(iii)* improve until late childhood, if not until adolescence in babble noise conditions (Elliott, 1979; Destoky et al., 2019; Vander Ghinst et al., 2019), and *(iv)* benefit from visual speech (Schwartz et al., 2004; Sumby and Pollack, 1954) since infancy (Patterson and Werker, 2003; Dodd, 1979), but increasingly more as age increases (Ross et al., 2011; Wightman et al., 2006).

Overall, these data suggest (i) that different aspects of CTS, whose behavioral relevance is well demonstrated, undergo different developmental trajectories, and (ii) that these trajectories depend on noise properties and availability of visual speech information. However, since previous studies focused on restricted age ranges and noise conditions, a detailed characterization of these trajectories is still lacking. The present magnetoencephalography (MEG) study aims at filling this gap by outlining the developmental trajectory of phrasal and syllabic CTS and speech comprehension, from early school age to early adulthood in various noise conditions, with or without visual speech information. Our research hypotheses were guided by *grain-size* proposals according to which children develop awareness of increasingly smaller phonological units with age (Anthony and Francis, 2005; Goswami and Ziegler, 2006). Extrapolating these proposals to supra-phonological units (Goswami, 2015; Gross et al., 2013), we hypothesized that the cortical tracking of large units such as phrases and sentences would mature faster during development compared with the tracking of smaller units such as syllables. Also, since coping with noise and leveraging visual speech information require the development and integration of additional mechanisms subtended by high-order associative neocortical areas that mature during late childhood (Gogtay et al., 2004), we hypothesize that corresponding developmental trajectories would be further delayed. Of note, how these maturation processes parallel structural changes (Fjell et al., 2015) was beyond the scope of the present study.

2. Material and methods

2.1. Participants

In total, 144 native French-speaking healthy right-handed children and young adults (age range: 5–27 years, 77 females) participated in this study. For some of the upcoming analyzes, participants were assigned to 5 age groups: 5–7 years ($n = 31$, 17 females), 7–8.5 years ($n = 34$, 17 females), 8.5–11.5 years ($n = 28$, 13 females), 11.5–18 years ($n = 27$, 15 females) and 18–27 years ($n = 25$, 15 females). The data collected from 73 of them was used in a previous study by our team (Destoky et al., 2020). Most of our participants were aged below 12, since SiN abilities essentially develop before that age. As a consequence, the 3 age groups of school-age children (5–7 years, 7–8.5 years and 8.5–11.5 years) span a narrower age range than the groups of teenagers (11.5–18 years) and young adults (18–27 years).

All participants had normal hearing according to pure-tone audiometry (i.e., hearing thresholds between 0 and 20 dB HL for 125, 250, 500, 1000, 2000, 4000 and 8000 Hz), and normal dichotic perception, speech, and SiN perception for their age (data missing for the 20 youngest participants) according to another test assessing speech perception in noise (Demanez et al., 2003).

The study had prior approval by the ULB-Hôpital Erasme Ethics Committee (Brussels, Belgium). Each participant or their legal representative gave written informed consent before participation. Participants were compensated with a gift card worth 25 euros for the neuroimaging assessment reported in the present study.

2.2. Stimuli

The stimuli were derived from 12 audiovisual recordings of 4 native French-speaking narrators (2 females, 3 recordings per narrator) telling a story for ~6-min (mean ± SD, 6.0 ± 0.8 min). Stories consisted of children's fairy tales; for more details, see our previous report (Destoky et al., 2020). In each video, the first 5 s were kept unaltered to enable participants to unambiguously identify the narrator's voice and face they were requested to attend. The remainder of the video was divided into 10 consecutive blocks of equal size that were assigned to 9 conditions. Two blocks were assigned to the *noiseless* condition in which the audio track was kept but the video was replaced by static pictures illustrating the story (mean ± SD picture presentation time across all videos, 27.7 ± 10.8 s). The remaining 8 blocks were assigned to 8 conditions in which the original sound was mixed with a background noise at 3 dB signal-to-noise ratio (SNR). This SNR was chosen as we assumed it was high enough to ensure children could cope with the noise and keep their attention to the story, and low enough to introduce non-negligible interference; both assumptions proved accurate *a posteriori*. There were 4 different types of noise, and each type of noise was presented once with visual speech information (the original video), and once without visual speech information (static pictures illustrating the story). The different types of noise differed in the degree of energetic and informational interference they introduced (Pollack, 1975). The least-energetic non-speech (i.e., non-informational) noise was a white noise high-pass filtered at 10,000-Hz. The most-energetic non-speech noise had its spectral properties dynamically adapted to mirror those of the narrator's voice ~1 s around. The different-gender babble noise was a 5-talker cocktail party noise recorded by individuals of gender different from the narrator's (i.e., a 5-male talker for female narrators, and vice-versa). The same-gender babble noise was a 5-talker cocktail party noise recorded by individuals of gender identical to the narrator's. The different- and same-gender babble noises introduced informational interferences and a similar degree of energetic masking (Destoky et al., 2020). Their distinction is however relevant since speech intelligibility is generally better when attended and interfering speech are uttered by different-gender talkers compared to same-gender talkers (Brungart, 2001; Bronkhorst, 2015), because on average, voice fundamental frequency and vocal tract length differ between males and females (Bronkhorst, 2015; Darwin et al., 2003). For both babble noises, the 5 individual noise components were obtained from a French audiobook database (http://www.litteratureaudio.com), normalized, and mixed linearly. The assignment of conditions to blocks was random, with the constraint that each of the 5 first and last blocks contained exactly 1 *noiseless* audio and each type of noise, 2 with visual speech and 2 without. Smooth audio and video transitions between blocks was ensured with 2-s fade-in and fade-out. Ensuing videos were grouped in 3 disjoint sets featuring one video of each of the narrators (total set

duration: 23.0, 24.3, 24.65 min), and there were 4 versions of each set differing in condition random ordering.

Fig. 1 illustrates the time-course of a video stimulus.

2.3. Experimental paradigm

During the imaging session, participants were laying on a bed with their head inside the MEG helmet. The lying position was chosen to maximize participants' comfort and reduce head movements, which would be expected to be higher in the youngest participants if they were sitting. Participants' brain activity was recorded while they were attending 4 videos (separate recording for each video) of a randomly selected set and ordering of the videos presented in a random order, and finally while they were at *rest* (eyes opened, fixation cross) for 5 min. They were instructed to watch the videos attentively, listen to the narrators' voice while ignoring the interfering noise, and remain as still as possible. After each video, they were asked 10 yes/no simple comprehension questions.

Videos were projected onto a back-projection screen placed vertically, ~120 cm away from the MEG helmet. The inner dimensions of the black frame were 35.2 cm (horizontal) and 28.8 cm (vertical), and the narrator's face spanned ~15 cm (horizontal) and ~20 cm (vertical). Participants could see the screen through a mirror placed above their head. In total the optical path from the screen to participants' eyes was ~150 cm. Sounds were delivered at 60 dB (measured at ear-level) through a MEG-compatible front-facing flat-panel loudspeaker (Panphonics Oy, Espoo, Finland) placed ~1 m behind the screen.

2.4. Data acquisition and processing

During the experimental conditions, participants' brain activity was recorded with MEG at the CUB Hôpital Erasme. MEG was preferred to electroencephalography for its higher spatial resolution (Baillet, 2017), and for its increased sensitivity to CTS (Destoky et al., 2019). Neuromagnetic signals were recorded with a whole-scalp-covering MEG system (Triux, Elekta) placed in a lightweight magnetically shielded room (Maxshield, Elekta), the characteristics of which have been described elsewhere (De Tiège et al., 2008). The sensor array of the MEG system comprised 306 sensors arranged in 102 triplets of one magnetometer and two orthogonal planar gradiometers. Magnetometers measure the radial component of the magnetic field, while planar gradiometers measure its spatial derivative in the tangential directions. MEG signals were band-pass filtered at 0.1–330 Hz and sampled at 1000 Hz.

We used 4 head-position indicator coils to monitor subjects' head position during the experimentation. Before the MEG session, we digitized the location of these coils and at least 300 head-surface points (on scalp, nose, and face) with respect to anatomical fiducials with an electromagnetic tracker (Fastrack, Polhemus).

Finally, subjects' high-resolution 3D-T1 cerebral images were acquired with a magnetic resonance imaging (MRI) scanner (MRI 3 T, Signa, General Electric) after the MEG session.

2.5. Data preprocessing

Continuous MEG data were first preprocessed off-line using the temporal signal space separation method implemented in MaxFilter software (MaxFilter, Neuromag, Elekta; 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). Head movement compensation is highly recommended when different age groups are compared (Larson and Taulu, 2017). To further suppress physiological artifacts, 30 independent components were evaluated from the data band-pass filtered at 0.1–25 Hz and reduced to a rank of 30 with principal component analysis. Independent components corresponding to heartbeat, eye-blink, and eye-movement artifacts were identified, and corresponding MEG signals reconstructed by means of the mixing matrix were subtracted from the full-rank data. Across subjects and conditions, the number of subtracted components was 3.45 ± 1.23 (mean ± SD across subjects and recordings). Finally, time points at timings 1 s

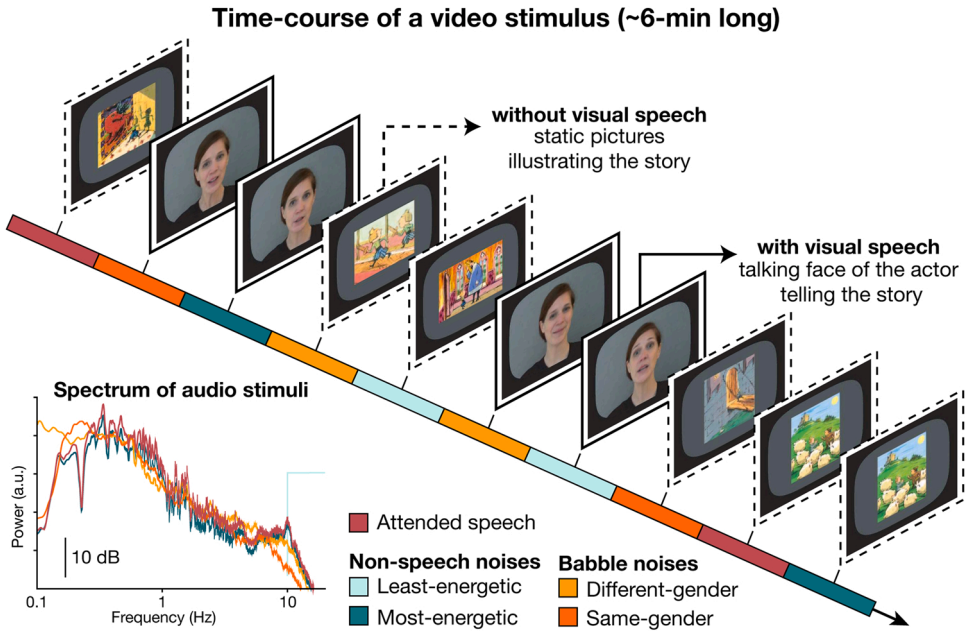


Fig. 1. Illustration of the time-course of a video stimulus. Videos lasted approximately 6 min and were divided into 10 blocks to which experimental conditions were assigned. There were two blocks of the noiseless condition, and eight blocks of speech-in-noise conditions: one block for each possible combination of the four types of noise and the two types of visual display.

around remaining artifacts were set to bad. Data were considered contaminated by artifacts when MEG amplitude exceeded 5 pT in at least one magnetometer or 1 pT/cm in at least one gradiometer.

