Received: 9 November 2021 | Revised: 31 May 2022 | Accepted: 11 June 2022

DOI: 10.1111/desc.13300

Developmental Science WILEY

RESEARCH ARTICLE

Tracking transitional probabilities and segmenting auditory sequences are dissociable processes in adults and neonates

Lucas Benjamin[1] | Ana Fló[1] | Marie Palu[1] | Shruti Naik[1] | Lucia Melloni[2,3] | Ghislaine Dehaene-Lambertz[1]

[1]Cognitive Neuroimaging Unit, CNRS ERL 9003, INSERM U992, CEA, Université Paris-Saclay, NeuroSpin Center, Gif-sur-Yvette, Île-de-France, France

[2]Department of Neuroscience, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Hessen, Germany

[3]Department of Neurology, NYU Grossman School of Medicine, New York City, New York, USA

Correspondence
Lucas Benjamin, Cognitive Neuroimaging Unit, CNRS ERL 9003, INSERM U992, CEA, Université Paris-Saclay, NeuroSpin center, 91191 Gif/Yvette, France.
Email: lucas.benjamin@cea.fr

Abstract

Since speech is a continuous stream with no systematic boundaries between words, how do pre-verbal infants manage to discover words? A proposed solution is that they might use the transitional probability between adjacent syllables, which drops at word boundaries. Here, we tested the limits of this mechanism by increasing the size of the word-unit to four syllables, and its automaticity by testing asleep neonates. Using markers of statistical learning in neonates' EEG, compared to adult behavioral performances in the same task, we confirmed that statistical learning is automatic enough to be efficient even in sleeping neonates. We also revealed that: (1) Successfully tracking transition probabilities (TP) in a sequence is not sufficient to segment it. (2) Prosodic cues, as subtle as subliminal pauses, enable to recover words segmenting capacities. (3) Adults' and neonates' capacities to segment streams seem remarkably similar despite the difference of maturation and expertise. Finally, we observed that learning increased the overall similarity of neural responses across infants during exposure to the stream, providing a novel neural marker to monitor learning. Thus, from birth, infants are equipped with adult-like tools, allowing them to extract small coherent word-like units from auditory streams, based on the combination of statistical analyses and auditory parsing cues.

KEYWORDS
EEG, language learning, neonates, prosody, sequence learning, statistical learning

Research Highlights

- Successfully tracking transitional probabilities in a sequence is not always sufficient to segment it.
- Word segmentation solely based on transitional probability is limited to bi- or tri-syllabic elements.
- Prosodic cues, as subtle as subliminal pauses, enable to recover chunking capacities in sleeping neonates and awake adults for quadriplets.

1 | INTRODUCTION

One of the main challenges encountered by infants to learn their native language and construct their lexicon is that words are rarely produced in isolation. Instead, words are embedded in sentences with no systematic silence or clear acoustic boundaries between them. Subtle acoustical markers such as the lengthening of the last syllable, pitch change, slowing-down of the syllabic rate and less coarticulation between syllables can signal words ending. But adults rely mainly on lexical knowledge and sentential context to retrieve words in their native language (Mattys et al., 2005) and in an unknown language, they have great difficulty in correctly segmenting sentences into words (Wakefield et al., 1974). However, when the experimental task is simplified by using an artificial stream of concatenated words, these acoustical cues can be used to discover the possible words as shown by their above-chance accuracy in forced-choice tasks (Bagou & Frauenfelder, 2018). Similarly, neonates can detect these subtle variations in a binary situation in which they have to discriminate pseudo-words constituted of syllables either coming from inside a word (e.g., /mati/ from *mathe**mati**cien*) or from two successive words (e.g., */mati/* from *pyja**ma ti**ssé*) (Christophe et al., 1994). However, the relative weights of these markers vary across languages (Ordin et al., 2017) and within a language (i.e., they depend on the position of the word in the sentence and interact with other prosodic features such as lexical stress). Therefore, the robustness of these word-boundary cues is commonly estimated as insufficient for infants to segment natural speech in successive word units.

A second mechanism, based on the analysis of the transitions between syllables, has thus been proposed. Within a word, syllables have a fixed order, whereas any syllable can follow the last syllable of a word. Thus, a word boundary corresponds to a drop in the transition probabilities (TP) between consecutive syllables. To prove that the concept could apply for word learning in infancy, Saffran, Aslin,and Newport (1996) used a mini-language of four tri-syllabic words and tested 8-month-old infants who listened for 3 mins to a continuous stream in which these words were concatenated with a flat intonation. The authors reported that infants were subsequently able to distinguish two different lists of isolated tri-syllables pseudo-words: one corresponding to the Words (i.e., ABC:TP were equal to 1 between each syllable) and the other to PartWords formed by the two last syllables of a word and the first syllable of the next word for example (i.e., BCA':TP were equal to 1 and 0.33). This result has been replicated multiple times (Black & Bergmann, 2017) and extended to non-linguistic stimuli (Saffran et al., 1999; Schön et al., 2008) and to the visual domain (Fiser & Aslin, 2002). Sensitivity to statistics in sequences is also observed in animals (Hauser et al., 2001; James et al., 2020; Toro & Trobalón, 2005) indicating that the capacity of extracting transitional probabilities between subsequent elements is a robust general mechanism. Additionally, it has been reported in asleep neonates (Fló et al., 2019; Fló, Benjamin, et al., 2022; Teinonen et al., 2009); and to some extent, in inattentive adults (Batterink & Choi, 2021; Benjamin et al., 2021; Fernandes et al., 2010; Toro et al., 2005). Yet the limits of this mechanism and the influence of development and expertise on the performances are still poorly known.

One of the limitations of statistical learning, already reported in the literature, is its interaction with alternative segmentation cues (Black & Bergmann, 2017), especially its embedding in prosodic units. A word cannot straddle a prosodic boundary. Therefore, even if two syllables are always presented in succession, they are not attributed to the same word if a prosodic boundary separates them. This property is observed in adults (Shukla et al., 2007) and in 5–8-month-old babies (Johnson & Tyler, 2010; Shukla et al., 2011). This result is not surprising given the importance of prosody to structure the speech signal. A hierarchy of prosodic units (Nespor & Vogel, 2006) roughly parallel to the syntactic tree is used to improve speech comprehension in adults and to favor language acquisition in infants. For example, even at 2 months of age, infants memorize better the phonetic content of a sentence with a well-formed prosodic contour relative to a word list (Mandel et al., 1994). This advantage can be explained because statistical computations are limited to a few elements within the prosodic unit, relieving memory. Prosodic units also provide perceptual anchors, which help infants note the reproducible location of certain words at their edges, such as articles or proper name. Finally, the higher frequency of function words relative to content words has also been proposed as anchors favoring word discovery (Hochmann et al., 2010). To succeed in the complex task of constructing a lexicon from natural speech, infants have a toolbox of procedures at their disposal, whose relative contributions are currently underspecified.

