Developmental Cognitive Neuroscience 51 (2021) 101010

Contents lists available at ScienceDirect

Developmental Cognitive Neuroscience

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

Oscillatory activity and EEG phase synchrony of concurrent word segmentation and meaning-mapping in 9-year-old children

Neus Ramos-Escobar [a,b], Emma Segura [a,b], Guillem Olivé [a,b], Antoni Rodriguez-Fornells [a,b,c,**,1], Clément François [d,*,1]

[a] Dept. of Cognition, Development and Educational Science, Institute of Neuroscience, University of Barcelona, L'Hospitalet de Llobregat, Barcelona, 08097, Spain
[b] Cognition and Brain Plasticity Group, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, 08097, Spain
[c] Catalan Institution for Research and Advanced Studies, ICREA, Barcelona, Spain
[d] Aix Marseille Univ, CNRS, LPL, 13100, Aix-en-Provence, France

A R T I C L E I N F O

Keywords:
Children
Speech segmentation
Meaning-mapping
EEG
Frequency-tagging
Scalp-based phase synchrony

A B S T R A C T

When learning a new language, one must segment words from continuous speech and associate them with meanings. These complex processes can be boosted by attentional mechanisms triggered by multi-sensory information. Previous electrophysiological studies suggest that brain oscillations are sensitive to different hierarchical complexity levels of the input, making them a plausible neural substrate for speech parsing. Here, we investigated the functional role of brain oscillations during concurrent speech segmentation and meaning acquisition in sixty 9-year-old children. We collected EEG data during an audio-visual statistical learning task during which children were exposed to a learning condition with consistent word-picture associations and a random condition with inconsistent word-picture associations before being tested on their ability to recall words and word-picture associations. We capitalized on the brain dynamics to align neural activity to the same rate as an external rhythmic stimulus to explore modulations of neural synchronization and phase synchronization between electrodes during multi-sensory word learning. Results showed enhanced power at both word- and syllabic-rate and increased EEG phase synchronization between frontal and occipital regions in the learning compared to the random condition. These findings suggest that multi-sensory cueing and attentional mechanisms play an essential role in children's successful word learning.

1. Introduction

When learning a new language, one must segment words from continuous speech and associate them with possible meanings. Recent cognitive models suggest that these processes may rely on statistical learning (SL), a domain-general mechanism that allows the computation of statistical regularities within and across sensory modalities (Frost et al., 2015; Räsänen and Rasilo, 2015). In the case of speech, infants, children, and adults can implicitly segment continuous streams of pseudowords by relying on the transitional probabilities between consecutive syllables (Cunillera et al., 2009; François et al., 2013, 2017a; Saffran et al., 1996). Redundant multi-sensory cues can facilitate speech segmentation (Cunillera et al., 2010a, b; Glicksohn and Cohen, 2013; Thiessen, 2010) and word-referent associations (Yu and Smith, 2007; Shukla et al., 2011). Interestingly, these redundant cues may capture learners' attention by highlighting temporally contiguous target words and referents (Axelsson et al., 2012; Horst and Samuelson, 2008). These results align well with previous studies proposing that the combination of attentional and associative processes may facilitate multi-modal word learning (Plunkett, 1997; Smith, 2000). The attention toward a visual event may trigger the association of the visual referent with the spoken word. This attentional capture may occur through cue facilitation, where the learner uses the referent's perceptual saliency as a cue for the binding process. At the same time, this process could rely on associative mechanisms that allow the binding of temporally contiguous multi-sensory events (Smith, 2000). In any case, enhancing visual

* Corresponding author at: Laboratoire Parole et Langage, CNRS, Aix-Marseille University, 5 avenue Pasteur, 13100, Aix-en-Provence, France.
** Corresponding author at: Dept. of Cognition, Development and Educational Science, Institute of Neuroscience, University of Barcelona, L'Hospitalet de Llobregat, Barcelona, 08097, Spain.
E-mail addresses: antoni.rodriguez@icrea.es (A. Rodriguez-Fornells), clement.francois@univ-amu.fr (C. François).
[1] Co-last authorship.

https://doi.org/10.1016/j.dcn.2021.101010
Received 22 October 2020; Received in revised form 25 August 2021; Accepted 26 August 2021
Available online 26 August 2021

attention to a referent by perceptual saliency or temporal contiguity may facilitate word learning. Overall, current empirical evidence and theoretical proposals converge that redundant cues might facilitate language learning.

Apart from the mechanisms mentioned above, other cognitive and linguistic factors have been proposed to explain children's ability to learn new words, such as implicit and perceptual learning (Reber, 1967; Gibson and Gibson, 1955), social context (Tomasello, 1995; Bloom, 2000) and pre-existing linguistic constraints (Markman, 1990; Carey and Bartlett, 1978). Hollich et al. (2000) proposed a compiled model of word acquisition positing that children's word-learning ability emerges from combining some of these factors and that the weight of those factors changes through development. During the initial stages of word learning, children would rely on domain-general cues (e.g., social and attentional cues). During later learning stages, they would rely on language-specific cues instead (e.g., phonotactic or grammatical rules). In this line, some authors suggest there exists a shift in the way children learn novel words during middle school years, from direct, explicit teaching to increased learning from context (Dickinson, 1984; Chall, 1983). For instance, 8- to 9-year-old children benefit more from multiple word-encounters within a story than 6−7-year-olds (Wilkinson and Houston-Price, 2013). In the third and fourth grades, children begin to distinguish between pseudowords and nonsense letter strings in an adult-like manner (Henderson and Chard, 1980; Lefton and Spragins, 1974). Children of this age also show evidence of adult-like semantic processing (Coch, 2015). Interestingly, implicit learning abilities also change during development and follow a U-shaped trajectory across the lifespan (Juhasz et al., 2019; Janacsek et al., 2012). Indeed, 7–9-year-old children show superior general skills than younger children (Juhasz et al., 2019), reaching a plateau at around 12 (Janacsek et al., 2012). Finally, orienting attention abilities, which are crucial for multi-sensory tasks, seem to develop between middle and late childhood (Pozuelos et al., 2014). Overall, several converging pieces of evidence suggest a knee point in language learning mechanisms during middle childhood.

Previous studies have used Electroencephalography (EEG) to decipher the neural substrates of speech segmentation based on SL in adults (Batterink and Paller, 2017; Cunillera et al., 2009; de Diego Balaguer et al., 2007; François et al., 2014, 2017b), infants (Saffran et al., 1996; Aslin et al., 1998) and newborns (François et al., 2017a; Teinonen et al., 2009). Functional magnetic resonance imaging studies (fMRI) have revealed that speech segmentation based on SL induces consistent functional activations of the superior temporal gyrus (STG), the left supramarginal gyrus, but also the left inferior frontal gyrus (Cunillera et al., 2009; McNealy et al., 2006). Furthermore, despite the importance of understanding how multimodal language learning operates, only one study has explored the neurophysiological correlates of parallel speech segmentation and meaning-mapping in adults (François et al., 2017b). In this study, adults were exposed to an audio-alone speech segmentation task and an audio-visual meaning-mapping task. After exposure to random and statistically structured streams, participants were tested on the online detection of mismatches, the explicit recall of words, and word-to-picture associations. Behavioral data showed better word recognition in the audio-visual compared to the audio condition. Furthermore, in both conditions the event-related brain potentials (ERPs) revealed a larger fronto-central N400 for the structured streams compared to the random ones. Crucially, the topographical differences between conditions suggested that additional resources over frontal regions might be necessary for the meaning-mapping.

To the best of our knowledge, no previous study has focused on brain oscillations during multimodal language learning in children. Brain oscillations play a crucial role in a variety of cognitive operations, such as stimulus processing (Thut et al., 2011; Worden et al., 2000), cognitive control (Canolty et al., 2006; Cunillera et al., 2012), and the development of abstract reasoning or verbal working memory (Power et al., 2012; Embury et al., 2019; Taylor et al., 2020). Several studies have used phase synchrony between brain regions to explore the involvement of long-range communication (Varela et al., 2001; Lachaux et al., 1999) during cognitive tasks that may require information flows between different brain regions (Buzsáki and Dragun, 2004). Interestingly, synchronization among brain regions is associated with learning (Fries, 2005) and the intrinsic hierarchical organization of brain oscillations allows to optimize stimulus processing (Lakatos et al., 2005). Moreover, one influential tenant is that speech processing is realized via the synchronization of neural activity to the speech signal (Giraud and Poeppel, 2012; Luo et al., 2010; Luo and Poeppel, 2007). Better alignment with external stimuli predicts better speech comprehension and audio-motor response (Assaneo et al., 2019; Peelle et al., 2013). Further research shows that neural synchronization to the stimulus only occurs for attended stimuli (Ding and Simon, 2012; Lakatos et al., 2013), whereas when the tracking of the stimulus is not possible, speech perception and comprehension is impoverished (Ahissar et al., 2001; Luo and Poeppel, 2007). These findings suggest that top-down attentional mechanisms modulate speech processing in children (Wang et al., 2014; Panda et al., 2020; Ríos-López et al., 2020). One way to investigate oscillatory brain activity is to use frequency-tagging analysis, which shows the fine neural synchronization in response to both auditory and visual stimuli (Doesburg et al., 2012; Kim et al., 2007; Keitel et al., 2014; Covic et al., 2017; Nozaradan et al., 2012). This method has been used to show that successful speech segmentation is associated with a peak at both syllable and word frequencies during exposure to statistically structured streams (Batterink and Paller, 2017; Buiatti et al., 2009; Farthouat et al., 2017; Henin et al., 2021; Ordin et al., 2020). However, EEG-based studies in infants and children are still scarce (François et al., 2013, 2017a; Kabdebon et al., 2015; Teinonen et al., 2009), and no previous study has used frequency-tagging combined with scalp-based phase synchronization analyses to explore multimodal word learning in children.