We extracted the temporal envelope of the attended speech (narrators' voice) using a state-of-the-art approach (Biesmans et al., 2017). Briefly, audio signals were bandpass filtered using a gammatone filter bank (15 filters centered on logarithmically-spaced frequencies from 150 Hz to 4000 Hz), and subband envelopes were computed using Hilbert transform, elevated to the power 0.6, and averaged across bands.

2.6. CTS quantification

For each condition and participant, a global value of CTS for the attended speech was evaluated for all left-hemisphere gradiometer sensors at once, and for all right-hemisphere gradiometer sensors at once. We assessed the left and right hemispheres separately because CTS is hemispherically asymmetric both in noiseless and SiN conditions (Vander Ghinst et al., 2016). Using the mTRF (multivariate Temporal Response Function) toolbox (Crosse et al., 2016b), we trained a decoder on MEG data to reconstruct speech temporal envelope, and estimated its Pearson correlation with real speech temporal envelope. This correlation is often referred to as the reconstruction accuracy, and it provides a global measure of CTS. We did this for MEG signals filtered at 0.2–1.5 Hz (phrasal rate, which also englobes sentential rate) (Destoky et al., 2019, 2020; Bourguignon et al., 2019) and 2–8 Hz (syllabic rate, which also englobes word rate) (Ding and Simon, 2012; O'Sullivan et al., 2014; Zion-Golumbic et al., 2013; Lalor and Foxe, 2010).

The decoder tested on a given condition was built based on MEG data from all the other conditions (i.e., based on about 20 min of data). This procedure was preferred over a more conventional cross-validation approach in which the decoder is trained and tested on separate chunks of data from the same condition for which at most ~2.4 min of data was available. It is based on the rationale that the different conditions do modulate response amplitude but not its topography and temporal dynamics. In practice, electrophysiological data were bandpass filtered at 0.2–1.5 Hz (phrasal rate) or 2–8 Hz (syllabic rate), resampled to 10 Hz (phrasal) or 40 Hz (syllabic) and standardized. The decoder was built based on MEG data from –500–1000 ms (phrasal) or from 0 ms to 250 ms (syllabic) with respect to speech temporal envelope. Filtering and delay ranges were as in previous studies for phrasal (Destoky et al., 2019; Bourguignon et al., 2019), and syllabic CTS (Ding and Simon, 2012; O'Sullivan et al., 2014; Zion-Golumbic et al., 2013; Lalor and Foxe, 2010). Regularization was applied to limit the norm of the derivative of the reconstructed speech temporal envelope (Crosse et al., 2016b), by estimating the decoder for a fixed set of ridge values ($\lambda = 2^{-10}, 2^{-8}, 2^{-6}, 2^{-4}, 2^{-2}, 2^{0}$). The regularization parameter was determined with a classical 10-fold cross-validation approach: the data is split into 10 segments of equal length, the decoder is estimated for 9 segments and tested on the remaining segment, and this procedure is repeated 10 times until all segments have served as test segment. The ridge value yielding the maximum mean reconstruction accuracy is then retained. The ensuing decoder was then used to reconstruct speech temporal envelope in the left-out condition. Reconstruction accuracy was then estimated in 10 disjoint consecutive segments. We then retained the mean of this reconstruction accuracy, leaving us with one value for all combinations of subjects, conditions, hemispheres, and frequencies of interest. Note that, as documented in previous studies (Destoky et al., 2020, 2022), most of the 73 youngest participants of the present study showed significant CTS.

2.7. Normalized CTS in SiN conditions

Based on CTS values, we derived the normalized CTS (nCTS) in SiN conditions as the following contrast between CTS in SiN (CTS_{SiN}) and noiseless ($CTS_{\mathrm{noiseless}}$) conditions:

$$nCTS = (CTS_{\mathrm{SiN}} - CTS_{\mathrm{noiseless}})/(CTS_{\mathrm{SiN}} + CTS_{\mathrm{noiseless}})$$

Such contrast presents the advantage of being specific to SiN processing abilities by factoring out the global level of CTS in the noiseless condition. However, it can be misleading when derived from negative CTS values (which may happen since CTS is an unsquared correlation value). For this reason, CTS values below a threshold of 10% of the mean CTS across all subjects, conditions and hemispheres were set to that threshold prior to nCTS computation. Thanks to this thresholding, the nCTS index takes values between –1 and 1, with negative values indicating that the noise reduces CTS.

2.8. Developmental trajectory of (n)CTS

We used repeated measures ANOVA to assess the effect of brain hemisphere (left vs. right) and age group on CTS in noiseless conditions (dependent variable). This analysis was run separately for phrasal and syllabic CTS.

We used the same approach to analyze nCTS values in SiN conditions, this time with two additional factors: type of noise (least-energetic non-speech, most-energetic non-speech, different-gender babble vs. same-gender babble) and type of visual input (with vs. without visual speech). For both phrasal and syllabic nCTS, Mauchly sphericity tests indicated non-sphericity for the effect and interactions including the factor "type of noise" ($p < 0.01$). For this reason, Greenhouse-Geisser corrections were applied when needed.

For statistically significant effects involving age group, we used a model fitting approach to estimate the developmental trajectory of (n)CTS averaged across irrelevant factors. This approach is explained here for CTS, but the same was used for nCTS. We fitted to individual values of CTS three models involving different types of dependence on age:

Constant model: $CTS(age) = CTC_{\mathrm{constant}}$

Linear model: $CTS(age) = CTS_0 + slope \times age$

Logistic model: $CTS(age) = CTS_{\mathrm{min}} + (CTS_{\mathrm{max}} - CTS_{\mathrm{min}})/(1+\exp(-k_a \times (age-age_{\mathrm{trans}})))$

The logistic model features an evolution of CTS with age from CTS_{min} to CTS_{max} with a transition at age_{trans} occurring at rate k_a. Following this model, the maturation of CTS values roughly starts at $age_{\mathrm{trans}} - 2.2/k_a$ and finishes at $age_{\mathrm{trans}} + 2.2/k_a$, corresponding to 10% and 90% of the evolution from CTS_{min} to CTS_{max} (respectively). We also report on the percentage of increase in CTS, which is obtained as $(CTS_{\mathrm{max}} - CTS_{\mathrm{min}})/CTS_{\mathrm{min}} \times 100\%$.

Parameters were estimated with the least-square criterion, so that their values for the constant and linear models were trivial to obtain. Parameters of the logistic model were estimated with *fminsearch* Matlab function.

The models were compared statistically with a classical *F* test.

A potential pitfall with the initial step of the above-described approach is that age, a continuous variable, was treated as a categorical factor comprising 5 levels (age groups) in the ANOVA. This comes with a potential decrease in statistical power and increase in risks of false-positives (Altman and Royston, 2006; MacCallum et al., 2002). Therefore, we cross-checked our results by running the same ANOVA four times, with participants grouped into 4, 5, 6, and 7 equally-populated age groups. In the results section, we report only the results for the initial analysis including the 5 predefined age groups, while confirming that statistical (non-)significance was corroborated by at least 3 of the 4 outcomes for alternative groupings. Of note, treating age as a continuous factor was prohibitively challenging given the expected non-linear relationship between age and (n)CTS that could depend on other factors.

2.9. Behavioral relevance of the CTS

We assessed the behavioral relevance of the tracking measures that showed significant maturation effects. For this, we estimated Spearman correlation between (n)CTS measures and comprehension scores, after having removed the—potentially non-linear—effect of age.

2.10. Interrelationship between phrasal and syllabic (n)CTS

We used Spearman correlation to determine the degree of association between phrasal and syllabic (n)CTS values. This was done for each hemisphere separately, for CTS in the noiseless condition and for nCTS pooled across least- and most-energetic noises. In view of determining the role of maturation in these associations, the same analysis was repeated after having removed the—potentially non-linear—effect of age from CTS and nCTS values. Then, a 95% confidence interval was built for the difference between correlation coefficients (without vs. with correction for age) using bias corrected and accelerated bootstrap (Efron and Tibshirani, 1994).

2.11. Source reconstruction

As a preliminary step to estimate brain maps of CTS, MEG signals were projected into the source space. 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. When a participant's MRI was missing ($n = 39$), which happened mainly for the youngest participants, we used that of another participant of roughly the same age, which we linearly deformed to best match head-surface points using the CPD (Coherent Point Drift) toolbox (Myronenko and Song, 2010) embedded in FieldTrip toolbox (Donders Institute for Brain Cognition and Behaviour, Nijmegen, The Netherlands, RRID:SCR_004849) (Oostenveld et al., 2011). The individual MRIs were segmented using Freesurfer software (Martinos Center for Biomedical Imaging, Boston, MA, RRID:SCR_001847) (Reuter et al., 2012). Then, 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 and Friston, 1999; Ashburner et al., 1997). This transformation was used to map a homogeneous 5-mm grid sampling the MNI (Montreal Neurological Institute) 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 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. Source signals were then reconstructed with Minimum-Norm Estimates inverse solution (Dale and Sereno, 1993).

We followed a similar approach to that used at the sensor level to estimate source-level CTS. For each grid point, we trained a decoder on the two-dimensional source time-series to reconstruct speech temporal envelope. Again, the decoder was trained on the data from all but one condition, and used to estimate CTS in the left-out condition. To speed up computation, the training was performed without cross-validation, with the ridge value retained in a sensor-space analysis run on all gradiometer sensors at once. This procedure yielded a source map of CTS for each participant, condition, and frequency range of interest; and because the source space was defined on the MNI brain, all CTS maps were inherently corregistered with the MNI brain. Hence, group-averaged maps were simply produced as the mean of individual maps within age groups, conditions and frequency ranges of interest.