Here we investigated another putative limitation of the statistical learning mechanism: the size of the words that can be learned. In fact, most, if not all, studies in infants have used tri-syllabic words. Is it due to particular experimental choices? Or is there a hard limit to segmentation based on statistical computation, especially in immature infants? If the latter, can subtle prosodic cues rescue segmentation and word learning, allowing memory processes to deploy (Fló et al., 2022)? To investigate these questions, we created a first artificial stream consisting of four quadri-syllabic words, pseudo-randomly concatenated without any prosodic cue, and a second one strictly identical to the first one but with a 25 ms pause between each word, every four syllables. In previous studies of artificial streams with this short pause, adults reported not noticing it and were at chance when they had to choose which of the two streams had pauses (Peña et al., 2002). Nevertheless, pauses significantly improved their performances (Buiatti et al., 2009; Peña et al., 2002). The pause was probably perceived as a vowel lengthening, a universal ending cue for words and musical segments (Tyler & Cutler, 2009). In adults, final syllable lengthening improved tri-syllabic word segmentation (Saffran et al., 1996). The authors proposed a putative hierarchy in using these cues, i.e., infants first rely on transitional probabilities, then notice that syllable lengthening coincides with a word ending to finally learn this new cue. Yet, this hypothesis remains untested because the relative contribution of transitional probabilities and this subtle prosodic cue was not assessed in this study.

We used high-density EEG (128 channels) to evaluate segmentation processes in neonates. EEG allows not only to observe different responses to test-words after learning, but also to track learning while neonates are listening to the artificial stream. As the syllables have exactly the same length, their perception creates a regular evoked

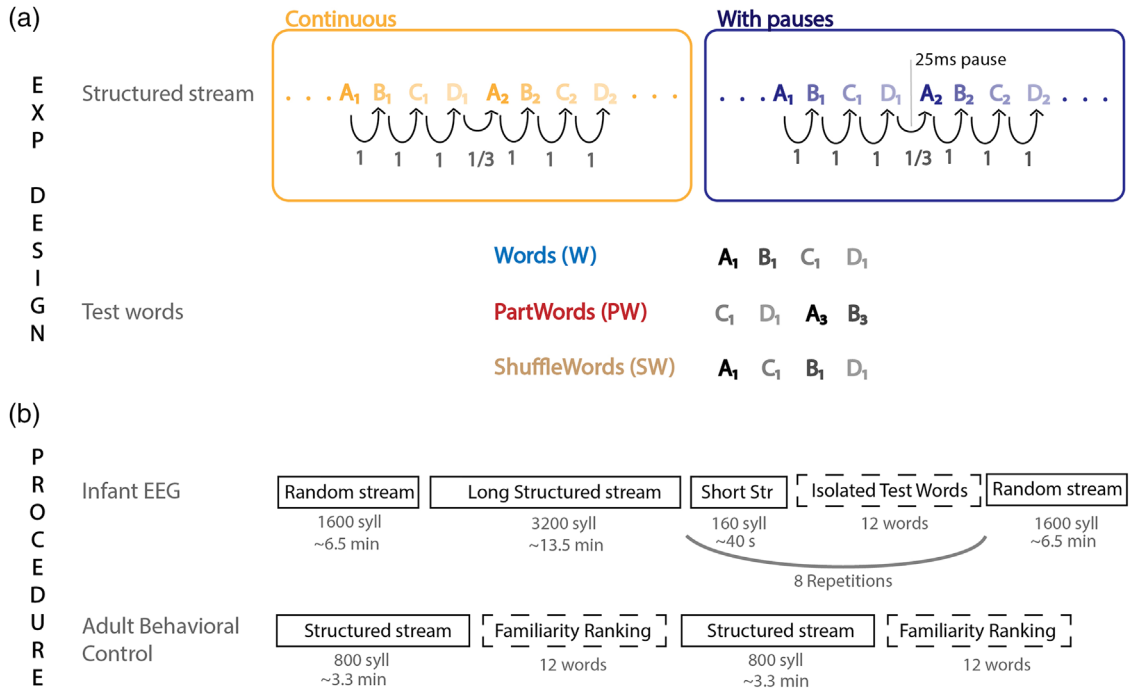

FIGURE 1 (a) Experimental design: Participants were first exposed to a structured stream comprising four quadri-syllabic words (called ABCD) then presented with 3 types of isolated test words. (b) Experimental procedure: Neonates were tested asleep using high-density EEG (128 channels) while they were presented with random stream, structured stream, and isolated words. Short Str: short structured streams were presented to the neonates to maintain learning. Adults were tested on a web platform. After familiarization with the structured stream, adults were asked to rank the familiarity on a scale (1–6). To avoid a bias to quadri-syllabic words, they were also presented with foils corresponding to three other types of bi-syllabic test words (see Section 5 for more details)

response, which is observed as a power increase at the frequency of the syllable presentation. If the syllables are perceived grouped in a quadri-syllabic word, the power should increase at ¼ of the syllable frequency (⅓ if tri-syllabic words are used). Such a power increase at the word frequency has indeed been reported for tri-syllabic words in adults (Batterink & Choi, 2021; Benjamin et al., 2021; Buiatti et al., 2009), in 6–8-month-old infants and in neonates (Choi et al., 2020; Fló et al., 2022; Kabdebon et al., 2015). We also presented neonates with "random" streams constituted of pseudo-randomly concatenated syllables, with and without a pause every four syllables, to control whether the pause itself was sufficient to induce a 4-syllable-rhythm. In adults, inserting such a pause in a random stream did not produce any increase of power at the pause frequency (Buiatti et al., 2009). Therefore, in the case of successful segmentation, we expected a significant power increase at the word frequency in the structured streams relative to the random streams. No change, or perhaps a decrease, was expected at the syllabic rate, in line with previous reports in adults in which perceiving the word induced a decrease of the entrainment at the syllabic rate (Batterink & Choi, 2021; Benjamin et al., 2021).

After the learning phase, three types of test-words were presented in isolation: Words, PartWords, and ShuffleWords (Figure 1). Successful word segmentation is commonly revealed by a significant difference between the measured response to Words and PartWords. In Words, all transitional probabilities between syllables equal 1, while in PartWords (straddling two words), a drop in transitional probabilities indicates an ill-formed word. In ShuffleWords, the two middle syllables of a Word were inverted, violating local position. Thus, while all the transitional probabilities were zero, all syllables were always heard in close proximity during the learning stream. This temporal proximity might induce memory errors and a wrong recognition of ShuffleWords as possible words. Indeed, in longer words of six-syllables, neonates are not able to detect a shuffle of the middle syllables, whereas they detect a shuffle of the edges syllables (Ferry et al., 2016).

Thus, our experimental design provides several markers of transitional probability computation and word segmentation that might be differently associated, opening the possibility to disentangle several steps or hypotheses of this classical learning task. *(H1: TP computation)* If infants computed TP and memorized the TP matrix, they should reject Words from Shuffle- Words (1+1+1 vs. 0+0+0) but marginally Words from PartWords (1+1+1 vs. 1+0.33+1). *(H2: segmentation)* Stream segmentation should create an increase of neural entrainment at the word frequency. *(H3: complete memorization of the word)* should create a difference between words on one side and PartWords and ShuffleWords on the other side. *(H4: memory errors)* If Words are segmented and swap errors occur, ShuffleWords should not differ from Words due to the temporal proximity of the syllables belonging to the same Word. *(H5: first syllable memorization only)*. This hypothesis could explain why words are preferred over PartWords in many statistical learning studies. As the typical trisyllabic paradigm compares Words (ABC) to PartWords (BCA'), the difference observed could result from

the encoding of the first syllable only (A vs. B). In a recent study with tri-syllabic words, we indeed observed an ERP difference between words and PartWords for the first syllable, whereas no difference was recorded when the last syllable was incorrect (Fló, et al., 2022).