Here, we explored concurrent speech segmentation and meaning-mapping in a large group of sixty 9-year-old children. Even though infants and adults often use environmental information to aid the extraction and acquisition of novel word forms (Smith et al., 2002; Johns et al., 2016), these two processes have traditionally been explored separately in the literature (Graf-Estes et al., 2007; Hay et al., 2011; Räsänen and Rasilo, 2015). Besides, language acquisition can be considered a joint inference problem for multiple linguistic elements that need to be learned in parallel (Johnson and Tyler, 2010; Lim et al., 2015). To our knowledge, only two studies in adults have shown that the statistical properties found in continuous speech together with the statistical consistency between speech and visual referents might be crucial to infer the possible meaning of a new word (Cunillera et al., 2010a, b; François et al., 2017b). However, the neurophysiological mechanisms involved when the two processes of segmenting new words and binding conceptual representations onto these newly isolated words are taking place at the same time are poorly understood in children.

With this purpose in mind, we recorded EEG during an audio-visual SL task. We applied frequency-tagging and scalp-based phase synchronization analyses to track the online modulations in neural synchronization and long-distance connectivity during learning. EEG phase synchronization analyses at the sensor level used the weighted Phase-Lag Index (wPLI), a measure of phase synchronization between two signals that avoids the problem of volume conduction (Vinck et al., 2011). Specifically, children were presented with a learning stream with consistent word-picture associations and a random stream with inconsistent ones. Based on previous results in adults (François et al., 2017b), the multimodal presentation of words and referents may facilitate word segmentation and the binding to its referent. The presence of this visual referent should induce attentional capture and facilitate the emergence of new multi-sensory memory traces. Therefore, we expected larger neural responses at both syllable and word frequencies for learning than for random streams over frontal regions as observed in previous ERP studies on speech segmentation based on SL (Cunillera et al., 2006, 2009; De Diego-Balaguer, 2007; François et al., 2014). We also expected an enhanced word-frequency response over occipital electrodes reflecting the perception of the pictures at the same frequency as words. Finally, we predicted a larger cross-talk between occipital (visual) and fronto-central electrodes (speech-segmentation) with stronger scalp-based connectivity between fronto-occipital regions in the learning than in the random condition.

2. Methods

2.1. Participants

A total of seventy-two children attending the fourth grade in two elementary schools in Barcelona participated in the study. Two children did not finish the task, and ten more were excluded due to noise in the recordings due to excessive movements (4), drowsiness (3), or background noise (3). Hence, a total of sixty children (mean age: 9 years and five months; 34 girls) were included in the study. All children had similar socioeconomic status determined from the parents' incomes (middle/upper class). They were Catalan-Spanish bilinguals and right-handed, except for four left-handed. A written informed consent was obtained from all the children's parents before the start of the study. School directors and teachers were also informed and collaborated in the organization of the experiment. The study was conducted following the local norms and guidelines for protecting human subjects and was approved by the local ethics committee.

2.2. Stimuli

We used the same audio-visual artificial language streams as those used in François et al. (2017b). Briefly, we created a synthetic speech stream containing four trisyllabic pseudowords ("tapiru", "bagoli"," gukido", "tokuda"; from now on, called words) that were concatenated in a pseudo-random order with no immediate repetition (see Fig. 1). We used the MBROLA speech synthesizer with the Spanish male database (es1) to create the stream. All phonemes had the same duration (116 ms) and pitch (200 Hz). Each word had a duration of 696 ms and was repeated 25 times in the stream. The language stream was presented four times with a short pause between each presentation for a total of 400 words and a duration of 4 min and 39 s. The transitional probability between syllables was 1.0 within words and .33 between words.

In addition to the statistically structured stream, a random stream was created by pseudo-randomly mixing the syllables. The random stream had the same duration as the structured one, but the transitional probabilities between syllables were .09. Therefore, EEG signals from this condition were used as a baseline.

Four different pictures were synchronously presented with the four words of the structured speech stream. The visual stimuli consisted of four 20 × 120 mm black-and-white drawings belonging to four different semantic categories (animals, vegetables, vehicles, and accessories; Snodgrass and Vanderwart, 1980) while controlling for name agreement (>80 %), familiarity (M = 3.4), and complexity (M = 3.1). Significantly, each word was associated with a single picture resulting in four fully consistent word-picture associations. In the random condition, four

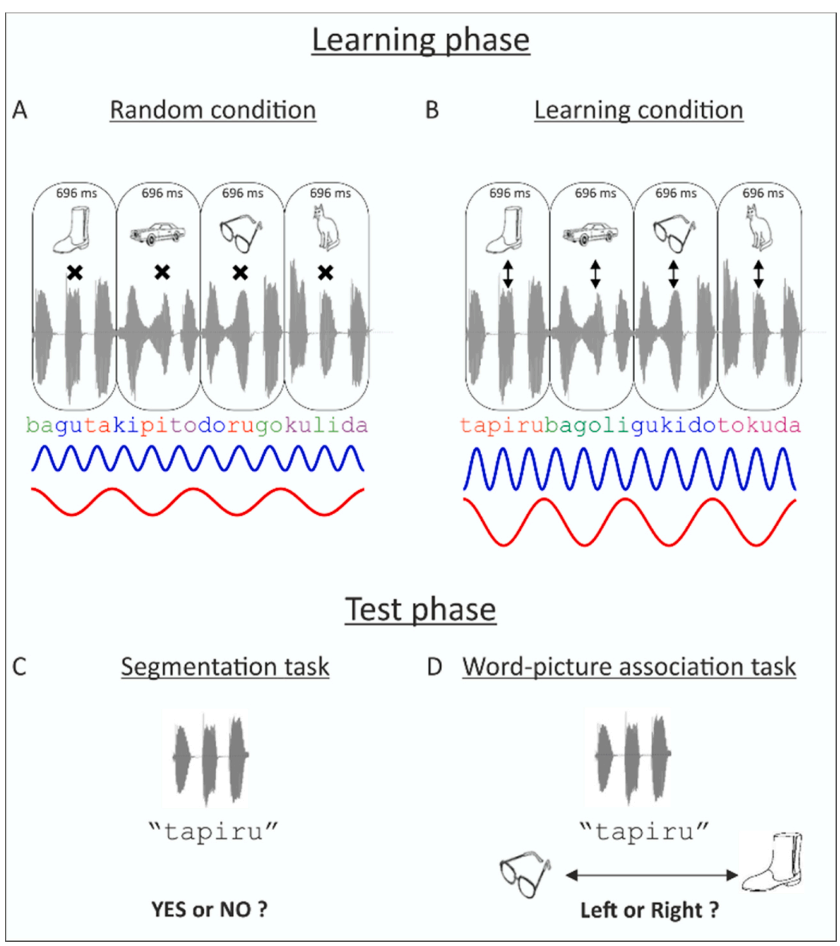

Fig. 1. Illustration of the audio-visual streams used in the random and learning conditions (A and B, respectively) with the corresponding neural response at syllable (blue) and word (red) frequency. The learning phases were immediately followed by the test phase assessing both word and word-picture association. C: Illustration of one trial from the segmentation task. D: Illustration of one trial from the word to picture association task.

different pictures [name agreement (>80 %), familiarity (M = 3.4), and complexity (M = 2.9)] were pseudo-randomly presented on the screen every 696 ms. However, because the random stream did not contain words, no association could be created.

2.3. Experimental procedure

Children were tested individually in a quiet room of their school in 2 separate sessions on two different days that included neuropsychological assessments and electrophysiological tests, respectively. The electrophysiological session lasted one hour and a half, and the neuropsychological session lasted 30 min.

In the neuropsychological session, children were evaluated with subtests of the NEPSY-II battery (Korkman et al., 2007) and the WISC-IV (Wechsler and Corral, 2015). Specifically, language functions were assessed using the verbal fluency and nonword repetition tasks from the NEPSY-II battery. Non-verbal problem solving was evaluated with the matrix-reasoning subtest and working memory with the digit span subtest from the WISC-IV. Overall, they performed within the normal range (Digit span: M = 15.56, SD = 2.89, mean normal score: 13–14; Matrix reasoning: M = 17.9, SD = 3.71, mean normal score 18–19; Non-word repetition: M = 41.18, SD = 2.65, expected range: 34–42; Verbal fluency: M = 33.28; SD = 6.33, expected range: 25–35; Letter fluency: M = 15.02, SD = 6.05, expected range: 11–20).

During the electrophysiological session, children were seated comfortably in a chair and performed different computer-based tasks while we EEG was recorded. Before the learning phase, children were explained to carefully listen to discover the words of an "alien" language. First, the random stream was presented four times (Random Condition), followed by four presentations of the structured stream that contained the novel words (learning condition). Between each stream, a short pause was done, and children were encouraged to continue learning. This pause was especially important in the random condition to avoid children becoming discouraged due to the impossibility of learning. After the learning phase, children were assessed with a speech segmentation and a word-to-picture association task (Fig. 1).

The level of performance for word segmentation was assessed with a lexical decision task. On each trial, children heard one of the four words from the language or a nonword. Each word was presented eight times, thus, leading to 32 trials. The nonwords were compiled by mixing the syllables of the words pseudo-randomly. Children had to decide whether the item was a word from the language or not. We chose to assess word recognition with a lexical decision task (LDT) to collect behavioral responses for both types of test items as was done in adults (François et al., 2017b).