We further identified the coordinates of local maxima in group-averaged CTS maps. Such local maxima of CTS are sets of contiguous voxels displaying higher CTS values than all neighboring voxels. We only report statistically significant local maxima of CTS, 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; Bourguignon et al., 2018; Wens et al., 2015).

Note that 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 overall 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). This said, the precise anatomical location of anterior frontal and temporal opercular sources might be limited due to the greater deformation in those regions (Wilke et al., 2002).

2.12. Significance of local maxima of CTS

The statistical significance of the local maxima of CTS observed in group-averaged maps for each age group, condition and frequency range of interest was assessed with a non-parametric permutation test that intrinsically corrects for multiple spatial comparisons (Nichols and Holmes, 2002). First, participant and group-averaged *null* maps of CTS were computed with the MEG signals and the voice signal in each story rotated in time by about half of story length (i.e., the first and second halves were swapped, thereby destroying genuine coupling but preserving spectral properties). The exact temporal rotation applied was chosen to match a pause in speech to enforce continuity. Group-averaged difference maps were obtained by subtracting *genuine* and *null* group-averaged CTS maps. Under the null hypothesis that CTS maps are the same whatever the experimental condition, the labeling *genuine* or *null* 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 CTS 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 CTS values in the right auditory cortex could bias the permutation distribution and overshadow weaker CTS values in the left auditory cortex, even if these were highly consistent across subjects. Therefore, the permutation test described above was conducted separately for left- and right-hemisphere voxels.

2.13. Effect of age group and conditions on CTS source location

We evaluated for each frequency range if sources of CTS tended to cluster according to some categories, refraining from directly contrasting maps of CTS values (Bourguignon et al., 2018). Five different categories were considered: (i) age-group category (5 age groups), (ii) visual category (with vs. without visual input), (iii) 3-noise category (noiseless vs. non-speech noises vs. babble noises), (iv) 2-noise category (noiseless and non-speech noises vs. babble noises), and (v) presence of noise category (noiseless vs SiN). For this analysis, we gathered the coordinates of all sources of CTS in all conditions (8 SiN and 2 instances of noiseless speech). For each (target) source and category we computed the proportion of the 10 closest sources (excluding those for the same condition within the same age group as the target source) sharing the same category as the target source, we divided that proportion by that

expected by chance (i.e., the total number of sources sharing the same category as the target source divided by the total number of sources), subtracted 1, and multiplied by 100%. The mean of these values for a given category across all sources indicates the increase in chance (in percent; compared with what is expected by chance) of finding another CTS source of that category in the close vicinity. For statistical assessment, this mean value was compared with its permutation distribution where the CTS sources were assigned to random labels (1000 permutations).

3. Results

3.1. How does the CTS evolve with age in the absence of noise?

We assessed with an ANOVA if CTS in the noiseless condition depended on the hemisphere and on the age group (Table S1). Phrasal CTS was higher in the right (0.44 ± 0.09; mean ± SD across subjects) than in the left hemisphere (0.40 ± 0.08), and was not modulated by age. Syllabic CTS was also higher in the right (0.092 ± 0.036) than in the left hemisphere (0.079 ± 0.034), but a significant interaction with age indicated that left- and right-hemisphere CTS underwent different developmental trajectories.

Fig. 2 and Table 1 illustrate the developmental trajectory of syllabic CTS. While both left- and right-hemisphere CTS were similar in children aged below 7, a maturation process starting at 7.7 ± 1.7 years increased right- but not left-hemisphere CTS by ~30%, plateauing at 10.4 ± 1.9 years.

3.2. How does noise impact the CTS, and how does this impact evolve with age?

We assessed with an ANOVA if nCTS in noise conditions depended on noise properties, hemisphere, visibility of the talker's lips and whether they evolve with age (Table S2).

Fig. 3 summarizes the results involving other factors than age for phrasal and syllabic nCTS. Overall, the impact of the different types of background noises was similar for phrasal and syllabic nCTS: while non-speech noises did not affect much CTS (nCTS was close to zero), babble noises substantially reduced CTS compared to non-speech and noiseless conditions. Contrastingly, the level of energetic masking introduced by either type of noise only mildly affected the nCTS. Such pattern was observed for both hemispheres, and irrespective of the availability of visual speech information. Nevertheless, phrasal nCTS in babble noise conditions was higher in the left than right hemisphere (−0.16 ± 0.20 vs. −0.20 ± 0.19) while the reverse was true for syllabic nCTS in all noise conditions (−0.10 ± 0.24 vs. −0.07 ± 0.22).

A beneficial effect of visual speech information was observed in all noise conditions except in the least challenging one (i.e., least-energetic non-speech) for phrasal nCTS, and in all noise conditions for syllabic nCTS. The way visual speech information modulated nCTS was stable across the age range ($ps > 0.2$ for interactions involving age and type of visual input).

Critically, the way different noises impacted both phrasal and syllabic nCTS differed between age groups (i.e., significant interactions involving age and noise), with an additional dependence on the hemisphere for syllabic but not phrasal nCTS.

Fig. 4 and Table 1 present the developmental trajectories for nCTS pooled across least- and most-energetic noises. The detailed results for all noise conditions separately are presented in Supplementary material (Fig. S1, and Table S3).

Phrasal nCTS increased with age for both non-speech and babble noises. The modulation in CTS was however only marginal in non-speech noise conditions (4.2% with a transition at 8.6 ± 0.0 years following a logistic model), to the point that our model-fitting approach did not deem the age-dependent models better than a constant model. This suggests a minimal evolution of phrasal nCTS in non-speech noises, at least at a SNR of 3 dB. As a slight nuance, the evolution was clearer when considering the most-energetic non-speech noise, and fully absent for the least-energetic non-speech noise (Fig. S1 and Table S3). In contrast, a clear maturation process starting at 5.4 ± 1.6 years increased phrasal nCTS in babble noises by ~79%, with a plateau at age 9.3 ± 0.9 years.

Syllabic nCTS also increased with age, following linear trajectories, with different patterns observed for non-speech and babble noises in the left and right hemispheres. That is, syllabic nCTS increased with age in both hemispheres for babble noise, and only in the left hemisphere for non-speech noise.

3.3. How does speech comprehension evolve with age, in noiseless and noise conditions?

Fig. 5 illustrates speech comprehension abilities in the different conditions, which were assessed using yes/no forced-choice questions after each video. Comprehension scores were computed as the percentage of correct answers to 4 questions in each noise condition (or 8 in the noiseless condition).

Comprehension scores differed between age groups in noiseless ($F(4,137) = 14.0$, $p < 0.0001$) and noise conditions ($F(4,137) = 26.1$, $p < 0.0001$), improving with age in both cases. The model fitting approach identified a sharp transition at 6.9 years for comprehension in silence, and absurd values (negative transition age) for comprehension in noise. Comprehension in noise conditions was also affected by noise

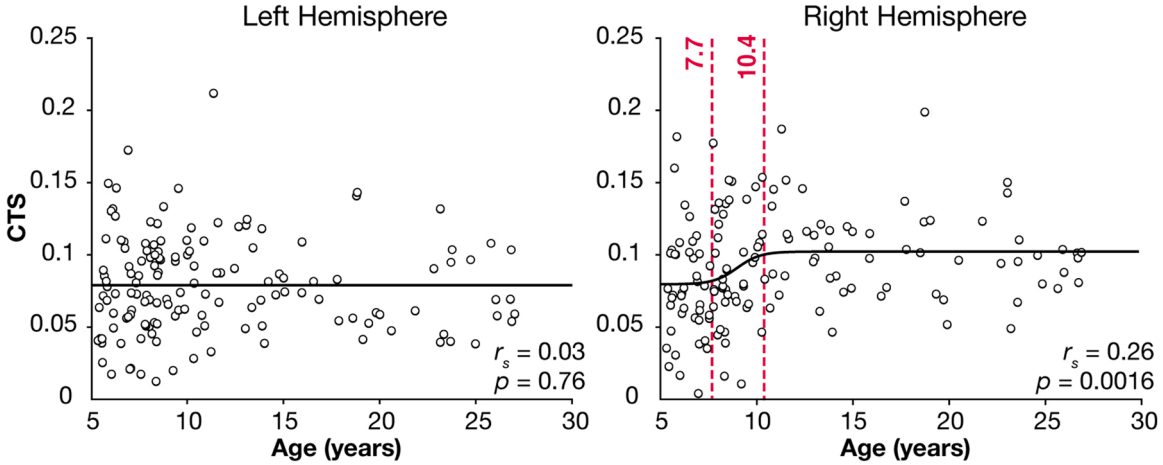


Fig. 2. Dependence on age of syllabic CTS in the noiseless condition. Dashed red lines indicate the beginning and the end of the maturation process.

Table 1
Parametric models of the dependence on age of (n)CTS values and speech comprehension scores. The number of participants (*n*) on which models were fitted was 144 for (n)CTS values and 142 for speech comprehension scores. Values of normalized CTS (nCTS) were pooled across conditions with and without visual speech, and across least- and most-energetic conditions. Values of phrasal nCTS were further pooled across hemispheres.