Finally, comparing the two groups of neonates, one listening to the stream without pauses (continuous group) and the other to the stream with pauses (with pauses group), should clarify the relative contribution of auditory parsing cues and transitional probabilities to word segmentation at that age. This comparison should disentangle whether pauses rescue segmentation, subsequently allowing the computation of transitional probabilities on smaller segments, or whether the computation of TP is done independently of the segmentation process.

Neonates are two-decades far from a mature state in terms of linguistic abilities but also in terms of memory capacities. To our knowledge, no adult equivalent of the paradigm proposed here was available. Thus, we collected adult behavioral data as a mature model of the mechanisms we explored in neonates. We adapted the paradigm to collect behavioral responses via a web-based procedure and shortened the habituation to avoid over-learning already reported in similar experiments in adults (Peña et al., 2002) (see Figure 1). Despite different procedures and learning indicators, the results were surprisingly congruent with those obtained in infants, especially showing comparable limitations.

2 | RESULTS

2.1 | Adults

Two groups of adults were tested online on a web platform ($n = 43$). After having listened to the continuous stream, or to the stream with 25 ms subliminal pauses depending on the group, the participants had to judge the familiarity of three types of words (Words, Part-Words and Shuffle-Words). Exposure (3.3 mn) and test were performed two times and results of the two tests were aggregated (see Section 5 and Figure S5 for results in each test). We analyzed the responses by items in a linear mixed-effects model in each group with FDR correction. For the stream without subliminal pause at the end of the word, Words and Part-Words were similarly rated ($p = 0.26$) and estimated more familiar than the ShuffleWords (W vs. SW $p < 0.001$, PW vs. SW $p < 0.01$). When subliminal pauses were added at the end of the words in the stream, all types of words were ranked differently (all $p < 0.001$) with the following order: Words were judged more familiar than PartWords themselves more familiar than ShuffleWords (Figure 1a,b). To better visualize the difference in segmentation performances between the two groups, we calculated the difference in mean familiarity ranking given by each participant to Words and PartWords, and performed an unpaired unidirectional, *t*-test $t(41) = 2.3$, $p = 0.013$. The segmentation effect, seen as a positive value in Figure 2c, was larger when subliminal pauses were present (Figure 2c).

Thus, adults were able to distinguish Words from PartWords, indicating that they had correctly segmented the stream only if helped by a subliminal acoustic cue. Yet even when there was no pause, they rejected ShuffleWords because of null transitional probabilities.

2.2 | Infant EEG experiment

EEG was recorded in two groups of healthy full-term neonates ($n = 52$ after rejection procedure, see Section 5) while they were listening to streams without pauses for the first group and with 25 ms pauses every four syllables for the second group. For each group, neonates were exposed first to ~7 min of a "random" stream in which syllables were randomly concatenated with a flat TP of 0.33 with and without a subliminal pause every four syllables depending on the group, followed by ~13.5 min of the word-structured stream. After the exposure learning phase, a test phase followed in which they were exposed to isolated quadri-syllables sequences (Words, PartWords, and ShuffleWords). To avoid interference with learning in the testing phase and to reinforce the learning of the structured materials, 40s-short segments of the structured stream were presented every 12 words during this phase. Finally, another ~7 min of the random stream was presented again after this testing phase. The division of the random stream into two periods was done to avoid a time confound in the comparison between random and structured streams. We used a longer exposure time than usual statistical learning experiments in order to perform pattern similarity analyses as done by Henin et al (2021).

2.2.1 | Neural entrainment

As described in other studies on neural entrainment (Fló, et al., 2022; Hochmann et al., 2010; Kabdebon et al., 2015), there was a significant increase in power and Phase Locking Value (PLV) at the frequency of the syllables presentation compared to neighboring frequencies in both groups (continuous and with pauses) and stream types (random and structured) (all *p*s < 0.05 FDR corrected). No interaction was observed between groups and streams indicating a similar signal-to-noise ratio and comparable experimental conditions in the two groups (all *p*s > 0.05 FDR corrected). The power analysis for each condition and group is presented in the supplementary material.

Stream segmentation should be revealed by a significant increase of power and/or PLV at the frequency of the words (i.e., ¼ of the syllabic frequency) relative to the random stream. In the first group (continuous stream), we failed to find this result, contrary to the second group (stream with pause), in whom a significant increase of both power and PLV was observed in several electrodes. Finally, the interaction between groups and streams was significant for both power and PLV on several electrodes (Figure 3, all *p*s < 0.05 cluster corrected, Figure S2 for the results for each stream in each group).

It has been described that the power at the syllabic rate decreased when adults segment the stream (Buiatti et al., 2009). However, we did not find any modulation of the power or PLV at the syllabic frequency in the structured stream compared to the random one. We also performed a time-course analysis of the neural entrainment at the

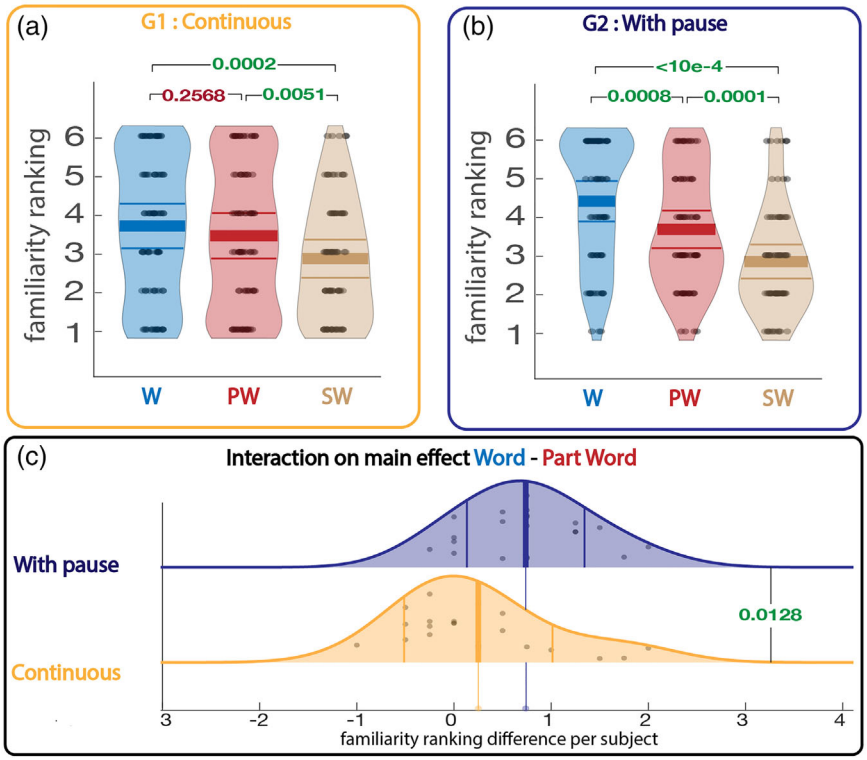

FIGURE 2 Adult behavioral results. (a,b) Results of the familiarity ranking tests in adults subjects for each item. Both test sessions were aggregated. Results for each session are presented in Figure S5. (c) Interaction at the subject level between both groups on the main effect of segmenting (Word - PartWord). Dots represent familiarity ranking difference between Words and PartWords in each subject. *p*-Value is estimated using one-way unpaired *t*-test

frequencies of interest over sliding time windows of 2 min with a 1.5 s step, similarly to Fló, et al. (2022). We observed no change at the word frequency along time for either group (Figure S3). The poor signal/noise ratio at these low frequencies might explain the poor sensitivity of this analysis.