Children's word to picture association performance was assessed with an associative word-picture matching task. On each trial, children heard one of the four words while two pictures were displayed on the computer screen. Each picture was presented three times, thus leading to 12 trials. They had to choose which picture (left or right) was associated with the word. Notably, both pictures appeared in the continuous streams. In both tasks, each trial was presented on the screen until children responded, forcing them to answer before moving to the subsequent trial.

To explore the role of age, we performed Pearson correlations between the behavioral performance in the speech segmentation and word-picture association tasks and the age at test. We performed additional correlations assessing the relationship between age and word recognition and nonword rejection separately for the speech segmentation task.

2.4. EEG data acquisition and analyses

The EEG signal was recorded at a sampling rate of 256 Hz using a Biosemi amplifier system (Biosemi ActiveTwo, Amsterdam University) from 16 active Ag-Cl scalp electrodes mounted on a child-sized elastic cap at standard positions (International 10/20 system sites: Fp1, Fp2, F3, F4, T7, C3, C4, T8, P3, P4, O1, O2, Fz, Cz, Pz, and Oz). Five additional electrodes were placed on the right and left mastoids for off-line re-referencing, at 1 cm from the right and left outer canthi and below the right eye to control blinks and lateral eye movements.

2.5. Frequency tagging analysis of spectral power

Continuous EEG recordings were re-referenced off-line to the average left and right mastoids and high-pass filtered at 0.1 Hz to remove slow drifts. Epochs of 5.57 s, thus, corresponding to the presentation of 8 words, were extracted to obtain a high frequency-resolution and were baseline corrected. This large time-window provided frequency bins equal to 0.12 Hz allowing to get one frequency bin centered on these two target frequencies. With such resolution, we obtained reasonable power estimations for both target frequencies (word frequency: 1.43 Hz and syllable frequency: 4.31 Hz). Epochs with activity exceeding baseline level by +/− 80 μV in the [−50 700 ms] period were excluded from the analyses. We used only one of every four epochs in the final average to avoid oversampling. We then transformed the obtained waveforms into the frequency domain using a discrete Fourier transformation (Matlab; Natick, MA). Finally, the power at both target frequencies was obtained for each child and electrode in the random and learning conditions. All the scripts were manually prepared for this analysis. For the statistical analysis, we used a repeated-measures ANOVA with four within-subject factors: Frequency (Syllable Frequency, Word Frequency), Condition (Learning, Random), region of interest (ROI; Frontal, Central, Parietal, Occipital), and Lateralization (Left, Central, Right).

2.6. Scalp-based phase synchronization analysis – weighted Phase-Lag Index

To assess the phase relationship between different brain regions during speech segmentation, we used the wPLI. We selected this measure because it allows to avoid volume conduction since zero-lag synchronization is left out from the analysis, and thus, the reported measure is not based on shared sources (Bastos and Schoffelen, 2016; Stam et al., 2007). We performed the same pre-processing as in the frequency-tagging analysis, computed the Fourier transform, and finally obtained the wPLI using Fieldtrip Toolbox (Oostenveld et al., 2011) run in Matlab. For the statistical analysis, we defined four different regions of interest (ROI): frontal (F3, F4, Fz), central (C3, C4, Cz), parietal (P3, P4, Pz), and occipital (O1, O2, Oz) to perform a repeated-measures ANOVA with three within-subjects factors: Frequency (Syllable Frequency, Word Frequency), Condition (Learning, Random), and ROI (Fronto-Occipital, Fronto-Central, Fronto-Parietal, Occipito-Central, Occipito-Parietal, Parietal-Central). Post-hoc *t*-tests were conducted to determine the source of significant interactions and were corrected for multiple comparisons using the Bonferroni-Holm FDR correction method to avoid false positives (Benjamini and Hochberg, 1995). Finally, and for display purpose only, the surviving significant *t*-values were plotted separately for each target frequency using the BrainNet Viewer (Xia et al., 2013).

3. Results

3.1. Behavioral results

Fig. 2 shows the behavioral data obtained in the speech segmentation and word-picture association tasks. For the speech segmentation task (Fig. 2A), the comparison of performance against chance level (50 %) showed that children performed significantly above chance level for both words (M = 63.4, SEM = 2.15) and non-words (M = 62.2, SEM = 1.78, both P's < .001). This result indicates that they were able to recall newly segmented words and to reject nonwords correctly. The level of

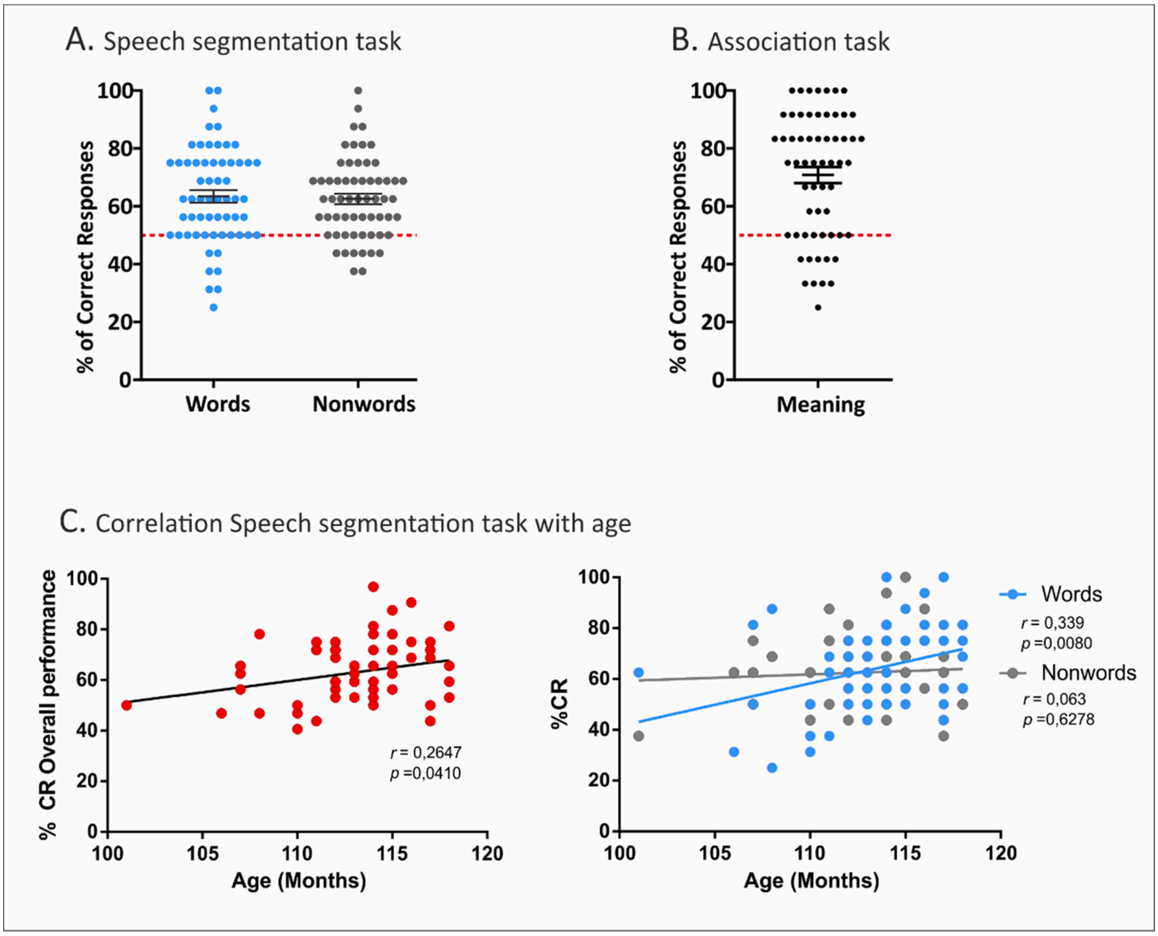

Fig. 2. Behavioral data. (A) Percentage of correct responses in the speech segmentation task for words and nonwords. (B) Percentage of correct responses in the word-picture association task (B). Dots represent individual values, and bars correspond to the mean and standard error of the mean (SEM) in each condition. The red dotted line represents chance level (50 %). (C) Scatter plot showing the correlations between age (in months) and the level of performance in the speech segmentation task (Left: overall performance; Right: separated performance for words and nonwords).

performance for words and non-words was not significantly different (t_{59} = 0.35; P = .72). For the word-picture association task, the mean percentage of correct responses (M = 70.49, SEM = 2.77) was significantly above chance level (t_{59} = 7.39; p. < .001), indicating that children were also able to associate the pictures to the segmented words.

We explored the role of age by computing Pearson correlations between the age at test (in months) and the overall performance in both the speech segmentation and word-picture association tasks. Results showed that age was positively associated with the overall performance in the segmentation task (r = .264; p = .041; see Fig. 2C) but not in the word-picture association task (r = .075; p = .564) even though these results did not survive the correction for multiple comparisons. Separated correlations for words and non-words in the segmentation task showed that age was positively associated with word recall (r = .339 p = .008; see Fig. 2D) while this was not the case for nonword rejection (r = .063; p = .627).