	Constant vs. Linear		Linear vs. Logistic		Constant vs. Logistic		Model
	F (1,*n*-2)	*p*	*F* (2,*n*-4)	*p*	*F* (3,*n*-4)	*p*	
Syllabic CTS in noiseless							
left hemisphere	0.17	0.69	1.87	0.16	1.30	0.28	0.080
right hemisphere	**3.98**	**0.048**	**4.18**	**0.017**	**4.17**	**0.0073**	0.079 + 0.023/(1 +exp(−1.6(age-9.0)))
Phrasal nCTS							
non-speech	2.06	0.15	2.09	0.10	2.09	0.13	–0.01 + 0.02/(1 +exp(–787(age–8.6)))
babble	**31.4**	**< 0.0001**	**20.5**	**< 0.0001**	**12.5**	**< 0.0001**	–0.37 + 0.27/(1 +exp(–1.1(age–7.3)))
Syllabic nCTS							
non-speech, left hem	**4.94**	**0.028**	0.77	0.47	**2.15**	**0.096**	–0.035 + 0.0039 × age
babble, left hem	**7.99**	**0.0054**	1.51	0.22	**3.69**	**0.014**	–0.25 + 0.0067 × age
non-speech, right hem	1.86	0.17	0.24	0.79	0.77	0.51	0.027
babble, right hem	**4.18**	**0.043**	0.002	1.00	1.38	0.25	–0.20 + 0.0049 × age
Speech (story) comprehension scores							
noiseless	**20.7**	**< 0.0001**	**18.6**	**< 0.0001**	**21.0**	**< 0.0001**	0.79 + 0.16/(1 +exp(–770(age–6.9)))
noise	**35.0**	**< 0.0001**	**94.3**	**< 0.0001**	**76.3**	**< 0.0001**	-28755.83 + 28756.78/(1 +exp(–0.87(age+7.8)))
difference Lip-Vid	***3.07***	***0.082***	***2.84***	***0.062***	**2.94**	**0.035**	–0.065 + 0.086/(1 +exp(–1.2(age–7.3)))

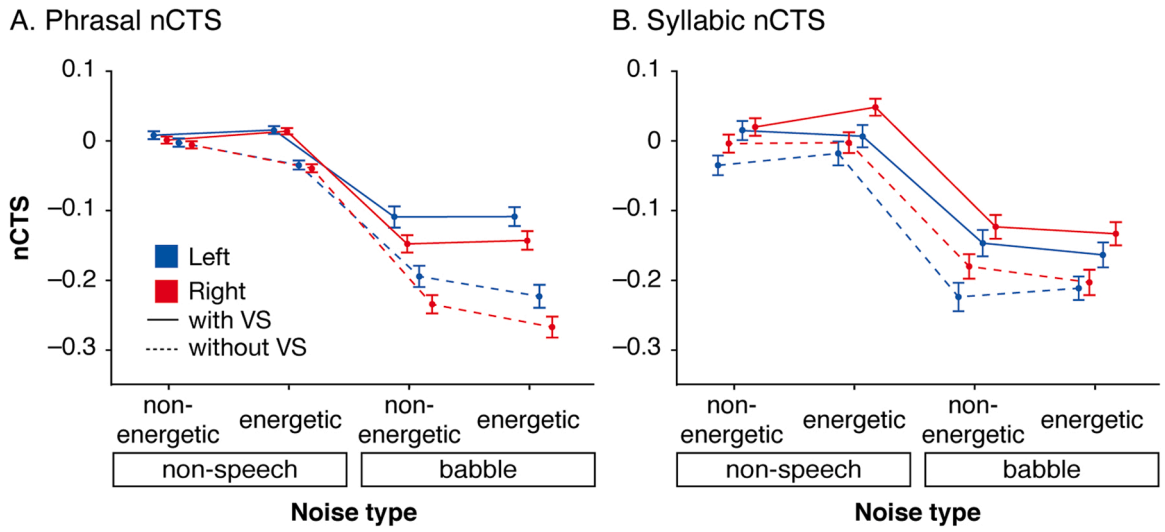


Fig. 3. Impact of the main effects on nCTS at phrasal (A) and syllabic rates (B). Mean and SEM (Standard Error of the Mean) values are displayed as a function of noise properties. The four traces correspond to conditions with (connected traces) and without (dashed traces) visual speech (VS), within the left (blue traces) and right (red traces) hemispheres. nCTS values are bounded between –1 and 1, with values below 0 indicating lower CTS in speech-in-noise conditions than in noiseless conditions.

properties ($F(3,411) = 7.31, p < 0.0001$). It was better when non-speech (92.3 ± 15.3%) compared to babble noises (88.5 ± 17.8%) were presented in the background. In fact, comprehension in non-speech noise conditions was similar ($t(141) = -0.37, p = 0.72$) to that in the noiseless condition (92.0 ± 11.2%), indicating that non-speech noise had no detrimental effect on the comprehension of the story (for our 3-dB SNR level). Finally, a marginally significant interaction between visual input and age group ($F(4,137) = 2.34, p = 0.059$) suggested that the benefit of visual speech for speech comprehension differed between age groups. The exploration of the boost in comprehension induced by visual speech is presented in Supplementary Fig. S2. No other significant effects were disclosed ($p > 0.1$).

3.4. Behavioral relevance of the CTS

Next, we used Spearman correlation to appraise the behavioral relevance of the tracking measures showing significant maturation effects (i.e., syllabic CTS in the right hemisphere, phrasal nCTS in babble noise, and syllabic nCTS in each hemisphere and in non-speech and babble noise conditions separately).

In the noiseless condition, this analysis revealed no significant association between syllabic CTS and speech comprehension ($ps > 0.3$ for left- and right-hemisphere CTS).

In SiN conditions, this analysis revealed a positive correlation between phrasal nCTS (averaged across hemispheres) and speech comprehension in babble noise conditions ($r_S = 0.22$; $p = 0.0074$; see Fig. 6), and no significant association between syllabic nCTS and speech comprehension, neither in non-speech ($ps > 0.5$ for left- and right-hemisphere CTS) nor in babble noise conditions ($ps > 0.9$).

3.5. Interrelationship between phrasal and syllabic (n)CTS

We assessed the degree of association between phrasal and syllabic (n)CTS with Spearman correlation (Table S4). This analysis revealed that phrasal and syllabic CTS in the noiseless condition were significantly positively related. Phrasal and syllabic nCTS values were also significantly positively related in the babble noise conditions but essentially not in the non-speech conditions. The degree of association was however limited, with maximum correlation coefficients of 0.36. The age factor did not appear to play a prominent role in these

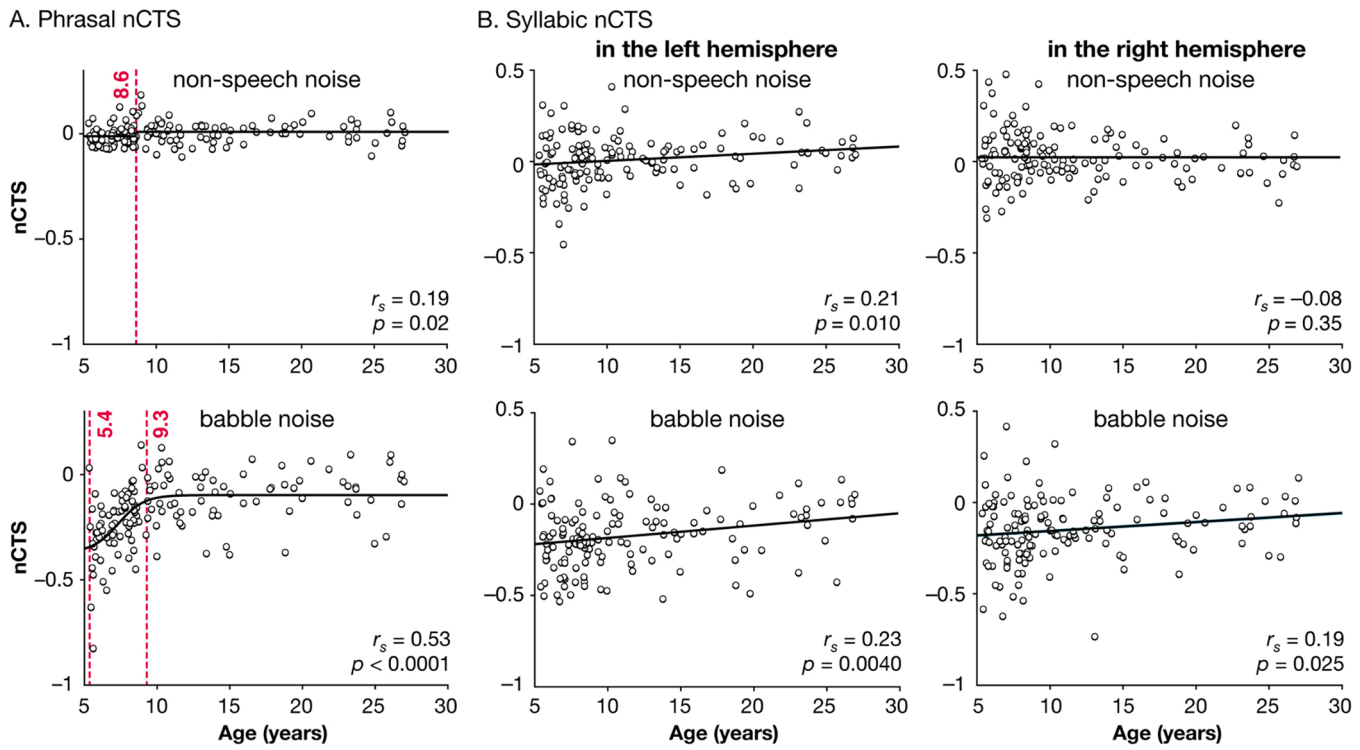


Fig. 4. Dependence on age of phrasal (A) and syllabic (B) nCTS. nCTS was pooled across conditions with and without visual speech, and across least- and most-energetic conditions. Phrasal nCTS was further pooled across hemispheres. Dashed red lines indicate the beginning and end of the maturation process.

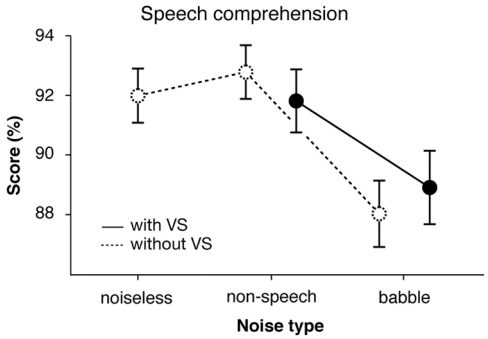


Fig. 5. Impact of the noise condition and of the presence or absence of concomitant visual speech on the speech comprehension scores, pooled across age groups. Vertical bars indicate SEM values.

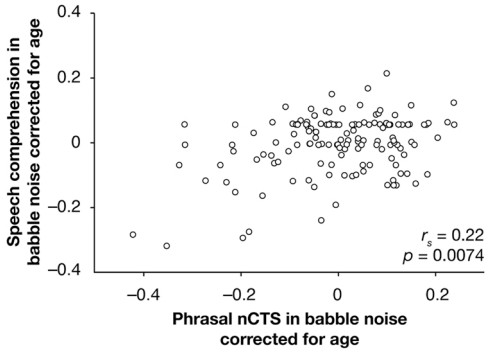


Fig. 6. Behavioral relevance of phrasal nCTS in babble noise. The speech comprehension scores and phrasal nCTS values were corrected for their nonlinear dependence on age. Presented values are positive when the initial values were above those of the fitted model and negative otherwise.

associations. Indeed, correcting for age had a significant impact only on the relation for left-hemisphere nCTS values in babble noise.

3.6. Sources of the CTS

We next identified the cortical sources underlying the CTS.