2.2.2 | Between-subjects correlation analysis

Because the exact same stream was used in each participant, we were able to analyze whether learning increased the global neural synchronization between neonates beyond neural entrainment. To do so, we tested whether the correlation between participants increased over time more when they listened to the structured stream. We then compared at each time the topography of each subject with the average of the other subjects' topographies at that time. We observed a progressive increase of the mean correlation between subjects in neural activity only in the second group with pauses (Figure 4a Left). Indeed, the increase was higher for the second group (with pauses) than the first one (continuous) (cluster corrected $p < 0.01$, time [88-820]s; Figure 4a). During the random streams, the correlations were flat relative to baseline in both groups (Figure 4b Left). To confirm this effect, we computed a linear regression of the variation of subject correlation with the group with time at the subject level during each stream and compared the slopes in the two groups. Only when neonates listened to the structured stream with pauses (second group), the slope was significantly positive and significantly greater than the same measure in the continuous group (Figure 4a Right). No difference was observed during the random streams (Figure 4b Right).

2.2.3 | Syllable pattern similarity analysis

In a recent paper, Henin et al. (2021) proposed that pattern similarity between syllables can vary with learning in a similar task. More specifically, using electrocorticography in epileptic adult patients who listened to a structured stream composed of the concatenation of four trisyllabic words, they computed different patterns of similarity between the 12 syllables. They took advantage of the high spatial resolution of electrocorticography and observed different clusters of electrodes sensitive either to TP transitions (low vs. high TP), the ordinal position (1st vs. 2nd vs. 3rd syllable), or the word identity (word 1 vs. word 2 vs. word 3 vs. word 4) in different brain areas. We computed a similar analysis on the responses to the syllables during the structured stream and showed that the similarity pattern for syllables belonging to the same words was significantly increased in the pause group compared to the continuous group ($p = 0.012$). However, we failed to find an increase in pattern similarity for low TP (DA') and ordinal position (AA', BB', CC', and DD') between the two groups. To investigate if the significant increase in similarity for syllables belonging to the same word was only due to an increase in high TP pairs or all pairs belonging to the same word, we separated the word condition in two sub-conditions: Consecutive (AB, BC, CD) and non-Adjacent (AC, BD & AD). Interestingly, Consecutive and non-adjacent pairs showed a significant increase

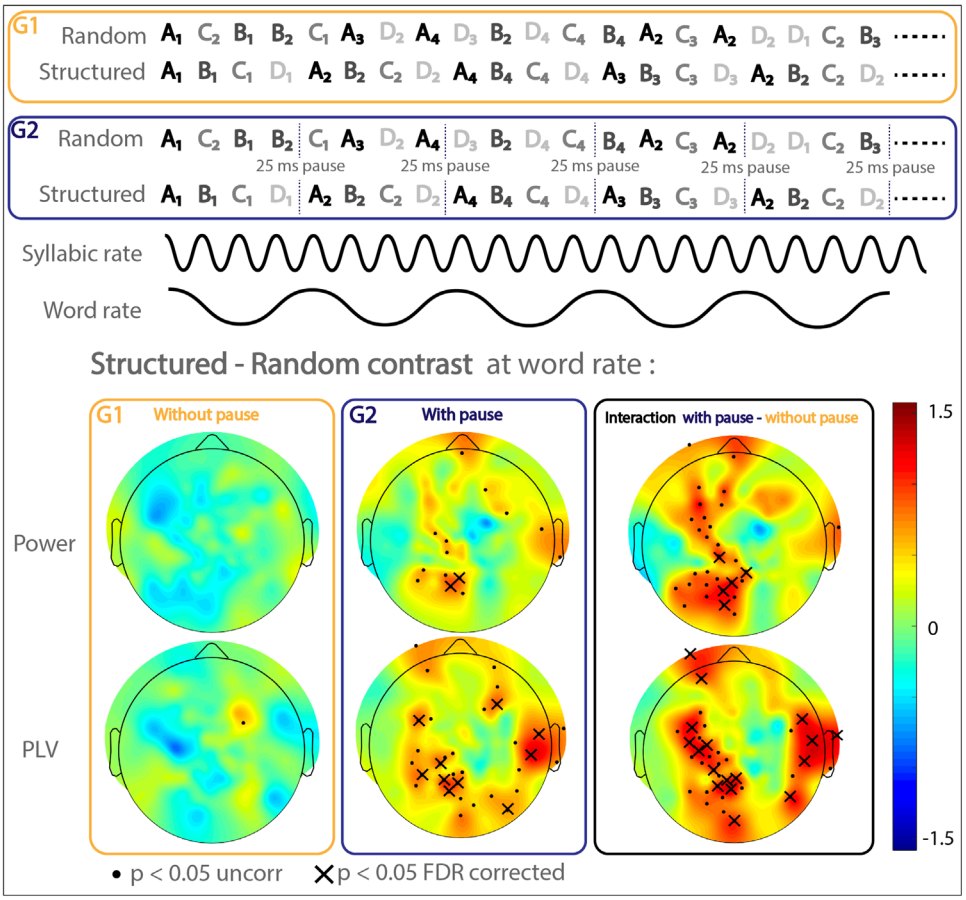

FIGURE 3 Neural entrainment analyses at the word frequency in the structured stream minus the random stream (Power and phase locking value [PLV]) in the two groups of neonates (continuous and with pauses). Top rows: the presentation of stimuli with a fixed duration evoked a reproducible time-locked neural response that can be recovered as a neural oscillation at the frequency of stimulation. If infants segment the structure streams based on the quadri-syllabic words, an increase at the word frequency should be observed relative to the random stream. It is what is seen in the second group of neonates (with pauses) who listened to the streams with sub-liminal pauses. The acoustical effect of pauses was controlled by also adding a pause every four syllables in the random stream in this group. The last column shows the interaction between groups and frequencies. Dots locate the electrodes showing a significant result at $p < 0.05$ uncorrected, and larger dots after cluster correction (cluster $p < 0.05$). Power during structured and random streams are presented separately in each group in Figure S2

in pattern similarity (both $p < 0.05$ FDR corrected) with pauses compared to the continuous group. The differences in pattern similarity between the two groups for each condition are reported in Figure 4c.