3.2. Electrophysiological results

3.2.1. Frequency tagging analysis

Fig. 3A and B display the frequency-tagging analysis results showing clear peaks at word and syllable frequencies. The peak to word frequency was higher than for syllable frequency [significant main effect of Frequency: F(1,59) = 11.49; p = .001]. As predicted, considering the rhythmic presentation of pictures (matching word length), we observed a clear peak at the word frequency over occipital ROI [significant Frequency x ROI interaction [F(1,177) = 22.09; p < .001]. Interestingly, some ROIs showed larger differences between Learning and Random conditions than others at both word and syllable frequency [evidenced by a non-significant main effect of Condition F(1,59) = 2.94; p = .092 but a significant Condition x ROI interaction F(3,177) = 9.18; p < .001 and a significant Condition x Frequency x ROI x Lateralization interaction: F(6,354) = 4.65; p = .007; Fig. 3A & B]. To further explore these effects, we performed separate repeated-measures ANOVAs for both word and syllable frequency.

At the word frequency, a clear peak was found over all ROIs with a strong response over occipital regions [main effect of ROI: F(3,177) = 18.14; p < .001]. When looking at differences between conditions across ROIs [significant Condition x ROI interaction: F(3,177) = 3.54; p = .029] we found that those differences occurred over frontal regions only [t_{59} = −2.31; p = 0.025] with an enhanced power in the learning compared to the random condition. Interestingly, despite the clear peak at occipital ROI, there were no differences between conditions in this region [t_{59} = 1.17; p = 0.247]. The Condition by ROI by Lateralization interaction was also significant [F(6,354) = 3.74; p = .016]. Further post-hoc analysis at the electrode level showed significantly enhanced power at word

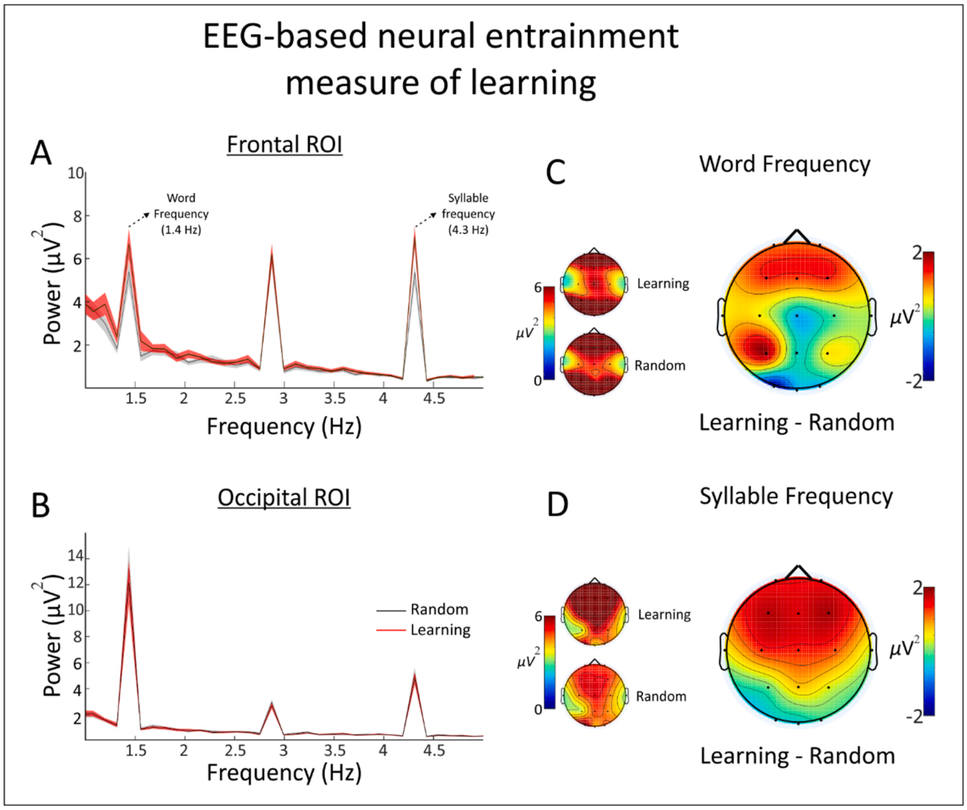

Fig. 3. Results of the frequency tagging analysis. Power spectrum of the EEG signal over frontal (A) and occipital (B) regions of interest (Mean and SEM). Red = learning condition, black = random condition. Topographical distribution of the power at word (C) and syllable frequency (D) for the single conditions and the difference between conditions (Learning-Random).

frequency in the learning compared to the random conditions at F4, Fz, F3 and P3 electrodes (F4: $t_{59} = -2.2$; $p = .032$; Fz: $t_{59} = -2.01$; $p = .048$; F3: $t_{59} = -2.45$; $p = .017$; P3: $t_{59} = -2.18$; $p = .033$; see Fig. 3C). However, none of these post-hoc comparisons survived the FDR correction for multiple comparison.

At the syllable frequency, we found clear peaks in all ROIs with a stronger response over frontal regions [main effect of ROI: $F(3,177) = 6.51$; $p = .002$]. Results also revealed an overall enhanced power in the learning compared to the random condition [main effect of Condition: $F(1,59) = 6.81$; $p = .01$]. Interestingly, this effect was differently distributed across ROIs [significant Condition x ROI interaction: $F(3,117) = 15.99$; $p < .001$] with enhanced power in the learning compared to the random condition over frontal ($t_{59} = -4.59$; $p < .001$) and central regions ($t_{59} = -3.44$; $p = .01$). No differences between the learning and random conditions were found over parietal ($t_{59} = -.64$; $p = .526$) and occipital ($t_{59} = 1.17$; $p = .247$) regions. The Condition by ROI by Lateralization interaction was also significant [$F(6,354) = 3.41$; $p = .011$]. Further post-hoc analysis at electrode level showed that the power at the syllable frequency was larger in the learning condition than in the random one (FDR correction, $p < .004$) at F4, Fz, F3, C3, Cz, and C4 electrodes (F4: $t_{59} = -4.66$; $p < .001$; Fz: $t_{59} = -4.11$; $p < .001$; F3: $t_{59} = -4.68$; $p < .001$; C3: $t_{59} = -3.77$; $p < .001$; Cz: $t_{59} = -3.11$; $p = .003$; C4: $t(59) = -3.02$; $p = .004$, see Fig. 3D).

3.2.2. *Scalp-based phase synchronization analysis*

Based on the frequency-tagging results, we expected a larger cross-talk between occipital (visual) and fronto-central electrodes or ROIs (speech segmentation) in the learning than in random condition, reflecting the involvement of visual cues in word binding and integration. Specifically, we expected stronger EEG phase synchronization between fronto-occipital electrodes at word frequency. Fig. 4A & B show the significant connections between electrodes at word and syllable frequency, respectively. The wPLI values were stronger at the syllable than at the word frequency [significant main effect of Frequency: $F(1,59) = 72.91$; $p < .001$]. Fig. 4C depicts the phase synchronization EEG spectra across ROIs, pointing to a higher phase synchronization strength in the learning compared to the random condition [main effect of Condition: $F(1,59) = 15.69$; $p < .001$]. As can be seen in Fig. 4D & E, this effect was differently distributed between ROIs [significant Frequency x Condition x ROI interaction: $F(1,59) = 3.26$; $p = .021$]. Further post-hoc *t*-tests conducted at the word frequency showed stronger EEG phase synchronization in the learning compared to the random condition in fronto-occipital ($t_{59} = 2.91$; $p = .005$), fronto-central ($t_{59} = 2.77$; $p = .008$), fronto-parietal ($t_{59} = 4.38$; $p < .001$) and parieto-central pairs of ROIs ($t_{59} = 2.91$; $p = .005$; FDR correction, $p < .008$; see Fig. 4D). For the syllable frequency we found significantly stronger phase synchronization in the learning compared to random condition in fronto-occipital ($t_{59} = 3.38$; $p = .001$) and occipito-central pairs of ROIs ($t_{59} = 2.98$; $p = .004$; FDR correction, $p < .004$; see Fig. 4E).

For the frequency-tagging and the phase synchronization analyses, we also included the factor age (in months) as a covariate. No significant main effect of age was found in these analyses [Frequency tagging: $F(12,47) = 0.73$; $p = .734$]; Phase synchronization: $F(12,47) = 1.29$; $p = .254$) and none of the interaction survived the correction for multiple comparisons.