Fig. 7 presents the grand average CTS map (mean across all factors), together with the location of the significant sources of CTS in all conditions. Globally, sources of phrasal CTS localized bilaterally in the mid-superior temporal gyrus (STG), in the ventral part of the inferior frontal gyrus (IFG; in partes opercularis, triangularis and orbitalis) and pre-central gyrus and, to a lower extent, in posterior temporal regions. Sources of syllabic CTS essentially localized bilaterally in tight clusters centered around Heschl gyrus and in the anterior part of the IFG (partes orbitalis and triangularis) and, for few of them, in the temporoparietal junction (TPJ) and inferior part of the precentral gyrus.

Next, we evaluated for each frequency range if sources of CTS tended to cluster according to age group, or different noise properties.

First, sources of phrasal CTS had among their 10 closest neighbors 62.9% more sources for the same age group than expected by chance ($p < 0.0001$). To better understand this effect, Fig. 8A presents the sources of phrasal CTS color-coded by age group. Sources in the right hemisphere tended to localize more posteriorly with increasing age. Other differences were more subtle and not characterized by clear age gradients or source presence from or before a given age (e.g., sources in the right posterior temporal region were not seen in age groups of 7–8.5 years and 18–27 years; sources in precentral gyri were not seen in age groups of 5–7 years and 11.5–18 years).

Second, sources of phrasal CTS for (i) babble noise conditions on the one hand, and (ii) non-speech noise and noiseless conditions on the other hand, had among their 10 closest neighbors 66.1% more sources for the same category (i.e., i or ii) than expected by chance ($p < 0.0001$). Fig. 8B presents the sources of phrasal CTS color-coded for the informational property of the noise. Sources in bilateral STG and IFG were more anterior for babble noise conditions than for non-speech noise and

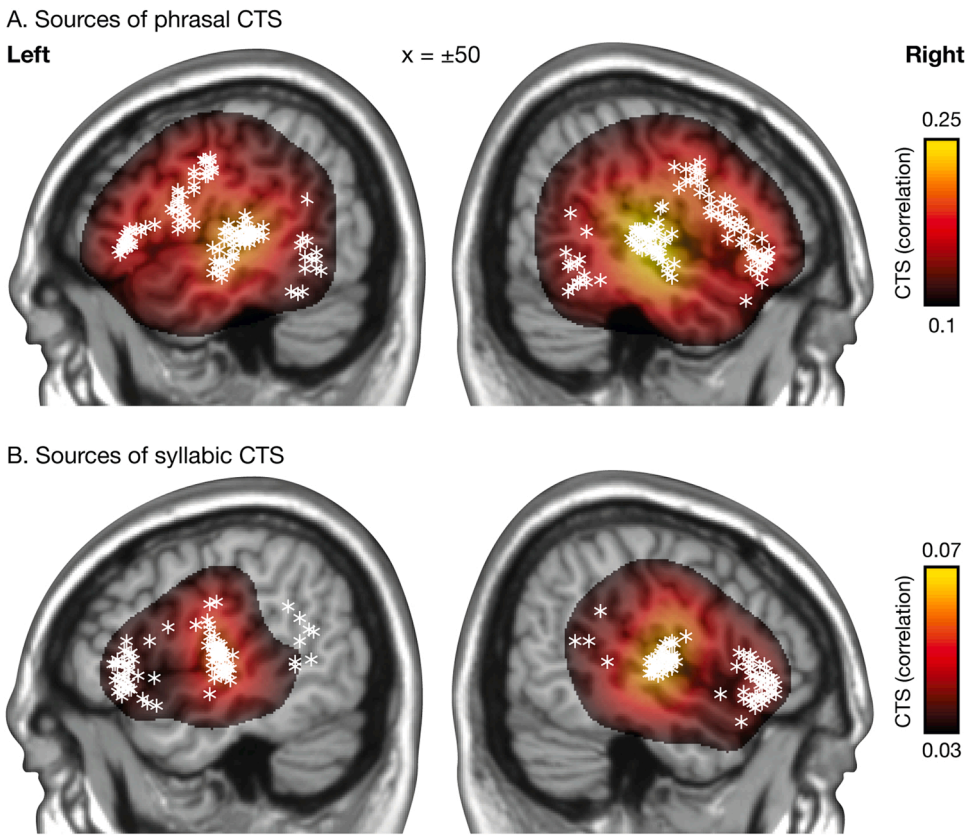


Fig. 7. Sources of phrasal (A) and syllabic (B) CTS in the left and right hemispheres. The overlays present the mean CTS values across all conditions and participants (regardless of age). Values at MNI coordinates |X|> 25 mm were projected orthogonally onto the parasagittal slice of coordinates |X| = 50 mm. The location of each significant source of CTS in each condition and age group is indicated with a white star (with the same projection scheme).

noiseless conditions. Sources in bilateral IFG localized in the pars orbitalis/triangularis for babble noise conditions, and in the pars triangularis/opercularis as well as in the inferior part of the precentral gyrus for non-speech noise and noiseless conditions. Finally, all sources in the posterior temporal areas were from babble noise conditions except for 3 right-sided sources from non-speech noise conditions.

Third, sources of syllabic CTS had among their 10 closest neighbors 76.8% more sources for the same age group than expected by chance ($p < 0.0001$). To better understand this effect, Fig. 8C presents the sources of syllabic CTS color-coded by age group. Paralleling the effect found for phrasal CTS, sources of syllabic CTS in the right hemisphere tended to localize increasingly more posteriorly with increasing age. Other subtler effects included the absence of source in TPJ in the oldest age group (18-27 years old), and more scattered source distributions along the ventrodorsal axis in the left Heschl gyrus in the two oldest age groups (11.5–18 and 18–27; sources reached the ventral bank of the STG and the ventral part of the postcentral gyrus).

Phrasal and syllabic CTS sources showed no significant tendency to cluster according to visual category ($p > 0.05$), and syllabic CTS sources showed no significant tendency to group according to noise categories ($p > 0.05$).

4. Discussion

This study characterizes the maturation of neurophysiological markers of the perception and understanding of natural connected speech in silence and in noise with or without visual speech information. Our results highlight that phrasal and syllabic CTS are rather distinct entities characterized by distinct developmental trajectories. While phrasal CTS in quiet conditions is adult-like from at least 5 years of age, syllabic CTS matures later in childhood. We also demonstrated two distinct neuromaturational effects related to the ability to perceive speech in babble noise: while the ability to maintain phrasal CTS matures rapidly between ~5 and ~9 years, a much slower maturation process improves the ability to maintain syllabic CTS in babble noise through childhood and into early adulthood. Visual speech information increased phrasal CTS mainly in babble noise conditions and syllabic CTS similarly in all noise conditions. These effects were not modulated by age. The results also reveal a limited impact of age on the cortical sources of phrasal and syllabic CTS.

4.1. Increase of syllabic but not phrasal CTS in silence during childhood

Our data revealed different developmental trajectories related to the capacity of the brain to track the fluctuations of speech temporal envelope at different frequencies, in noiseless conditions. While phrasal CTS is adult-like from at least 5 years of age, syllabic CTS matures later, between 7.5 and 10.5 years. This difference in developmental trajectory is well in line with grain-size proposals (Anthony and Francis, 2005; Goswami and Ziegler, 2006) extended to linguistic units we have proposed. Indeed, phrasal CTS is considered to partly reflect prosodic (Bourguignon et al., 2013; Keitel et al., 2018) and linguistic (Ding et al., 2016; Kaufeld et al., 2020) processing of large speech units. Contrastingly, syllabic CTS would reflect parsing of syllable rhythms (Ding et al., 2016), and the sensitivity to this basic unit of speech (Ghitza, 2013) would be at the basis of efficient phonemic processing (Giraud and Poeppel, 2012). Importantly, syllabic CTS is considered a lower-level process tightly related to the acoustic features of the auditory input (Ding and Simon, 2014; Molinaro and Lizarazu, 2018). The limited degree of interrelationship between phrasal and syllabic (n)CTS we observed further supports the view that these two aspects of CTS tag distinct processes.

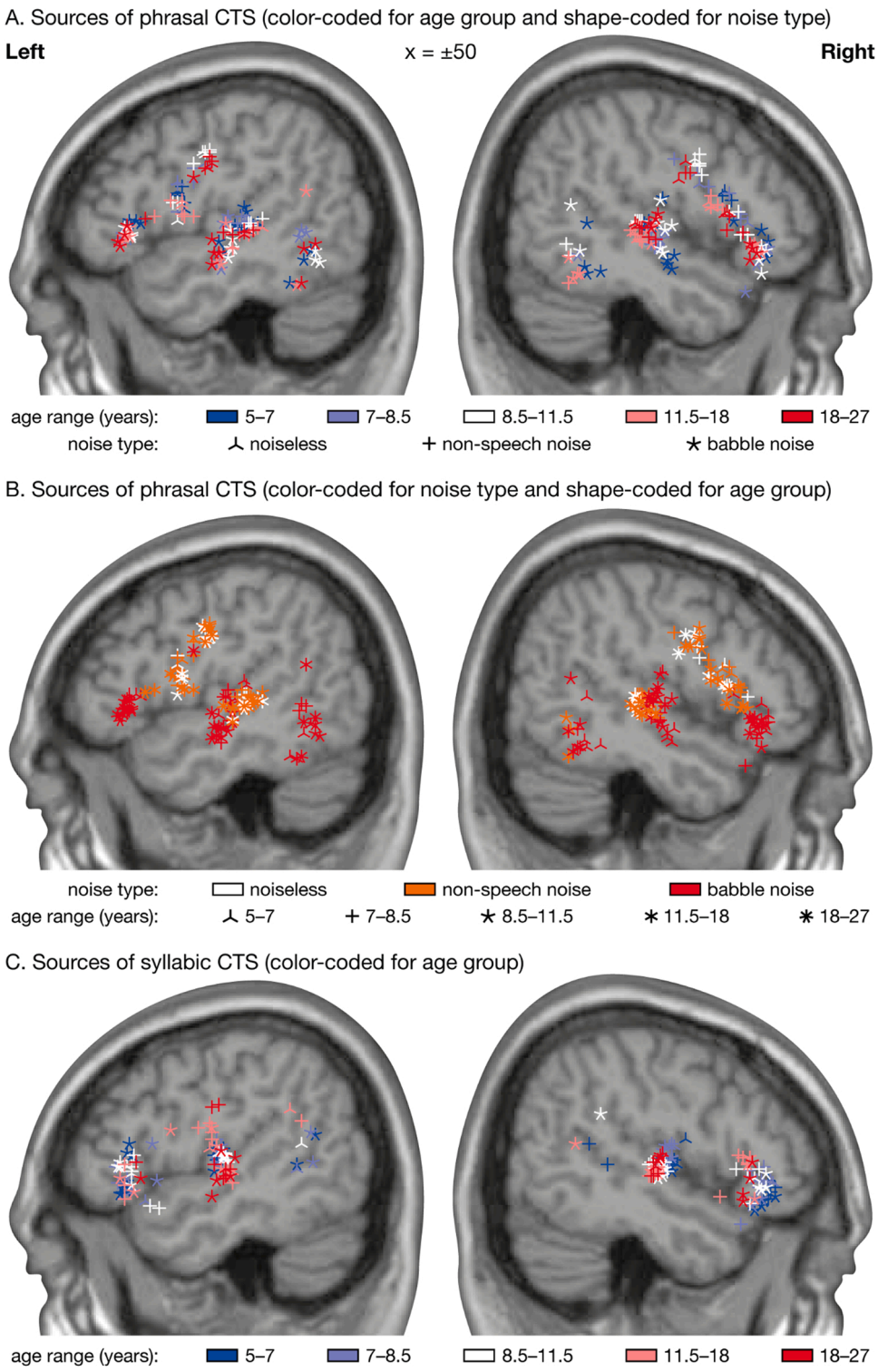


Fig. 8. Sources of CTS color-coded for age group (A, phrasal; C, syllabic) and for the informational property of the noise (B, phrasal); the other property being shape-coded.