2.2.4 | ERP analysis

For this analysis, we used the auditory localizer to determine regions of interest (ROI) for each group (see Section 5). A click preceding each word allows to isolate the two poles of the auditory response in each group and the time series of the electrodes of these data-driven ROIs were averaged together before we performed permutation cluster-based analyses on time as implemented in Fieldtrip (Oostenveld et al., 2011). In the continuous group, the response was flatter for ShuffleWords compared to the other two conditions. When the conditions were contrasted two by two, we only observed that Word versus ShuffleWord tended to differ during two time-periods in the frontal ROI ([1308–1652]ms, $p = 0.03$; [1780–2000]ms, $p = 0.05$) but only trending in the occipital ROI ($0.05 < p < 0.1$). Informed by the adult results, we compared Words and PartWords against ShuffleWords—the only condition rejected by the adults. This contrast revealed a significant difference in the frontal ROI (time: [1253–1644]ms, $p = 0.025$), and a trend was observed in the occipital ROI ($0.05 < p < 0.1$) (Figure 5 first panel). In the group with pauses, as shown in the second row of Figure 5, the response to Words seems to stand out from the other two conditions. ERPs to Words indeed significantly differed from those to PartWords in the two ROIs (all $p < 0.05$ Frontal: [1224–1780]ms, Occipital: [1196–1768]ms), and from those to Shuffle words in the occipital ROI ([1196–1768]ms) whereas two clusters were showing a trend ($0.05 < p < 0.1$) in the frontal ROI.

We also tested the interaction between groups and the main effect Words – Part Words and found significant interactions ($p < 0.05$) on both the frontal and occipital ROIs (Figure 5 last panel). Using a permutation cluster-based method (field trip cluster analysis) between 250 and 2000 ms without previous ROI extractions, we obtained similar results (Figure S4).

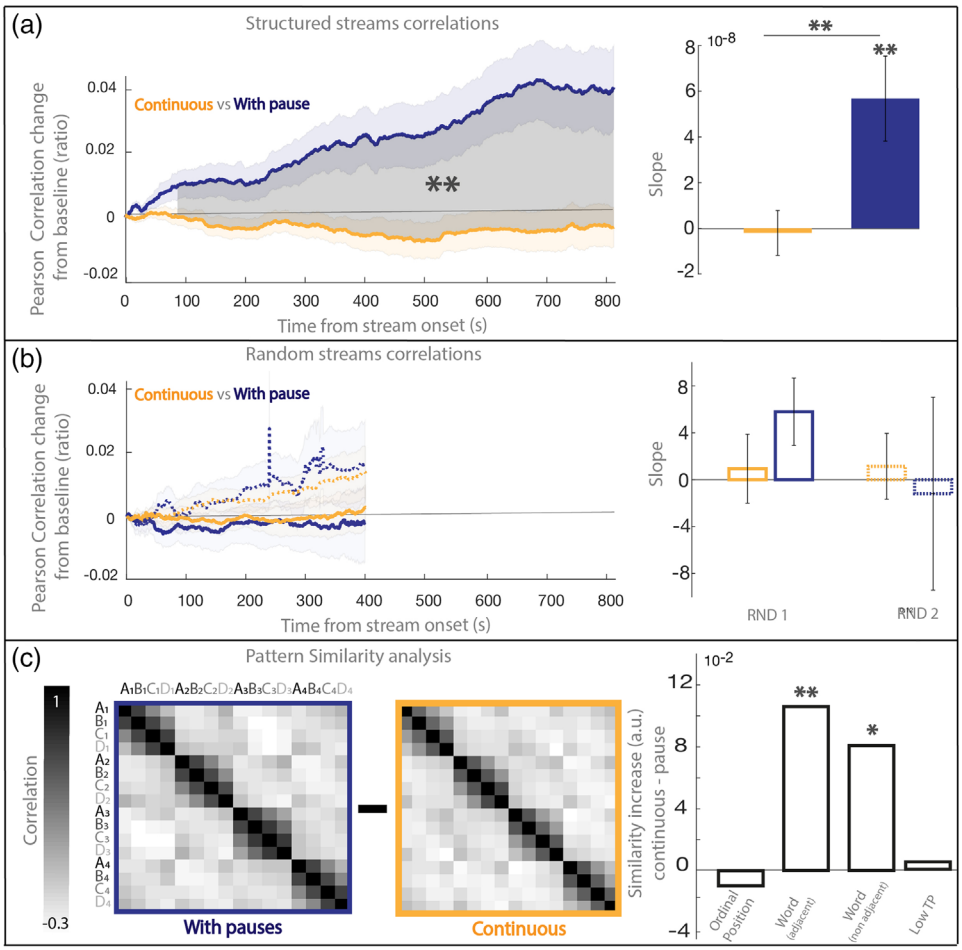

FIGURE 4 Correlation analysis. (a) Comparison of the correlation between neonates in the two groups during the structured stream. Left: Evolution of the correlation across neonates with time. Right: comparison of the slopes of the linear regression with time in each group (orange: continuous, blue: with pauses). (b) Similar analysis for the first (plain lines) and second random (dotted lines) streams in both groups. RND = Random. (c) Pattern Similarity analysis: We computed the increase of pattern similarity between the EEG response to each syllable in the two groups during the structured stream. The similarity significantly increased for syllables belonging to the same word (for adjacent pairs: AB, BC, CD, and non-adjacent pairs AC, BD, and AD)

3 | DISCUSSION

In natural speech, many signal-derived cues might assist segmentation (Wakefield et al., 1974), but none is robustly consistent to be systematically used by infants. Therefore, the computation of the transitions probabilities between syllables has been proposed as a possible solution (Saffran, Aslin, et al., 1996). We presented here a stream comprising quadri-syllabic words to investigate the efficiency of this strategy for longer words. Whereas tri-syllabic words are easily extracted from a flat speech stream using TP between syllables in adults (Saffran, Newport, et al., 1996), infants (Saffran, Aslin, et al., 1996), and even sleeping neonates (Fló, et al., 2022), this single cue seemed insufficient here for quadri-syllabic words even in awake vigilant adults, revealing a clear limitation of the statistical learning mechanism in a segmentation task. Whereas the power increase at the syllabic frequency was large during all streams, we did not record a significant neural entrainment at the word frequency in the continuous group contrary to what we obtained in a similar paradigm with tri-syllabic words. Because the "noise" level due to the background neural activity is exponentially growing with low frequencies in EEG data, notably in neonates, the lack of neural entrainment for quadruplets might have been due to a lack of sensitivity of the method at 1 Hz, compared to 1.33 Hz in the case of triplets. However, the significant word entrainment in the group listening to the stream with pauses and the interaction between both groups confirm that neural entrainment is a sensitive method even at these low frequencies.