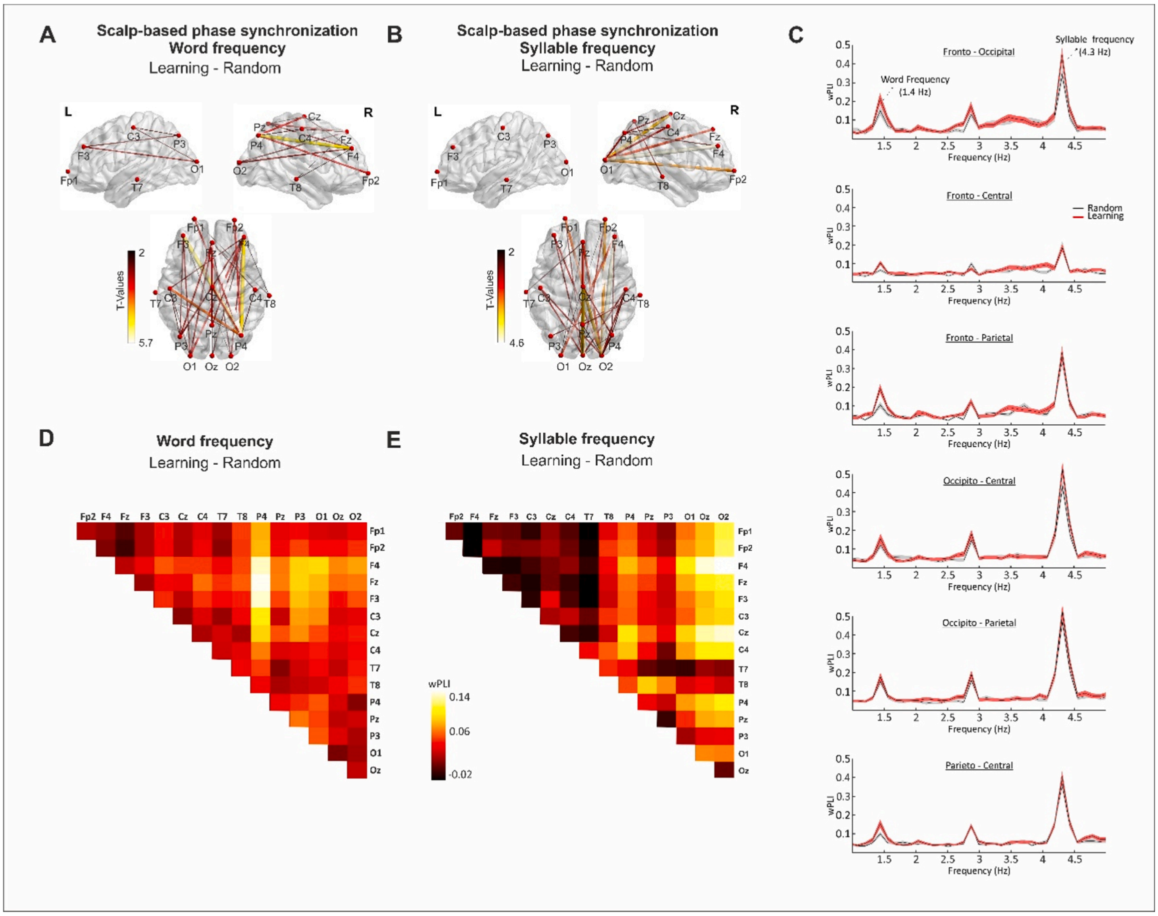

Fig. 4. Results of the scalp-based phase synchronization analysis. Statistical scalp-based connectivity maps for the learning - random contrast at the word (A) and the syllable frequency (B). Note that the analyses were performed in the sensor space. The color scale indicates t-values with light colors indicating strong EEG phase synchronization differences. C) Phase synchronization EEG spectra between the different ROIs (mean and SEM). Red = learning condition, black = random condition. D & E) EEG phase synchronization matrices at the word (D) and syllable frequency (E) for each pair of electrodes.

4. Discussion

The present study investigated the functional role of brain oscillations during concurrent speech segmentation and meaning acquisition in sixty 9-year-old children. Specifically, we used frequency tagging and EEG phase synchronization analyses to assess neural tracking and long-distance scalp-based phase synchronization during an audio-visual SL task. Behavioural results showed that children successfully learned the four novel word-forms and their associated meaning. Frequency-tagging analysis of scalp EEG data revealed enhanced power at word and syllable frequencies in the learning compared to the random condition. Furthermore, we found evidence of enhanced long-distance EEG phase synchronization between frontal and parieto-occipital ROIs at both target frequencies. These converging results suggest that attentional mechanisms modulate word segmentation and meaning-mapping during children's multimodal word learning induced by consistent visual cues and thus, facilitate learning (Hollich et al., 2000).

At the behavioural level, the results showed that children's performance was above chance level for both the speech segmentation and the word-to-picture association tasks. Even though the children only learned four words, these results are the first to show that 9-year-old children can successfully perform speech segmentation and meaning-mapping concurrently. This finding is in line with previous reports showing that redundant cues such as prosody can trigger word-object associations in infants (Plunkett, 1997; Smith, 2000; Shukla et al., 2011). The present results also replicate our previous study in adults in which the presence of visual referents induced refined memory traces of the correct syllabic patterns (François et al., 2017b). Nonetheless, the level of performance in the word-picture association task may appear relatively low, with only 70 % of correctly recognized associations. Several factors may account for this surprisingly low performance in the association task. First, it is important to keep in mind that the task is not as easy as it seems due to the relatively high syllabic rate (4.4 Hz). Second, children were not explicitly instructed to learn the meaning of the words nor that the pictures corresponded to the word onset and offset. Instead, they were told to listen carefully to the stream to discover the words of an "alien" language. Thus, the instructions were directed toward segmentation and not meaning-mapping leaving children to discover the relationship between the pictures displayed on the screen and the auditory words by themselves. Finally, the behavioural tasks (lexical decision task or 2AFC) used in SL experiments are known to generate interferences created by the presentation of the illegal items (François et al., 2012). Further research may use implicit reaction time tasks as recently described in studies with adults (Batterink et al., 2015; François et al., 2012). Overall, our results suggest that 9-year-old children and adults exhibit relatively similar behavioural patterns of explicit recall.

Despite the relatively low variability in age, we explored the association between age and behavioural measures of word segmentation and meaning-mapping. We observed that age in months predicted the overall performance in the speech segmentation task with older children exhibiting higher performance than their younger peers. Interestingly, implicit learning mechanisms, which contribute to SL, seem to follow a U-shaped developmental trajectory across the lifespan, with the knee point observed around nine years of age (Juhasz et al., 2019; Janacsek et al., 2012). Juhasz et al. (2019) found that 7–9 years-old children show better procedural learning abilities than adolescents or adults. Janacsek et al. (2012) also reported that SL reaches a plateau after 12 years of age, with a peak of performance observed between 4- and 12-years. Other studies have shown that visual SL abilities may gradually increase between 5 and 12 years of age (Arciuli and Simpson, 2011; Raviv & Arnon, 2018; but see Shufaniya and Arnon, 2018 for differences between auditory linguistic and non-linguistic inputs). Thus, the present results might converge with the idea that SL in the linguistic domain is still improving in 9-year-old children. Importantly, the audio-visual task used here contained not only statistical information but also visual cues that facilitated the segmentation of possible words. Therefore, it is difficult to conclude to which extent the observed correlations indicate an improvement of SL with age, or whether they are due to an increasing capacity to use multiple cues when learning new words (Cunillera et al., 2010a). Nonetheless, we observed that age predicted word recognition but did not predict nonword rejection in the speech segmentation task. This last result suggests that the recall of novel word forms is more dependent on age compared to nonwords rejection. Again, these results should be interpreted carefully due to the limited range of age explored as we tested only one single group of children. Such a narrow range of age limits the possibility to elaborate on the changes in cue-weighting that supports language acquisition across development. Further studies are needed to appropriately track the developmental trajectories of concurrent speech segmentation and meaning-mapping as well as its neural signatures.

The frequency tagging analysis revealed clear peaks at both syllable and word frequencies during exposure to the streams. At both target frequencies, we found enhanced power in the learning compared to the random condition. However, the differences survived FDR correction only for the syllable frequency suggesting that differences between conditions were higher when tracking the syllables than the words. The synchronized activity for the syllable frequency showed a frontal distribution, confirming previous ERP data obtained in adults (François et al., 2017b). Using a similar experimental design, François et al. (2017b) found that audio-visual streams elicited a larger fronto-central ERP component (FN400) in the learning than in the random condition. Previous electrophysiological studies of speech segmentation in adults revealed maximal N400 effects over fronto-central regions (Cunillera et al., 2006; 2009; François et al., 2014; de Diego Balaguer et al., 2007). Importantly, the level of neural alignment to the speech streams was significantly stronger at the word than at the syllable frequency. This result confirms previous findings in adults that show temporal alignment of brain activity with speech and musical stimuli (Luo et al., 2010; Luo and Poeppel, 2007; Nozaradan et al., 2012). Besides, we focused on EEG raw power and observed enhanced activity in the learning condition compared to the random at the word- and syllable frequency. While some studies have also focused on raw power (Buiatti et al., 2009; Nozaradan, 2014), others have used inter-trial phase coherence (ITC) where the phase is also considered (Buiatti et al., 2009; Batterink and Paller, 2017; Henin et al., 2021). Results based on ITC analyses showed higher coherence in the learning than in the random condition at word frequency but the opposite pattern was observed at syllable frequency with higher ITC values in the random than in the learning condition. Several reasons may account for these differences between previous studies and the present one.

First, methodological differences may explain the different patterns of results. A recent study used both measures to explore the oscillatory activity during an auditory SL task and found similar results (Ordin et al., 2020). While ITC provides information about the synchronization precision across trials, raw power informs on the intensity of the brain response. Therefore, although ITC and power measures are related, they provide different, not overlapping information (Wöstmann et al., 2017). Second, our study is the first to explore oscillatory activity in children during speech segmentation. Thus, maturational brain differences between children and adults may also explain the divergent patterns of results mentioned above. For instance, children's neuronal alignment to speech differs from that of adults. A recent study found that cortical tracking of speech at the syllable rate (4−8 Hz) predicted speech-in-noise discrimination ability in 6- to 9-year-old children. Notably, the detrimental effect of increasing auditory noise on the level of neural tracking was stronger in children than in adults suggesting that children may exhibit an immature cortical tracking of speech (Vander Ghinst et al., 2019). Third, Hollich et al. (2000) proposed that the relative weight of linguistic and non-linguistic cues in word learning facilitation changes over the course of development. Specifically, while children might be biased to rely on domain-general cues during the initial stages of word-learning, they are likely to use language-specific information (e.g., grammatical rules) during later learning stages. Thus, developmental changes in speech brain-tracking and attentional factors such as the reliance on and weighting of different cues may account for the differences between adults and children (Ríos-López et al., 2020; Moore et al., 1999). In the present study, visual cues may have facilitated the discovery of words by prompting participants to focus on the first syllable of the words (synchronized with the picture presentation), which might, in turn, modulate sensory processing (Parasuraman, 1980; Woldorff et al., 1993). The high response to word-frequency over occipital regions further supports this idea (see Fig. 4B). Because the pictures were presented at the same frequency as the words, this neuronal response over occipital regions presumably reflects the neural alignment to pictures. Steady-state visual evoked potentials studies consistently showed that attention toward the stimulus could modulate the neuronal response to that stimulus (Hillyard et al., 1998; Keitel et al., 2013; Morgan et al., 1996; Kim et al., 2007). Therefore, the attentional capture by visual referent may enhance sensory processing and facilitate the overall word learning process.