Our finding that phrasal CTS is adult-like from at least 5 years of age supports the view that tracking of slow phrasal and prosodic stress patterns is a foundational process that might be present since birth (Attaheri et al., 2021), and remains stable across middle adulthood (Decruy et al., 2019). This result is not so surprising if one embraces the view that phrasal CTS partly underpins prosodic speech processing (Bourguignon et al., 2013). Indeed, young infants already use such information to parse speech into words and phrases (Speer and Ito, 2009). Accordingly, other neurophysiological markers of brain processing of prosody in speech (i.e., closure positive shift) were reported in 6-months old in relation to brief pause detection but also to pitch variations (Holzgrefe-Lang et al., 2018). The result of stable phrasal CTS from 5 years on is also compatible with the view that phrasal CTS reflects lexical and syntactic processing. Indeed, typically developing children of that age possess basic syntactic skills (Lee et al., 2020) such as the ability to understand relative clauses (Corrêa and Sicuro, 1995; Kidd and Bavin,

2002). In addition, the stability of phrasal CTS across the age range we investigated indicates that potential methodological differences between children and adults—such as head size and head movements—did not significantly impact our CTS estimates.

We found evidence for a developmental boost in syllabic CTS in quiet conditions in the right- (but not left-) hemisphere. This boost mostly occurred between the age of 7.5 and 10.5 years, and signified the start of right-hemisphere dominance for syllabic CTS. This transition suggests that, although operational early in life (Attaheri et al., 2021; Ortiz Barajas et al., 2021), temporal parsing of the speech signal at the syllabic level refines with brain maturation. And indeed, children aged below 10 are less accurate than adults at identifying syllable boundaries when these are defined only by amplitude modulations in speech temporal envelope (Cameron et al., 2018). The right-hemispheric dominance in noiseless conditions observed for syllabic CTS after age ~10 and for phrasal CTS is consistent with previous findings in children and adults (Abrams et al., 2008, 2009; Bourguignon et al., 2013; Gross et al., 2013; Molinaro et al., 2016; Molinaro and Lizarazu, 2018; Ríos-López et al., 2020; Vander Ghinst et al., 2019). It is even at the core of the *asymmetric sampling in time hypothesis*, which argues that prosodic and syllabic information are preferentially processed in the right hemisphere, while phonemic information is preferentially processed in the left hemisphere or bilaterally (Poeppel, 2003). As previously argued (Giraud and Ramus, 2013), the fact that language brain functions become asymmetric in the course of development suggests asymmetry is a hallmark of maturity.

4.2. Development of the neurophysiological basis of speech perception in noise

Our results highlight two distinct neuromaturational effects related to the ability to perceive speech in babble noise. First, the ability to maintain phrasal CTS in babble noise matures rapidly between ~5 and ~9 years, with a marked transition at age ~7. Second, a much slower maturation process—best characterized by a linear progression with age—improves the ability to maintain syllabic CTS in babble noise through childhood and into early adulthood. Following the rationale developed in the previous subsection, our results indicate a rapid maturation at age ~7 of the neurophysiological mechanisms at play in processing prosodic and suprasegmental linguistic information in natural connected speech in babble noise, and a slower, progressive maturation into early adulthood of the mechanisms involved in the extraction of hierarchically lower syllabic, phonemic or even acoustic information from speech in babble noise. This is well in line with our working hypotheses: neuronal processing of larger linguistic units (words and phrases) develops before that of smaller syllabic units, and coping with noise necessitates additional processes that mature later on.

The maturational time-course of the ability to maintain phrasal CTS in babble noise closely parallels that of the ability to recognize words in the presence of two-talker speech. The latter improves progressively from 5 to 10 years of age, reaching adult-like levels at age 11 (Hall et al., 2002; Leibold and Buss, 2013; Bolia et al., 2000). This maturation trajectory is specific to informational noise since speech recognition in speech-shaped noise is close to adult-like already at age 5 (Hall et al., 2002; Leibold and Buss, 2013), as was phrasal CTS in non-speech noise in our data. This suggests that maintenance of phrasal CTS reflects a range of processes involved in the ability to perceive and understand linguistic chunks larger than syllables or words in babble noise. This interpretation is further supported by the similar developmental trajectory of our measures of SiN comprehension, and by the finding that CTS resistance to babble noise is positively related to speech comprehension after having accounted for age.

The degree of maintenance of phrasal CTS in babble noise could actually underpin the maturation of auditory stream formation, which is the process of grouping together sounds from the same source (Darwin and Hukin, 1999). Forming auditory streams is a challenging aspect of speech perception in noise, and failure in forming streams seems to explain the behavioral difficulties understanding speech from among two same-gender talkers (Bronkhorst, 2015). From the point of view of development, the ability to form auditory streams based on frequency separation appears to be immature at age 5–8, and adult-like at age 9–11 (Sussman et al., 2007). Since these developmental milestones match well with those found for phrasal CTS in babble noise in our study, the way babble noise impacts phrasal CTS could represent an electrophysiological signature of the ability to form auditory streams.

The slow maturation of the ability to maintain syllabic CTS in noise closely parallels the evolution of phonemic perception in noise. Although such slow evolution was not evident in our phonemic perception test (see Supplementary Material), it is clearly seen in normative data for this test where twice more items were used to assess an even larger sample of participants than ours (Demanez et al., 2003). In that study, phonemic perception in noise improved steadily from age 5, topped in the 15–19 year group, and then decreased in the subsequent age ranges, the first of which was overly broad (20–49 years) unfortunately. Our data therefore provide a neurophysiological ground for the slow maturation of phonemic perception in noise. It also suggests a more important role of the left hemisphere since maturation was not observed in the non-speech noise condition in the right hemisphere. This is well in line with the classical dominant role of the left hemisphere for language comprehension.

4.3. Impact of visual speech on CTS

Our data did not reveal any evidence for a maturation of the boost in CTS afforded by visual speech across the tested age range. This is somewhat surprising since audiovisual integration processes mature rather slowly, with adult-like performance reached after age 12 for some tasks (Ross et al., 2011; Wightman et al., 2006; Barutchu et al., 2010). A potential explanation could be that the SNR we used (3 dB) was too high to confer a significant advantage to older participants (Lalonde and Werner, 2021).

The analysis of our comprehension scores hinted at a transition between age 6 and 9 in the ability to leverage visual speech to enhance comprehension. This is in line with the observation that at around 6.5 years of age, children start to benefit from having phonetic knowledge about severely degraded speech sounds when asked to match such a sound with a visual speech video (Baart et al., 2015). Possibly then, a CTS boost induced by visual speech may be driven by audiovisual congruence detection, an ability that is already observed at 2 months of age (Patterson and Werker, 2003; Dodd, 1979). Although this suggestion provides an interesting avenue for future work, it is currently rather tentative as processing congruence in audiovisual speech seems to start at around 200 ms (Stekelenburg and Vroomen, 2007), can take several hundreds of milliseconds (Baart et al., 2017; Arnal et al., 2009) and therefore overlaps in time with other processes (such as processing of lexico-semantic information) that are difficult to disentangle.

In the S1 discussion, we elaborate further on the beneficial effects of visual speech on CTS we observed across all age ranges.

4.4. Recruited neural network and impact of maturation and noise

Our results showed that source configuration was affected by age for both phrasal and syllabic CTS, and by informational noise properties for phrasal CTS.

The effect of age for both phrasal and syllabic CTS appeared to be mainly explained by an anterior-to-posterior shift (of about 1 cm) of right-hemisphere sources from youngest (5–7 years) to oldest (18–27 years) age groups. Whether this shift reflects a genuine developmental effect is difficult to tell since changes in brain anatomy from childhood to adulthood induce small, but consistent, age-dependent errors in the normalization of individual brains to a template (Wilke et al., 2002). Besides these unclear effects of age, our results rather emphasize the close similarity in location of cortical generators of CTS across the

investigated age range. This is in line with a host of findings indicating that the architecture of the language network is settled from age 3, with subsequent maturation essentially refining bottom-up communication and specialization of each node of the network (Skeide and Friederici, 2016).

In S2 Discussion, we discuss the interesting effect of noise properties on the configuration of CTS sources.

4.5. Limitations

We manipulated several properties of the noise but not all of those known to impact SiN perception. It is therefore worth noting that the developmental trajectories we report for CTS resistance to noise are valid only for the conditions we explored, and might be affected by other aspects of the listening condition, much like the maturation timeline of behavioral effects depends on the number of speakers making up the background noise (Hoen et al., 2007), noise intensity (Wightman and Kistler, 2005), or availability of spatial cues (Wightman et al., 2010).

Furthermore, our results were obtained in native French speakers with various levels of proficiency in other languages (mainly Dutch, which is taught in most French-speaking Belgian schools). It would be interesting to determine whether similar developmental trajectories are observed in other languages than French, which is a syllable-timed alphabetic language with a highly predictable stress pattern. Indeed, it has been hypothesized that multiple language properties could impact CTS (Lallier et al., 2017). Multilingualism may also affect the way humans perceive their own language (Gorba, 2019). However, these effects are rather subtle (Gorba, 2019) and in a previous study, the degree of proficiency in a second language did not affect CTS in the native language (Lizarazu et al., 2021). This makes it unlikely that the diversity in our participants' experience with other languages impacted our results.