The word segmentation failure is also not explained by the higher number of syllables to be memorized (16 syllables here vs. 12 for tri-syllabic words) and the larger word size since the same material, just with the addition of sub-liminal pauses, rescued the word extraction process. Furthermore, we recorded several other indicators of learning in the second group who listened to the stream with pauses: First, the increase in power and phase locking-value observed at the word frequency in the structured stream was not related to the mere presence

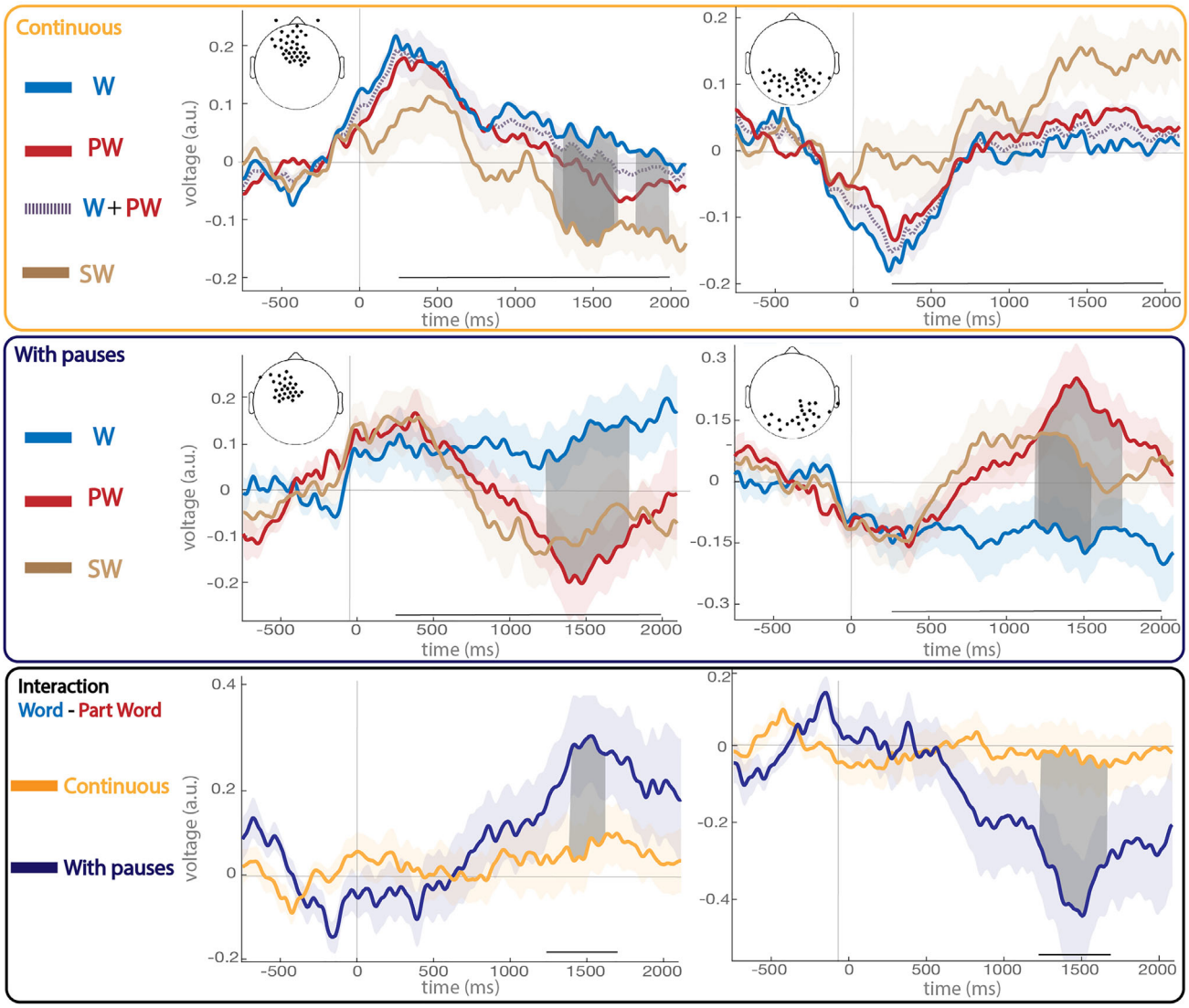

FIGURE 5 ERP analysis. Grand average ERPs to the test-words in both groups and to the word minus part word difference. The ROIs correspond to the two poles of the response to the auditory localizer preceding the test word in each group. Dark gray areas identify significant temporal clusters. Light shaded areas surrounding the thick lines represent the standard error across neonates. W = Words (ABCD), PW = Part Words (CDA'B') and SW = ShuffleWords (ACBD). Gray lines at the bottom of the plots indicate the time windows on which statistics were performed

of a pause but to a genuine learning process, as it was not observed for the random stream that also included pauses every four syllables. Second, neural synchrony increased between participants only for the structured stream with pauses, further suggesting that neonates were following a similar learning process constraining their brain state in this condition. Again, this phenomenon was only observed for the structured stream and not for the random stream with pauses. It underscores that it was not a general difference between the two groups of neonates but was related to the learning process engaged when they listened to the structured stream. Third, ERPs to Words and PartWords were significantly different after the stream with pauses, a classical indicator of word segmentation in this type of paradigm. Finally, adults also ranked Words higher than PartWords when they listened to streams with pauses relative to streams without pauses, confirming an undeniable advantage for the former over the latter. All these indicators of successful segmentation were not only lacking when the pause cue was not present, but for all of them, the differences between the two groups were significant in both infants and adults.

Yet, even if participants were not able to segment the words in the continuous structured stream, both adults and neonates rejected ShuffleWords, which contained the exact same syllables as the Words, but in the wrong order. This result reveals that the participants computed TP and were not misled by the temporal proximity of the syllables, but this computation was not sufficient to trigger stream segmentation. Interestingly, attentive adults appeared not better than sleeping neonates in the task: They also failed to segment the stream without the help of acoustical cues. Thus, tracking TP does not always result in word segmentation.

3.1 | Word segmentation based on statistical learning is limited by the word size

It was proposed that the computation of the TP might be used to segment a speech stream, either through boundary markers—a TP drop creates a prediction error, and the surprise allows to memorize the syllable following the drop (i.e., the first syllable of the following word)—or because adjacent events acquired a similar representation. However, neither the local drop of TP nor the temporal proximity within a chunk were sufficient to structure the stream, not even after 13 mins of exposure when the unit size was four syllables (1s long). On the contrary, sleeping neonates perceived a tri-syllabic rhythm in the same circumstances and only after 2 mins of exposure, and memorized the set of possible first syllables (Fló, et al., 2022). It remains possible that longer exposure to the continuous stream might eventually allow word segmentation. However, compared to the segmentation of tri-syllabic words tested under similar conditions, both neonates and adults had considerable difficulty performing the task with quadri-syllabic words.

The opposite hypothesis is that the neonates might have learned as quickly as in the case of trisyllabic words but that as time passed, this learning faded away because even low probability transitions became familiar. This overlearning effect has been reported in adults (Peña et al., 2002). The analysis of the neural entrainment along time of exposure was not sensitive to figure out the learning timeline even in the group with pauses, probably because of the very low signal to noise ratio in low frequencies. However, the group difference in the correlation between neonates' recordings increased from around the first minute of exposure showing that the two groups were diverging early on between a learning condition (stream with pauses) and a no-learning condition (continuous stream).