The EEG phase synchronization analyses based on the wPLI revealed enhanced long-distance phase synchronization strength between frontal-occipital and frontal-parietal ROIs at word frequency in the learning compared to the random condition. This result suggests that attentional modulation induced by consistent visual cues may facilitate learning. This idea fits well with recent results obtained during an audio-visual movie visualization task and showing increased frontal-occipital phase synchrony in the sensor space when children listened to well-formed sentences compared to semantically incongruous sentences (Panda et al., 2020). Previous studies using similar methodologies have also shown that the neural tracking of a specific talker in higher-order auditory regions differs between attended and non-attended speech input (Ding and Simon, 2012; Lakatos et al., 2013). Moreover, endogenous fluctuations of brain excitability may drive differences in behavior during an attentional task (Helfrich et al., 2018). These reports suggest that voluntary attention to a specific stimulus modulates neural tracking and, therefore, stimulus encoding. Studies on infant's gaze-following showed that word learning differences arise from different object-looking behaviors (Yu and Smith, 2012). Besides, multiple forms of visual attention may support statistical word-referent learning in different ways (see Talsma et al., 2010 for a review; Smith, 2000). Therefore, the fronto-occipital phase connectivity patterns found in the present study may reflect attentional modulations arising from visual cues and targeting frontal and temporal auditory speech brain regions. These attentional modulations may be crucial for optimizing multimodal word-learning mechanisms (Smith, 2000; Plunket, 1997; Hollich et al., 2000). However, the spatial resolution of the 16-electrode EEG system used here is somewhat limited. Further studies with a better spatial resolution must be conducted to shed more light on the brain mechanisms involved in multi-sensory word learning.

Interestingly, the frequency tagging and the phase connectivity analyses revealed unexpected results. We observed an enhanced syllable frequency power in the learning compared to the random condition and stronger long-distance scalp-based connectivity values for the syllable than for the word frequency. Although these results may seem surprising, Buiatti et al. (2009) reported similar findings in adults. In their study, EEG frequency tagging results obtained during an auditory speech segmentation task showed larger differences between random and learning conditions in the syllable than word frequency. Here, children may have initially tried to generate predictions about possible word candidates based on the visual cues during early exposure to the random streams. However, the inconsistencies between the visual referents signaling trisyllabic words and the random auditory stream may have interfered with these predictions by continuously decreasing the level of attention. For the syllable frequency, we observed a reduction in power in the random compared to the learning condition. We also found a decreased EEG phase synchronization strength between frontal-occipital and occipito-central pairs of ROIs in the random condition. An interesting question for future research is whether attentional modulation during SL is due to participants' awareness of impossible learning or if it is internally self-regulated without active decision-making processes. Indeed, intrinsic self-regulated learning mechanisms that select rich learning environments in which the best sources of information exist (see Begus et al., 2016) might explain why infants and children seem to be sensitive to the reliability of information during learning, paying particular attention to reliable speakers during language acquisition (Jaswal and Neely, 2006; Koenig and Harris, 2005; Sabbagh and Baldwin, 2001). For instance, 17-month-old infants show a decreased attention when presented with an unlearnable compared to a learnable artificial grammar (Gerken et al., 2011). Similarly, 4-year-old but not 3-year-old children can predict the reliability of an informant during novel word learning based on previous experience (Koenig and Harris, 2005). In our paradigm, the presence of multiple conflicting cues (unreliable association between visual and auditory information) may have automatically turned off their learning mode by decreasing their attentional resources during the random condition. However, future studies are needed to explore this hypothesis further.

An important concern about neural oscillations is whether the external stimuli directly induce them by phase-resetting or whether they merely reflect highly phase-locked brain responses to the repeated stimulus (see Obleser and Kayser, 2019 for a discussion). This debate is especially relevant as neural oscillations have been proposed to play a mechanistic role in language processing and speech segmentation (Giraud and Poeppel, 2012). Some researchers tried to discern the issues using indirect measures such as sub-threshold stimulation (Oever et al., 2017) and found that oscillation activity still arises in this condition. Others studied the oscillation activity after the stimulus presentation based on the idea that an oscillator would continue after the presentation of the stimulus (Lakatos et al., 2013). Recently, Doelling et al. (2019) confirmed this prediction by showing that the phase lag between the stimulus and the neural response was more variable in an evoked model than in an oscillatory model. Taken together, these findings support the hypothesis that neural oscillations reflect the coupling of an intrinsic oscillator and external stimuli rather than a succession of evoked responses by the stimuli.

Some methodological limitations are worth considering. First, volume conduction may introduce spurious correlations in the phase synchronization analysis when performed in the sensor space (Bastos and Schoffelen, 2016). However, as done in recent MEEG studies (Chaturvedi et al., 2019; Imperatori et al., 2020; Zakharov et al., 2020; Reiterer et al., 2011), we used the wPLI, which is known to decrease the impact of these confounding factors (Vinck et al., 2011). Second, as opposed to effective connectivity measures, the PLI used here provides an index of coordinated activity only and does not imply causal or directional influence as Granger causality does (Hesse et al., 2003; Brovelli et al., 2004). Third, we analyzed functional connectivity patterns at the sensor level, which allows the characterization of functional interactions between different electrodes but does not provide direct insights about the neural sources of those signals (Friston, 1994, 2011). Thus, further studies with more electrodes will allow source-level analyses that may extend our results at the scalp level.

5. Conclusion

To our knowledge, this is the first study investigating the functional role of brain oscillations and long-distance scalp-based phase synchronization during multimodal SL in children. Our results show (i) that 9-year-old children can simultaneously segment words and associate them onto meaning, (ii) that the presence of multimodal cues triggers speech segmentation and meaning-mapping process in children, (iii) that long-distance EEG phase synchronization between frontal and occipital pairs of electrodes contribute to multimodal word-learning. These results favor an integrative view of language learning, where multi-sensory cues and attentional mechanisms interact to enhance learning.

Data statement

The data that support the findings of this study are available on reasonable request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

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.

Acknowledgments

We thank all the children and their families for participating in the study. We also thank the directors and teachers of the two schools where the children were tested and the two anonymous reviewers for their constructive comments on a previous version of the manuscript. We thank D. Cucurell, G. Cardonna, and H. Tajadura for their help in different steps of the study. This research has been supported by a Spanish MINECO project (PSI2015-69132P) to CF. Finally, this work, carried out within the Institut Convergence ILCB (ANR-16-CONV-0002), has benefited from support from the French government, managed by the French National Agency for Research (ANR) and the Excellence Initiative of Aix-Marseille University (A*MIDEX).

References

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., Merzenich, M.M., 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc. Natl. Acad. Sci. 98 (23), 13367–13372.

Arciuli, J., Simpson, I.C., 2011. Statistical learning in typically developing children: the role of age and speed of stimulus presentation. Dev. Sci. 14 (3), 464–473.

Aslin, R.N., Saffran, J.R., Newport, E.L., 1998. Computation of conditional probability statistics by 8-month-old infants. Psychol. Sci. 9 (4), 321–324.

Assaneo, M.F., Ripollés, P., Orpella, J., Lin, W.M., de Diego-Balaguer, R., Poeppel, D., 2019. Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. Nat. Neurosci. 22 (4), 627–632.

Axelsson, E.L., Churchley, K., Horst, J.S., 2012. The right thing at the right time: why ostensive naming facilitates word learning. Front. Psychol. 3, 88.

Bastos, A.M., Schoffelen, J.M., 2016. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. Front. Syst. Neurosci. 9, 175.

Batterink, L.J., Paller, K.A., 2017. Online neural monitoring of statistical learning. Cortex 90, 31–45.

Batterink, L.J., Reber, P.J., Neville, H.J., Paller, K.A., 2015. Implicit and explicit contributions to statistical learning. J. Mem. Lang. 83, 62–78.

Begus, K., Gliga, T., Southgate, V., 2016. Infants' preferences for native speakers are associated with an expectation of information. Proc. Natl. Acad. Sci. 113 (44), 12397–12402.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57 (1), 289–300.

Bloom, P., 2000. How Children Learn the Meanings of Words. MIT Press, Cambridge.

Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., Bressler, S.L., 2004. Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality. Proc. Natl. Acad. Sci. 101 (26), 9849–9854.

Buiatti, M., Peña, M., Dehaene-Lambertz, G., 2009. Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. Neuroimage 44 (2), 509–519.