We have used natural connected speech as auditory material. Although this adds to the ecological validity of our results, it makes it difficult to resolve the development of brain functions supporting multiple distinct aspects of language. For example, phrasal CTS taps in brain function supporting linguistic (syntactic, lexical, grammatical) as well as paralinguistic (prosody) information. Studies relying on carefully synthesized speech in which, e.g., prosody is removed would be needed in this regard (Ding et al., 2016).

Characterization of behavioral performance was suboptimal. We only asked simple comprehension questions, and comprehension scores suffered ceiling effects. This may explain why the link between CTS and behavior, which has been well documented in other studies (Ahissar et al., 2001; Luo and Poeppel, 2007; Peelle et al., 2013; Vanthornhout et al., 2018), was either weak (for phrasal nCTS) or non-significant (for syllabic nCTS) in our study despite having a sample size ($n = 144$) that largely surpasses that of previous studies. A more extensive neuropsychological assessment of language processing abilities could have further supported the behavioral relevance of the multiple developmental effects we identified.

No MRI was available for 27% of the participants, and for those, source reconstruction was performed based on another participant's MRI linearly deformed to match the digitized head surface. Using a well-matched MRI was reported to lead to maximal errors of about 1 cm (Gohel et al., 2017). However, since the direction of the error is expected to be random across participants, the overall effect of missing MRIs on group-level maps of CTS should be akin to that of an additional smoothing. This is unlikely to have affected our results since our source-space analysis focused on source localization rather than peak CTS amplitude.

Finally, we have used a cross-sectional design to characterize the developmental trajectories of CTS. Adopting a longitudinal design, which is more logistically challenging, could have proved more insightful.

5. Conclusion

This study reveals distinct developmental trajectories for the neuronal processing of prosodic/syntactic (phrasal CTS) and syllabic/phonemic/acoustic information (syllabic CTS), and depending on the presence and the type of background noise. Overall, our results indicate that cortical processing of large linguistic units matures before that of smaller units, and that additional neuromaturational milestones need reaching for such processing to be optimal in adverse noise conditions. Unexpectedly, although visual speech information boosted the ability of the brain to track speech in noise, such boost was not affected by brain maturation. Finally, the ability to maintain phrasal tracking in noise was positively related to speech comprehension. These results therefore indicate that CTS tags behaviourally relevant neural mechanisms that progressively mature with age and experience following the trajectory presumed by grain-size proposals. Thus, the modulation of CTS by noise provides objective neurodevelopmental markers of multiple aspects of speech processing in noise.

CRediT authorship contribution statement

Conceptualization: JB, FD, TC, MVG, MBa, NM, XDT, MBo. Methodology: JB, MN, VW, MBo. Investigation: JB, MN, FD, AR, NT, MBo, Supervision: XDT, MBo, Writing—original draft: JB, MN, TC, Writing—review & editing: FD, MVG, VW, AR, NT, MBa, NM, XDT, MBo.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data underlying the study are available at https://osf.io/4uzrm/.

Acknowledgments

Florian Destoky, Julie Bertels and Mathieu Bourguignon 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). Maxime Niesen has been supported by the Fonds Erasme (Brussels, Belgium). Xavier De Tiège is Clinical Researcher at the Fonds de la Recherche Scientifique (F.R.S.-FNRS, Brussels, Belgium). Mathieu Bourguignon has been supported by the Marie Skłodowska-Curie Action of the European Commission (grant 743562). The MEG project at the CUB Hôpital Erasme and this study were financially supported by the Fonds Erasme (Research convention "Les Voies du Savoir", Brussels, Belgium). The PET-MR project at the CUB Hôpital Erasme is supported by the Association Vinçotte Nuclear (AVN, Brussels, Belgium).

Appendix A. Supplementary information

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dcn.2022.101181.

References

Abrams, D.A., Nicol, T., Zecker, S., Kraus, N., 2008. Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. J. Neurosci. 28, 3958–3965.
Abrams, D.A., Nicol, T., Zecker, S., Kraus, N., 2009. Abnormal cortical processing of the syllable rate of speech in poor readers. J. Neurosci. 29, 7686–7693.
Ahissar, E., et al., 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc. Natl. Acad. Sci. U. S. A. 98, 13367–13372.
Altman, D.G., Royston, P., 2006. The cost of dichotomising continuous variables. BMJ 332, 1080.

Anthony, J.L., Francis, D.J., 2005. Development of Phonological Awareness. Curr. Dir. Psychol. Sci. 14, 255–259.
Arnal, L.H., Morillon, B., Kell, C.A., Giraud, A.-L., 2009. Dual neural routing of visual facilitation in speech processing. J. Neurosci. 29, 13445–13453.
Ashburner, J., Friston, K.J., 1999. Nonlinear spatial normalization using basis functions. Hum. Brain Mapp. 7, 254–266.
Ashburner, J., Neelin, P., Collins, D.L., Evans, A., Friston, K., 1997. Incorporating prior knowledge into image registration. Neuroimage 6, 344–352.
Attaheri, A., et al., 2021. Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. BioRxiv 329326.
Baart, M., Bortfled, H., Vroomen, J., 2015. Phonetic matching of auditory and visual speech develops during childhood: Evidence from sine-wave speech. Journal of Experimental Child Psychology 129, 157–164. https://doi.org/10.1016/j.jecp.2014.08.002.
Baart, M., Lindborg, A., Andersen, T.S., 2017. Electrophysiological evidence for differences between fusion and combination illusions in audiovisual speech perception. Eur. J. Neurosci. 46, 2578–2583.
Baillet, S., 2017. Magnetoencephalography for brain electrophysiology and imaging. Nat. Neurosci. 20, 327–339.
Barutchu, A., et al., 2010. Audiovisual integration in noise by children and adults. J. Exp. Child Psychol. 105, 38–50.
Biesmans, W., Das, N., Francart, T., Bertrand, A., 2017. Auditory-inspired speech envelope extraction methods for improved EEG-based auditory attention detection in a cocktail party scenario. IEEE Trans. Neural Syst. Rehabil. Eng. 25, 402–412.
Bolia, R.S., Nelson, W.T., Ericson, M.A., Simpson, B.D., 2000. A speech corpus for multitalker communications research. J. Acoust. Soc. Am. 107, 1065–1066.
Bourguignon, M., et al., 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.
Bourguignon, M., Baart, M., Kapnoula, E.C., Molinaro, N., 2019. Lip-reading enables the brain to synthesize auditory features of unknown silent speech. J. Neurosci. https://doi.org/10.1523/JNEUROSCI.1101-19.2019.
Bronkhorst, A.W., 2015. The cocktail-party problem revisited: early processing and selection of multi-talker speech. Atten. Percept. Psychophys. 77, 1465–1487.
Brungart, D.S., 2001. Informational and energetic masking effects in the perception of two simultaneous talkers. J. Acoust. Soc. Am. 109, 1101–1109.
Cameron, S., et al., 2018. The parsing syllable envelopes test for assessment of amplitude modulation discrimination skills in children: development, normative data, and test-retest reliability studies. J. Am. Acad. Audiol. 29, 151–163.
Corbin, N.E., Bonino, A.Y., Buss, E., Leibold, L.J., 2016. Development of open-set word recognition in children: speech-shaped noise and two-talker speech maskers. Ear Hear 37, 55–63.
Corrêa, L.M.S., Sicuro, L.M., 1995. Corrêa, An alternative assessment of children's comprehension of relative clauses. J. Psycholinguist. Res. 24, 183–203.
Crosse, M.J., Butler, J.S., Lalor, E.C., 2015. Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions. J. Neurosci. 35, 14195–14204.
Crosse, M.J., Di Liberto, G.M., Lalor, E.C., 2016a. Eye can hear clearly now: inverse effectiveness in natural audiovisual speech processing relies on long-term crossmodal temporal integration. J. Neurosci. 36, 9888–9895.
Dale, A.M., Sereno, M.I., 1993. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. J. Cogn. Neurosci. 5, 162–176.
Darwin, C.J., Hukin, R.W., 1999. Auditory objects of attention: the role of interaural time differences. J. Exp. Psychol. Hum. Percept. Perform. 25, 617–629.
Darwin, C.J., Brungart, D.S., Simpson, B.D., 2003. Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. J. Acoust. Soc. Am. 114, 2913–2922.
De Tiège, X., et al., 2008. Recording epileptic activity with MEG in a light-weight magnetic shield. Epilepsy Res 82, 227–231.
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, 601–615.
Demanez, L., Dony-Closon, B., Lhonneux-Ledoux, E., Demanez, J.P., 2003. Central auditory processing assessment: a French-speaking battery. Acta Otorhinolaryngol. Belg. 57, 275–290.
Destoky, F., Philippe, M., Bertels, J., Verhasselt, M., Coquelet, N., Vander Ghinst, M., Wens, V., De Tiège, X., Bourguignon, M., 2019. Comparing the potential of MEG and EEG to uncover brain tracking of speech temporal envelope. Neuroimage 184, 201–213.
Destoky, F., et al., 2020. Cortical tracking of speech in noise accounts for reading strategies in children. PLoS Biol. 18, e3000840.
Destoky, F., et al., 2022. The role of reading experience in atypical cortical tracking of speech and speech-in-noise in dyslexia. Neuroimage 253, 119061.
Ding, N., Simon, J.Z., 2012. Emergence of neural encoding of auditory objects while listening to competing speakers. Proc. Natl. Acad. Sci. U. S. A. 109, 11854–11859.
Ding, N., Simon, J.Z., 2013. Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. J. Neurosci. 33, 5728–5735.
Ding, N., Simon, J.Z., 2014. Cortical entrainment to continuous speech: functional roles and interpretations. Front. Hum. Neurosci. 8, 311.
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.
Dodd, B., 1979. Lip reading in infants: attention to speech presented in- and out-of-synchrony. Cogn. Psychol. 11, 478–484.
Donhauser, P.W., Baillet, S., 2020. Two distinct neural timescales for predictive speech processing. e9 Neuron 105, 385–393. e9.
Efron, B., Tibshirani, R.J., 1994. An Introduction to the Bootstrap. CRC Press.
Elliott, L.L., 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.
Fjell M., A., Grydeland, H., Krogsrud K., S., Walhovd B., K., 2015. Development and aging of cortical thickness correspond to genetic organization patterns. PNAS 112 (50), 15462–15467.
Fuglsang, S.A., Dau, T., Hjortkjær, J., 2017. Noise-robust cortical tracking of attended speech in real-world acoustic scenes. Neuroimage 156, 435–444.
Ghitza, O., 2013. The theta-syllable: a unit of speech information defined by cortical function. Front. Psychol. 4, 138.
Giordano, B.L., et al., 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.
Giraud, A.-L., Ramus, F., 2013. Neurogenetics and auditory processing in developmental dyslexia. Curr. Opin. Neurobiol. 23, 37–42.
Gogtay, N., et al., 2004. Dynamic mapping of human cortical development during childhood through early adulthood. Proc. Natl. Acad. Sci. U. S. A. 101, 8174–8179.
Gohel, B., Lim, S., Kim, M.-Y., Kwon, H., Kim, K., 2017. Approximate Subject Specific Pseudo MRI from an Available MRI Dataset for MEG Source Imaging. Front. Neuroinformatics 11.
Gorba, C., 2019. Bidirectional influence on L1 Spanish and L2 English stop perception: The role of L2 experience. J. Acoust. Soc. Am. 145, EL587.
Goswami, U., 2015. Sensory theories of developmental dyslexia: three challenges for research. Nat. Rev. Neurosci. 16, 43–54.
Goswami, U., Ziegler, J.C., 2006. A developmental perspective on the neural code for written words. Trends Cogn. Sci. 10, 142–143.
Gramfort, A., et al., 2014. MNE software for processing MEG and EEG data. Neuroimage 86, 446–460.
Gross, J., et al., 2013. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS Biol. 11, e1001752.
Hall 3rd, J.W., Grose, J.H., Buss, E., Dev, M.B., 2002. Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. Ear Hear 23, 159–165.
Hämäläinen, M.S., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Med. Biol. Eng. Comput. 32, 35–42.
Hoen, M., et al., 2007. Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. Speech Commun. 49, 905–916.
Holzgrefe-Lang, J., Wellmann, C., Höhle, B., Wartenburger, I., 2018. Infants' processing of prosodic cues: electrophysiological evidence for boundary perception beyond pause detection. Lang. Speech 61, 153–169.
Horton, C., D'Zmura, M., Srinivasan, R., 2013. Suppression of competing speech through entrainment of cortical oscillations. J. Neurophysiol. 109, 3082–3093.
Kaufeld, G., et al., 2020. Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. J. Neurosci. 40, 9467–9475.
Keitel, A., Gross, J., Kayser, C., 2018. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. PLoS Biol. 16, e2004473.
Kidd, E., Bavin, E.L., 2002. English-speaking children's comprehension of relative clauses: evidence for general-cognitive and language-specific constraints on development. J. Psycholinguist. Res 31.
Lallier, M., Molinaro, N., Lizarazu, M., Bourguignon, M., Carreiras, M., 2017. Amodal atypical neural oscillatory activity in dyslexia. Clin. Psychol. Sci. 5, 379–401.
Lalonde, K., Werner, L.A., 2021. Development of the mechanisms underlying audiovisual speech perception benefit. Brain Sci. 11.
Lalor, E.C., Foxe, J.J., 2010. Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. Eur. J. Neurosci. 31, 189–193.
Larson, E., Taulu, S., 2017. The importance of properly compensating for head movements during MEG acquisition across different age groups. Brain Topogr. 30, 172–181.
Lee, Y.S., Ahn, S., Holt, R.F., Schellenberg, E.G., 2020. Rhythm and syntax processing in school-age children. Dev. Psychol. 56, 1632–1641.
Leibold, L.J., 2017. Speech perception in complex acoustic environments: developmental effects. J. Speech Lang. Hear. Res. 60, 3001–3008.
Leibold, L.J., Buss, E., 2013. Children's identification of consonants in a speech-shaped noise or a two-talker masker. J. Speech, Lang., Hear. Res. 56, 1144–1155.
Lizarazu, M., Carreiras, M., Bourguignon, M., Zarraga, A., Molinaro, N., 2021. Language Proficiency Entails Tuning Cortical Activity to Second Language Speech. Cereb. Cortex 31, 3820–3831.
Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54, 1001–1010.
Luo, H., Liu, Z., Poeppel, D., 2010. Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. PLoS Biol. 8, e1000445.
MacCallum, R.C., Zhang, S., Preacher, K.J., Rucker, D.D., 2002. On the practice of dichotomization of quantitative variables. Psychol. Methods 7, 19–40.
Mesgarani, N., Chang, E.F., 2012. Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485, 233–236.
Meyer, L., 2018. The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. Eur. J. Neurosci. 48, 2609–2621.
Meyer, L., Gumbert, M., 2018. Synchronization of electrophysiological responses with speech benefits syntactic information processing. J. Cogn. Neurosci. 30, 1066–1074.