3.2 | Rescuing segmentation with sub-liminal pauses

Adding a subliminal pause at the end of the word radically affects the performances at both ages. Although not consciously perceived, pauses act as other word boundary markers (e.g., lengthening of the last syllable, pitch drop) that neonates can perceive (Christophe et al., 1994). Our result demonstrates that such a word boundary marker is not only perceived but is effectively used to segment a stream from birth on, that is, before infants have perceived many isolated words. The use of word-ending cues, at least when it is a pause as here, does not need that infants first learn words as it was postulated (Saffran, Newport, et al., 1996) but is part of the auditory/linguistic perceptive system. This observation is in agreement with the proposal of a hierarchical framework in weighting the multiple segmentation cues (Wakefield et al., 1974) and the subordination of statistical learning to many other cues, such as coarticulation (Fernandes et al., 2007), prosodic contour (Shukla et al., 2007; Shukla et al., 2011), and top-down contextual parsing (Wang et al., 2020). However, in adults as exposure lengthens and the absolute frequency of all possible transitions increased, the familiarity advantage for Words relative to Part-Words created by the pauses faded away (Figure S5) suggesting that the weight attributed to each parameter might not be strictly hierarchized but dependent on the strength of the evidence provided.

We also observed in infants that similarity between syllables within the word was increased relative to the continuous stream without pauses (Figure 4c). Not only similarity between adjacent syllables within a word was stronger in the stream with pauses than without pauses, but similarity also increased between more distant syllables belonging to the same word. We cannot disentangle whether this increase in similarity between syllables in a word induced the segmentation as in a clustering strategy that is opposed to a bracketing strategy in which splitting points are looked for (Swingley, 2005), or the reverse, that is, because syllables were perceived in the same chunk, their similarity increased.

It is also interesting to note that perceiving the stream at a more complex level of representations increased neural synchrony between the neonates. Whereas the syllabic rate itself, which affects many channels (see Figure 2 in appendix), already creates a strong and similar entrainment across participants, it is not this low-level cue that was predominant in the neural synchrony between neonates but the perception of a higher level of organization of the stream. This synchrony measure probably captures a wider cross-subject convergence beyond neural entrainment at the two frequencies of interest in specific channels. It reveals that neonates' brain states are not purely entrained by the physical features of the stimulation, which remain similar along the stream but also constrained by learning mechanisms that led to more synchronous responses across neonates.

Finally, the performances between the test phase during which isolated quadri-syllabic sequences were presented were also massively affected by the stream condition, suggesting that once segmentation was done, memory encoding was improved. Words and PartWords were indeed only discriminated after the stream with pauses. However, in a similar experimental paradigm but after a stream of concatenated tri-syllabic words, Words were recognized since the first syllable (Fló et al., 2022), whereas here, the difference was developing from around 500 ms to become significant only after the end of the word. The lack of first syllable effect was confirmed by the absence of difference between PartWords and ShuffleWords, although the latter started with a correct first syllable. It is also consistent with the lack of similarity increase between both groups within the set of first syllables (Figure 4c right), which contrasts with the result reported in adults by Henin et al. (2021). Thus, contrary to the tri-syllabic stream, the ordinal position of the syllables was not encoded, and the difference between correct and incorrect chunks took longer to develop.

3.3 | Why a sharp distinction between tri and quadri-syllabic words?

The differences, in terms of neural entrainment during familiarization and ERPs during test in infants as the drop of performances in adults, between our two word-segmentation studies (Fló, etal., 2022), in which we used a similar paradigm except that the word size increased from

3 to 4 syllables, raised interesting questions regarding both word segmentation during the stream and subsequent memory encoding of the word unit.

Although this experiment does not directly test this question, we propose that recovering words in a stream is based on short-term memory (STM). Indeed, if TP between syllables can be locally computed within the auditory cortex, the integration of the successive syllables within a word requires a longer temporal window of integration. The sharp difference between tri- and quadri-syllabic words seems reminiscent of the 4 ± 1 unit limit of the auditory STM (Cowan, 2001) and suggests that the TP drop leading to word segmentation might only be noticeable when all the elements of a word plus the next syllable are present at once in the STM. Several studies suggest that adults use STM, and more specifically working memory, in such statistical learning tasks. For instance, their performance improves when speech is slowed down, an observation at odds with a decay-time in a purely sensory buffer that should be detrimental as the time between syllables increases, but in favor of maintenance of the successive syllabic items (Palmer & Mattys, 2016). Performances also drop when participants perform a concurrent two-back task (Palmer & Mattys, 2016) suggesting competition for general resources. These observations are coherent with the activations reported by Henin et al. (2021) along the dorsal linguistic pathway, and notably in the inferior frontal region. In the neonates, no explicit rehearsal was possible because they were asleep and, in any case, unable at that age to repeat syllables, but even in adults, STM effects may remain implicit (Hassin et al., 2009). If statistical learning is improved when adult participants are actively doing the task, the task remains feasible when they are distracted and unaware of the task (Fernandes et al., 2007; Palmer & Mattys, 2016).

The similar drop in performance in neonates and awake linguistically productive adults suggests a structural limitation in the number of items that can be stored in the STM. This limit of four in STM has been proposed as explaining several higher order linguistic observations, such as the size of phrasal verbs and idioms predominantly used in spoken languages such as English, the mean length of continuous discourse without pauses (Green, 2017), and the drop in mutual information scores after four words in many languages (Pothos & Juola, 2007). It also seems compatible with the observed word length inferior to four syllables in many languages (Zipf, 1935, Sigurd et al, 2004), suggesting that this chunk size limitation we observe here might be fine-tuned to real language word size. This limitation also reveals that TP computation might not be robust enough to be the proposed general-purpose mechanism for word segmentation in all speech streams without being complemented by other indices.

If neural entrainment during the stream reflects the chunking and word encoding, the ERPs to the isolated chunks in the test phase tested the participants' recognition and familiarity with the different conditions. In the tri-syllabic experiment (Fló et al., 2022), neonates during the test-phase were no more sensitive to TP (i.e., no distinctive ERP response for triplets containing a TP = 0) and were reacting to an incorrect first syllable. Here, they were rejecting ShuffleWords, thus were still sensitive to TP, but did not react particularly fast to the incorrect first syllable (Word vs. Part-Word), suggesting a more general response to the global familiarity of the word rather than noticing a particular error. In adults, Henin et al. (2021) confirmed using similarity analyses on ECOG recordings, that the ordinal position of the syllables was encoded. Adults are nevertheless better than neonates, encoding not only which syllables were first but also which were second and which were last. Here, we tried similar analyses in the neonates' data despite the sparser resolution of EEG. We observed an increase of similarity of the ERPs to the adjacent and non-adjacent syllables belonging to the same word in the stream with pause compared to the continuous stream. However, we found no evidence of an increase of similarity between the words first syllables. Thus, the particular status of the first syllables observed in neonates in the case of tri-syllablic words (Fló, et al., 2022) had no support in this study when quadri-syllabic words were used. This result might just reflect a lack of power of our analysis, or it might be explained by the difference in perception of the drop of TP in a tri-syllabic word stream. The TP drop, which can induce a surprise following a prediction error, might favor encoding these syllables at a particular position (i.e., the first position of the next word). These results might thus suggest that depending on the segmenting cue, different memory processes are engaged in neonates and that TP computation might favor a more precise encoding of the chunking elements, starting with the first syllable and progressing from one syllable to the next. Such an intriguing hypothesis should be further tested in experiments specifically designed to contrast these two cues and the word-size at this age and also in adults.