Buzsáki, G., Dragun, A., 2004. Neuronal oscillations in cortical networks. Science 304, 1926–1929.
Canolty, R.T., Edwards, E., Dalal, S.S., Soltani, M., Nagarajan, S.S., Kirsch, H.E., Berger, M.S., Barbaro, N.M., Knight, R.T., 2006. High gamma power is phase-locked to theta oscillations in human neocortex. Science 313 (5793), 1626–1628.
Carey, S., Bartlett, E., 1978. Acquiring a single new word. Pap. Rep. Child Lang. Dev. 15, 17–29.
Chall, J.S., 1983. Stages of Reading Development. McGraw-Hill, New York.
Chaturvedi, M., Bogaarts, J.G., Kozak, V.V., Hatz, F., Gschwandtner, U., Meyer, A., et al., 2019. Phase lag index and spectral power as QEEG features for identification of patients with mild cognitive impairment in Parkinson's disease. Clin. Neurophysiol. 130 (10), 1937–1944.
Coch, D., 2015. The N400 and the fourth grade shift. Dev. Sci. 18 (2), 254–269.
Covic, A., Keitel, C., Porcu, E., Schröger, E., Müller, M.M., 2017. Audio-visual synchrony and spatial attention enhance processing of dynamic visual stimulation independently and in parallel: a frequency-tagging study. Neuroimage 161, 32–42.
Cunillera, T., Càmara, E., Toro, J.M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., Pujol, J., Rodríguez-Fornells, A., 2009. Time course and functional neuroanatomy of speech segmentation in adults. Neuroimage 48 (3), 541–553.
Cunillera, T., Laine, M., Càmara, E., Rodríguez-Fornells, A., 2010a. Bridging the gap between speech segmentation and word-to-world mappings: evidence from an audio-visual statistical learning task. J. Mem. Lang. 63 (3), 295–305.
Cunillera, T., Càmara, E., Laine, M., Rodríguez-Fornells, A., 2010b. Speech segmentation is facilitated by visual cues. Q. J. Exp. Psychol. 63 (2), 260–274.
Cunillera, T., Fuentemilla, L., Periañez, J., Marco-Pallarès, J., Krämer, U.M., Càmara, E., Münte, T.F., Rodríguez-Fornells, A., 2012. Brain oscillatory activity associated with task switching and feedback processing. Cogn. Affect. Behav. Neurosci. 12 (1), 16–33.
De Diego-Balaguer, R.D.D., Toro, J.M., Rodriguez-Fornells, A., Bachoud-Lévi, A.C., 2007. Different neurophysiological mechanisms underlying word and rule extraction from speech. PLoS One 2 (11), e1175.
Dickinson, D.K., 1984. First impressions: children's knowledge of words gained from a single exposure. Appl. Psycholinguist. 5 (4), 359–373.
Ding, N., Simon, J.Z., 2012. Emergence of neural encoding of auditory objects while listening to competing speakers. Proc. Natl. Acad. Sci. 109 (29), 11854–11859.
Doelling, K.B., Assaneo, M.F., Bevilacqua, D., Pesaran, B., Poeppel, D., 2019. An oscillator model better predicts cortical entrainment to music. Proc. Natl. Acad. Sci. 116 (20), 10113–10121.
Doesburg, S.M., Vinette, S.A., Cheung, M.J., Pang, E.W., 2012. Theta-modulated gamma-band synchronization among activated regions during a verb generation task. Front. Psychol. 3, 195.
Embury, C.M., Wiesman, A.I., Proskovec, A.L., Mills, M.S., Heinrichs-Graham, E., Wang, Y.P., Calhoun, V.D., Stephen, M.J., Wilson, T.W., 2019. Neural dynamics of verbal working memory processing in children and adolescents. NeuroImage 185, 191–197.
Farthouat, J., Franco, A., Mary, A., Delpouve, J., Wens, V., De Beeck, M.O., De Tiège, X., Peigneux, P., 2017. Auditory magnetoencephalographic frequency-tagged responses mirror the ongoing segmentation processes underlying statistical learning. Brain Topogr. 30 (2), 220–232.
François, C., Tillmann, B., Schön, D., 2012. Cognitive and methodological considerations on the effects of musical expertise on speech segmentation. Ann. N. Y. Acad. Sci. 1252 (1), 108–115.
François, C., Chobert, J., Besson, M., Schön, D., 2013. Music training for the development of speech segmentation. Cereb. Cortex 23 (9), 2038–2043.
François, C., Jaillet, F., Takerkart, S., Schön, D., 2014. Faster sound stream segmentation in musicians than in nonmusicians. PLoS One 9 (7), e101340.
François, C., Teixidó, M., Takerkart, S., Agut, T., Bosch, L., Rodriguez-Fornells, A., 2017a. Enhanced neonatal brain responses to sung streams predict vocabulary outcomes by age 18 months. Sci. Rep. 7 (1), 1–13.
François, C., Cunillera, T., Garcia, E., Laine, M., Rodriguez-Fornells, A., 2017b. Neurophysiological evidence for the interplay of speech segmentation and word-referent mapping during novel word learning. Neuropsychologia 98, 56–67.
Fries, P., 2005. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn. Sci. 9 (10), 474–480.
Friston, K.J., 1994. Functional and effective connectivity in neuroimaging: a synthesis. Hum. Brain Mapp. 2 (1-2), 56–78.
Friston, K.J., 2011. Functional and effective connectivity: a review. Brain Connect. 1 (1), 13–36.
Frost, R., Armstrong, B.C., Siegelman, N., Christiansen, M.H., 2015. Domain generality versus modality specificity: the paradox of statistical learning. Trends Cogn. Sci. 19 (3), 117–125.
Gerken, L., Balcomb, F.K., Minton, J.L., 2011. Infants avoid' labouring in vain'by attending more to learnable than unlearnable linguistic patterns. Dev. Sci. 14 (5), 972–979.
Gibson, J.J., Gibson, E.J., 1955. Perceptual learning: differentiation or enrichment? Psychol. Rev. 62 (1), 32.
Giraud, A.L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15 (4), 511.
Glicksohn, A., Cohen, A., 2013. The role of cross-modal associations in statistical learning. Psychon. Bull. Rev. 20 (6), 1161–1169.
Graf-Estes, K., Evans, J.L., Alibali, M.W., Saffran, J.R., 2007. Can infants map meaning to newly segmented words? Statistical segmentation and word learning. Psychol. Sci. 18 (3), 254–260.
Hay, J.F., Pelucchi, B., Graf Estes, K.G., Saffran, J.R., 2011. Linking sounds to meanings: infant statistical learning in a natural language. Cogn. Psychol. 63, 93–106.
Helfrich, R.F., Fiebelkorn, I.C., Szczepanski, S.M., Lin, J.J., Parvizi, J., Knight, R.T., Kastner, S., 2018. Neural mechanisms of sustained attention are rhythmic. Neuron 99 (4), 854–865.
Henderson, L., Chard, J., 1980. The reader's implicit knowledge of orthographic structure. In: Frith, U. (Ed.), Cognitive Processes in Spelling. Academic Press, New York, pp. 85–116.
Henin, S., Turk-Browne, N., Friedman, D., Liu, A., Dugan, P., Flinker, A., Doyle, W., Devinsky, O., Melloni, L., 2021. Learning hierarchical sequence representations across human cortex and hippocampus. Sci. Adv. 7 (8), eabc4530.
Hesse, W., Möller, E., Arnold, M., Schack, B., 2003. The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies. J. Neurosci. Methods 124 (1), 27–44.
Hillyard, S.A., Vogel, E.K., Luck, S.J., 1998. Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 353 (1373), 1257–1270.
Hollich, G., Hirsh-Pasek, K., Golinkoff, R.M., 2000. I. What does it take to learn a word? Monogr. Soc. Res. Child Dev. 65 (3), 1–16.
Horst, J.S., Samuelson, L.K., 2008. Fast mapping but poor retention by 24-month-old infants. Infancy 13 (2), 128–157.
Imperatori, L.S., Cataldi, J., Betta, M., Ricciardi, E., Ince, R.A., Siclari, F., Bernardi, G., 2020. Cross-participant prediction of vigilance stages through the combined use of wPLI and wSMI EEG functional connectivity metrics. SleepJ 1–14.
Janacsek, K., Fiser, J., Nemeth, D., 2012. The best time to acquire new skills: age-related differences in implicit sequence learning across the human lifespan. Dev. Sci. 15 (4), 496–505.
Jaswal, V.K., Neely, L.A., 2006. Adults don't always know best: preschoolers use past reliability over age when learning new words. Psychol. Sci.
Johns, B.T., Dye, M., Jones, M.N., 2016. The influence of contextual diversity on word learning. Psychon. Bull. Rev. 23 (4), 1214–1220.
Johnson, E.K., Tyler, M.D., 2010. Testing the limits of statistical learning for word segmentation. Dev. Sci. 13 (2), 339–345.
Juhasz, D., Nemeth, D., Janacsek, K., 2019. Is there more room to improve? The lifespan trajectory of procedural learning and its relationship to the between-and within-group differences in average response times. PLoS One 14 (7), e0215116.
Kabdebon, C., Pena, M., Buiatti, M., Dehaene-Lambertz, G., 2015. Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. Brain Lang. 148, 25–36.
Keitel, C., Andersen, S.K., Quigley, C., Müller, M.M., 2013. Independent effects of attentional gain control and competitive interactions on visual stimulus processing. Cereb. Cortex 23 (4), 940–946.
Keitel, C., Quigley, C., Ruhnau, P., 2014. Stimulus-driven brain oscillations in the alpha range: entrainment of intrinsic rhythms or frequency-following response? J. Neurosci. 34 (31), 10137–10140.
Kim, Y.J., Grabowecky, M., Paller, K.A., Muthu, K., Suzuki, S., 2007. Attention induces synchronization-based response gain in steady-state visual evoked potentials. Nat. Neurosci. 10 (1), 117–125.
Koenig, M.A., Harris, P.L., 2005. Preschoolers mistrust ignorant and inaccurate speakers. Child Dev. 76 (6), 1261–1277.
Korkman, M., Kirk, U., Kemp, S., 2007. NEPSY-II. Madrid: Pearson. [Spanish adaptation by Universidad de Sevilla, FIVAN and Departamento I_D Pearson Clinical and Talent Assessment (2014). Pearson Publ., Madrid.
Lachaux, J.P., Rodriguez, E., Martinerie, J., Varela, F.J., 1999. Measuring phase synchrony in brain signals. Hum. Brain Mapp. 8 (4), 194–208.
Lakatos, P., Shah, A.S., Knuth, K.H., Ulbert, I., Karmos, G., Schroeder, C.E., 2005. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J. Neurophysiol. 94 (3), 1904–1911.
Lakatos, P., Musacchia, G., O'Connel, M.N., Falchier, A.Y., Javitt, D.C., Schroeder, C.E., 2013. The spectrotemporal filter mechanism of auditory selective attention. Neuron 77 (4), 750–761.
Lefton, L.A., Spragins, A.B., 1974. Orthographic structure and reading experience affect the transfer from iconic to short-term memory. J. Exp. Psychol. 103, 775–781.
Lim, S.J., Lacerda, F., Holt, L.L., 2015. Discovering functional units in continuous speech. J. Exp. Psychol. Hum. Percept. Perform. 41 (4), 1139.
Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54 (6), 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 (8), e1000445.
Markman, E.M., 1990. Constraints children place on word meanings. Cogn. Sci. 14 (1), 57–77.
McNealy, K., Mazziotta, J.C., Dapretto, M., 2006. Cracking the language code: neural mechanisms underlying speech parsing. J. Neurosci. 26 (29), 7629–7639.
Moore, C., Angelopoulos, M., Bennett, P., 1999. Word learning in the context of referential and salience cues. Dev. Psychol. 35 (1), 60.
Morgan, S.T., Hansen, J.C., Hillyard, S.A., 1996. Selective attention to stimulus location modulates the steady-state visual evoked potential. Proc. Natl. Acad. Sci. 93 (10), 4770–4774.
Nozaradan, S., 2014. Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. Philos. Trans. R. Soc. B Biol. Sci. 369 (1658), 20130393.
Nozaradan, S., Peretz, I., Mouraux, A., 2012. Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. J. Neurosci. 32 (49), 17572–17581.
Obleser, J., Kayser, C., 2019. Neural entrainment and attentional selection in the listening brain. Trends Cogn. Sci. 23 (11), 913–926.