Meyer, L., Henry, M.J., Gaston, P., Schmuck, N., Friederici, A.D., 2017. Linguistic bias modulates interpretation of speech via neural delta-band oscillations. Cereb. Cortex 27, 4293–4302.
Molinaro, N., Lizarazu, M., 2018. Delta(but not theta)-band cortical entrainment involves speech-specific processing. Eur. J. Neurosci. 48, 2642–2650.
Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., Carreiras, M., 2016. Out-of-synchrony speech entrainment in developmental dyslexia. Hum. Brain Mapp. 37, 2767–2783.
Muzik, O., Chugani, D.C., Juhász, C., Shen, C., Chugani, H.T., 2000. Statistical parametric mapping: assessment of application in children. Neuroimage 12, 538–549.
Myronenko, A., Song, X., 2010. Point set registration: coherent point drift. IEEE Trans. Pattern Anal. Mach. Intell. 32, 2262–2275.
Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15, 1–25.
O'Sullivan, J.A., et al., 2014. Attentional selection in a cocktail party environment can be decoded from single-trial EEG. Cereb. Cortex 25, 1697–1706.
Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011 (156869).
Ortiz Barajas, M.C., Guevara, R., Gervain, J., 2021. The origins and development of speech envelope tracking during the first months of life. Dev. Cogn. Neurosci. 48, 100915.
Park, H., Ince, R.A.A., Schyns, P.G., Thut, G., Gross, J., 2018. Representational interactions during audiovisual speech entrainment: Redundancy in left posterior superior temporal gyrus and synergy in left motor cortex. PLoS Biol. 16, e2006558.
Patterson, M.L., Werker, J.F., 2003. Two-month-old infants match phonetic information in lips and voice. Dev. Sci. 6, 191–196.
Peelle, J.E., Gross, J., Davis, M.H., 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb. Cortex 23, 1378–1387.
Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as "asymmetric sampling in time. Speech Commun. 41, 245–255.
Pollack, I., 1975. Auditory informational masking. S5–S5 J. Acoust. Soc. Am. 57. S5–S5.
Power, A.J., Colling, L.J., Mead, N., Barnes, L., Goswami, U., 2016. Neural encoding of the speech envelope by children with developmental dyslexia. Brain Lang. 160, 1–10.
Puschmann, S., et al., 2017. The right temporoparietal junction supports speech tracking during selective listening: evidence from concurrent EEG-fMRI. J. Neurosci. 37, 11505–11516.
Reiss, A.L., Abrams, M.T., Singer, H.S., Ross, J.L., Denckla, M.B., 1996. Brain development. Gend. IQ Child. *Brain* 119, 1763–1774.
Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 61, 1402–1418.
Rimmele, J.M., Golumbic, E.Z., 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.
Ríos-López, P., Molinaro, N., Bourguignon, M., Lallier, M., 2020. Development of neural oscillatory activity in response to speech in children from 4 to 6 years old. Dev. Sci. 23.
Ross, L.A., et al., 2011. The development of multisensory speech perception continues into the late childhood years. Eur. J. Neurosci. 33, 2329–2337.
Sanes, D.H., Woolley, S.M.N., 2011. A behavioral framework to guide research on central auditory development and plasticity. Neuron 72, 912–929.
Schwartz, J.-L., Berthommier, F., Savariaux, C., 2004. Seeing to hear better: evidence for early audio-visual interactions in speech identification. Cognition 93, B69–B78.
Shield, B.M., Dockrell, J.E., 2008. The effects of environmental and classroom noise on the academic attainments of primary school children. J. Acoust. Soc. Am. 123, 133–144.
Simon, J.Z., 2015. The encoding of auditory objects in auditory cortex: insights from magnetoencephalography. Int. J. Psychophysiol. 95, 184–190.
Skeide, M.A., Friederici, A.D., 2016. The ontogeny of the cortical language network. Nat. Rev. Neurosci. 17, 323–332.
Speer, S.R., Ito, K., 2009. Prosody in first language acquisition - acquiring intonation as a tool to organize information in conversation. Lang. Linguist. Compass 3, 90–110.
Stekelenburg, J.J., Vroomen, J., 2007. Neural correlates of multisensory integration of ecologically valid audiovisual events. J. Cogn. Neurosci. 19, 1964–1973.
Sumby, W.H., Pollack, I., 1954. Visual contribution to speech intelligibility in noise. J. Acoust. Soc. Am. 26, 212–215.
Sussman, E., Wong, R., Horváth, J., Winkler, I., Wang, W., 2007. The development of the perceptual organization of sound by frequency separation in 5–11-year-old children. Hear. Res. 225, 117–127.
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.
Vander Ghinst, M., et al., 2016. Left Superior Temporal Gyrus Is Coupled to Attended Speech in a Cocktail-Party Auditory Scene. J. Neurosci. 36, 1596–1606.
Vander Ghinst, M., et al., 2019. Cortical Tracking of Speech-in-Noise Develops from Childhood to Adulthood. J. Neurosci. 39, 2938–2950.
Vanthornhout, J., Decruy, L., Wouters, J., Simon, J.Z., Francart, T., 2018. Speech intelligibility predicted from neural entrainment of the speech envelope. J. Assoc. Res. Otolaryngol. 19, 181–191.
Wens, V., et al., 2015. A geometric correction scheme for spatial leakage effects in MEG/EEG seed-based functional connectivity mapping. Hum. Brain Mapp. 36, 4604–4621.
Wightman, F., Kistler, D., Brungart, D., 2006. Informational masking of speech in children: auditory-visual integration. J. Acoust. Soc. Am. 119, 3940–3949.
Wightman, F.L., Kistler, D.J., 2005. Informational masking of speech in children: effects of ipsilateral and contralateral distracters. J. Acoust. Soc. Am. 118, 3164–3176.
Wightman, F.L., Kistler, D.J., O'Bryan, A., 2010. Individual differences and age effects in a dichotic informational masking paradigm. J. Acoust. Soc. Am. 128, 270–279.
Wilke, M., Schmithorst, V.J., Holland, S.K., 2002. Assessment of spatial normalization of whole-brain magnetic resonance images in children. Hum. Brain Mapp. 17, 48–60.
Zion-Golumbic, E., Schroeder, C.E., 2012. Attention modulates "speech-tracking" at a cocktail party. Trends Cogn. Sci. 16, 363–364.
Zion-Golumbic, E.M., et al., 2013. Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party. Neuron 77, 980–991.