Oever, S., Schroeder, C.E., Poeppel, D., Van Atteveldt, N., Mehta, A.D., Mégevand, P., Zion-Golumbi, E., Zion-Golumbic, E., 2017. Low-frequency cortical oscillations entrain to subthreshold rhythmic auditory stimuli. J. Neurosci. 37 (19), 4903–4912.
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.
Ordin, M., Polyanskaya, L., Soto, D., Molinaro, N., 2020. Electrophysiology of statistical learning: exploring the online learning process and off-line learning product. Eur. J. Neurosci. 51 (9), 2008–2022.
Panda, E.J., Emami, Z., Valiante, T.A., Pang, E.W., 2020. EEG phase synchronization during semantic unification relates to individual differences in children's vocabulary skill. Dev. Sci., e12984
Parasuraman, R., 1980. Effects of information processing demands on slow negative shift latencies and N100 amplitude in selective and divided attention. Biol. Psychol. 11 (3–4), 217–233.
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 (6), 1378–1387.
Plunkett, K., 1997. Theories of early language acquisition. Trends Cogn. Sci. 1 (4), 146–153.
Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59 (3), 2142–2154.
Pozuelos, J.P., Paz-Alonso, P.M., Castillo, A., Fuentes, L.J., Rueda, M.R., 2014. Development of attention networks and their interactions in childhood. Dev. Psychol. 50 (10), 2405.
Räsänen, O., Rasilo, H., 2015. A joint model of word segmentation and meaning acquisition through cross-situational learning. Psychol. Rev. 122 (4), 792.
Raviv, L., Arnon, I., 2018. The developmental trajectory of children's auditory and visual statistical learning abilities: modality-based differences in the effect of age. Dev. Sci. 21 (4), e12593.
Reber, A.S., 1967. Implicit learning of artificial grammars. J. Verbal Learn. Verbal Behav. 6 (6), 855–863.
Reiterer, S.M., Pereda, E., Bhattacharya, J., 2011. On a possible relationship between linguistic expertise and EEG gamma band phase synchrony. Front. Psychol. 2, 334.
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., e12947
Sabbagh, M.A., Baldwin, D.A., 2001. Learning words from knowledgeable versus ignorant speakers: links between preschoolers' theory of mind and semantic development. Child Dev. 72 (4), 1054–1070.
Saffran, J.R., Aslin, R.N., Newport, E.L., 1996. Statistical learning by 8-month-old infants. Science 274 (5294), 1926–1928.
Shufaniya, A., Arnon, I., 2018. Statistical learning is not age-invariant during childhood: performance improves with age across modality. Cogn. Sci. 42 (8), 3100–3115.
Shukla, M., White, K.S., Aslin, R.N., 2011. Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. Proc. Natl. Acad. Sci. 108 (15), 6038–6043.
Smith, L.B., 2000. From knowledge to knowing: real progress in the study of infant categorization. Infancy 1 (1), 91–97.
Smith, L.B., Jones, S.S., Landau, B., Gershkoff-Stowe, L., Samuelson, L., 2002. Object name learning provides on-the-job training for attention. Psychol. Sci. 13 (1), 13–19.
Snodgrass, J.G., Vanderwart, M., 1980. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. J. Exp. Psychol. Hum. Learn. Memory 6 (2), 174.
Stam, C.J., Nolte, G., Daffertshofer, A., 2007. Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. Hum. Brain Mapp. 28 (11), 1178–1193.
Talsma, D., Senkowski, D., Soto-Faraco, S., Woldorff, M.G., 2010. The multifaceted interplay between attention and multi-sensory integration. Trends Cogn. Sci. 14 (9), 400–410.
Taylor, B.K., Embury, C.M., Heinrichs-Graham, E., Frenzel, M.R., Eastman, J.A., Wiesman, A.I., Wang, Y.P., Calhoun, V.D., Stephen, J.M., Wilson, T.W., 2020. Neural oscillatory dynamics serving abstract reasoning reveal robust sex differences in typically-developing children and adolescents. Dev. Cogn. Neurosci. 42, 100770.
Teinonen, T., Fellman, V., Näätänen, R., Alku, P., Huotilainen, M., 2009. Statistical language learning in neonates revealed by event-related brain potentials. BMC Neurosci. 10 (1), 1–8.
Thiessen, E.D., 2010. Effects of visual information on adults' and infants' auditory statistical learning. Cogn. Sci. 34 (6), 1093–1106.
Thut, G., Schyns, P., Gross, J., 2011. Entrainment of perceptually relevant brain oscillations by non-invasive rhythmic stimulation of the human brain. Front. Psychol. 2, 170.
Tomasello, M., 1995. Joint attention as social cognition. Joint attention: Its origins and role in development, 103130.
Vander Ghinst, M., Bourguignon, M., Niesen, M., Wens, V., Hassid, S., Choufani, G., Jouamäki, V., Hari, R., Goldman, S., De Tiège, X., 2019. Cortical tracking of speech-in-noise develops from childhood to adulthood. J. Neurosci. 39 (15), 2938–2950.
Varela, F., Lachaux, J.P., Rodriguez, E., Martinerie, J., 2001. The brainweb: phase synchronization and large-scale integration. Nat. Rev. Neurosci. 2 (4), 229–239.
Vinck, M., Oostenveld, R., Van Wingerden, M., Battaglia, F., Pennartz, C.M., 2011. An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. Neuroimage 55 (4), 1548–1565.
Wang, J., Gao, D., Li, D., Desroches, A.S., Liu, L., Li, X., 2014. Theta–gamma coupling reflects the interaction of bottom-up and top-down processes in speech perception in children. Neuroimage 102, 637–645.
Wechsler, D., Corral, S., 2015. WISC-IV: escala de inteligencia de Wechsler para niños-IV. Pearson.
Wilkinson, K.S., Houston-Price, C., 2013. Once upon a time, there was a pulchritudinous princess…: the role of word definitions and multiple story contexts in children's learning of difficult vocabulary. Appl. Psycholinguist. 34 (3), 591–613.
Woldorff, M.G., Gallen, C.C., Hampson, S.A., Hillyard, S.A., Pantev, C., Sobel, D., Bloom, F.E., 1993. Modulation of early sensory processing in human auditory cortex during auditory selective attention. Proc. Natl. Acad. Sci. 90 (18), 8722–8726.
Worden, M.S., Foxe, J.J., Wang, N., Simpson, G.V., 2000. Anticipatory biasing of visuospatial attention indexed by retinotopically specific α-bank electroencephalography increases over occipital cortex. J. Neurosci. 20 (6). RC63-RC63.
Wöstmann, M., Fiedler, L., Obleser, J., 2017. Tracking the signal, cracking the code: speech and speech comprehension in non-invasive human electrophysiology. Lang. Cogn. Neurosci. 32 (7), 855–869.
Xia, M., Wang, J., He, Y., 2013. BrainNet Viewer: a network visualization tool for human brain connectomics. PLoS One 8 (7), e68910.
Yu, C., Smith, L.B., 2007. Rapid word learning under uncertainty via cross-situational statistics. Psychol. Sci. 18 (5), 414–420.
Yu, C., Smith, L.B., 2012. Embodied attention and word learning by toddlers. Cognition 125 (2), 244–262.
Zakharov, I., Tabueva, A., Adamovich, T., Kovas, Y., Malykh, S., 2020. Alpha band resting-state EEG connectivity is associated with non-verbal intelligence. Front. Hum. Neurosci. 14, 10.