3.4 | Similarity between adults and neonate cognitive abilities

Despite very different measuring methods and attentional state in this set of experiments, the results in neonates and adults pointed to similar successes and failures in terms of TP computation and stream segmentation. This is somehow surprising given the fact that many of the structures that support sequence learning (Henin et al., 2021)—hippocampus, dorsal linguistic pathway, the superior temporal region—change rapidly in the first year of life; but the classic assumption that immature means poorly functional is increasingly challenged by brain imaging methods that provide markers of learning in young children. FMRI remains difficult in infants, but some results support the hypothesis of early efficiency despite immaturity. In a recent, paper Ellis et al. tested 3–24 month-old infants on a statistical learning task in the visual domain with fMRI and reported activation in the hippocampus associated with segmentation. Dehaene-Lambertz et al. (2002) reported activations in temporal and frontal areas in 3-month-olds listening to speech showing that regions usually reported in adults during statistical learning tasks (Henin et al., 2021) are, to some extent, already functional in infants.

A major distinction between adults and neonates seems to be the capacity of computing such a task during sleep. Indeed, with both three and quadri-syllabic experiments, we showed that sleeping neonates were able to process and segment the streams, under the correct circumstances. However, recent studies report a learning failure in

sleeping adults even for tri-syllabic words (Farthouat et al., 2018; Batterink & Zhang, 2022), and their learning remained limited to bi-syllabic words, that is, to classical associative learning. Infants might perform better than adults during sleep due to the different organization of sleep-wake cycles. At this age, sleep comprises only two clear stages, quiet (~40% of a sleep cycle at birth) and active sleep (50%–60% of a sleep-cycle at birth) with many micro-arousal periods within and between sleep stages (Scher, 2008). The short periods of wakefulness are immediately followed by active sleep, which is the equivalent of REM sleep in adults. In adults, learning has been shown to exist during REM (Andrillon & Kouider, 2016) and also that a task started during wake can continue during REM (Andrillon et al., 2016), opening the possibilities that neonates might learn and consolidate more efficiently than later, thanks to the closer wake-REM sleep alternations.

3.5 | Methodological considerations

Together, our results show that behavioral subjective ranking and EEG analyses provide powerful tools to investigate statistical learning and segmenting tasks. There was a neat congruency between the behavioral results in adults and the neural markers observed in neonates. Moreover, EEG data enables the investigation of such questions in preverbal and non-verbal subjects with different levels of attention (e.g., neonates, sleeping subjects, comatose patients). Power and PLV during the stream as well as ERP during isolated test words, were already proposed as reliable neural markers in this task (Kabdebon et al., 2015; Fló, Benjamin, et al., 2022). However, to our knowledge, between-subject correlation as a function of time had not been shown to capture learning in infants successfully. Our results confirm that despite the noise in infant EEG data, a significant part of the variance cannot be only explained by low-level bottom-up activation to external stimuli but instead by a more sustained learning effect. Although this first attempt might have been still noisy, we might hope that this method could more accurately quantify the average amount of learning of a group over time or even characterize learning at the subject level. One drawback of this method is that, to compare across subjects, all subjects have to be exposed to the exact same stimuli, which presents a risk of confound in the experimental design. Here we minimized this risk by taking two precautions. We first carefully designed and balanced the auditory material on acoustic aspects (see SI). Secondly, we ran two groups with a minimal change (a subliminal pause every four syllables) so that, if any bias persists, it would be the same in both groups and thus cannot explain differences between groups.

We also implemented what we believe to be an improvement for ERP analysis. Before the presentation of isolated words, we presented a short audio click as an auditory localizer. In this way, we were able to extract ROIs for analysis with a data-driven approach instead of literature driven. We performed a cluster based-permutation analysis on all data against zero during the click presentation to extract the auditory ERP ROI. Moreover, this localizer cluster was representative of the auditory response in this particular group of subjects taking into account non-relevant variations due to (1) Experimental conditions such as the placement of the net on the infant's head which is more variable at this age due to birth-related head deformation, and can introduce between groups differences; (2) eventually long-tail effects of the previous trials on the topography that can affect the baseline.

4 | CONCLUSION

Human neonates display sequence learning abilities even during sleep, based on TP computations and segmenting helped by acoustic/prosodic cues. The similarities with adults' successes and failures were remarkable, revealing early powerful capacities to process speech. A speech stream is not a uniform landscape for infants, but different cues might help them to chunk it into smaller units, opening the possibility to discover the linguistic regularities and the productive properties of speech.

5 | MATERIALS AND METHODS

5.1 | Behavioral experiment

5.1.1 | Participants

A total of 43 adults were recruited via social media and mailing (21 males, age distribution = [18-25]: 9, [25–40]:16, [40–60]: 17, 60+: 1]) with no reported auditory issue or language related troubles. They were randomly assigned to one of the streams with the instruction to carefully listen for ~3 min to a nonsense language composed of nonsense words that they have to learn because they will have to answer questions on the words afterward. The learning/test procedure was repeated twice.

The study was coded in javascript using jspsych toolbox (de Leeuw, 2015) and played audio mp3 pre-loaded and pre-created in MATLAB (see below) to avoid latencies during the presentation. Subjects voluntarily participated on their computer. They were asked to wear headphones, sit in a quiet environment, and stay focused during the whole task.

The Ethical research committee of Paris Saclay University approved the protocol under the reference CER-Paris-Saclay-2019-063.

5.1.2 | Stimuli

All speech stimuli were generated with the MBROLA text-to-speech software (Dutoit et al., 1996) using French diphones. The duration of all syllables was equalized to 250 ms with flat intonation and no coarticulation between syllables. Each experiment was composed of 800 syllables (3.3 mn) of an artificial monotonous stream of concatenated syllables that correspond to the four possible words randomly concatenated with the only restriction that the same word could not be presented twice in a row. The same vocabulary (sixteen syllables) was used in the two streams, with and without pause. In the stream with

pause, a 25-ms pause was inserted every 4-syllables (total duration 3.4 mn). All streams were ramped up and down during the first and last 5 s to avoid the start and end of the streams serving as perceptual anchors. We used the same syllables and words for the infant experiment. To avoid phonological similarity effects that could bias toward one or the other condition, Words and PartWords were reversed for half of the subjects.

In a previous experiment with similar streams with and without 25 ms pauses, Peña et al. (2002) showed that participants were at chance when they had to choose which of the two streams had pauses. To confirm that the pauses were not consciously perceived, eight adults listened to 20 streams (40 syllables – 10 s) presented randomly (10 without pauses and 10 with a 25 ms-pause every four syllables) and were unable to indicate which stream had pauses or not (mean = 49% (range [40, 59]%); $p = 0.89$).

5.1.3 | Procedure

After listening to the structured stream, participants were asked to rank the familiarity of the individual words (from "Completely unfamiliar" to "Completely familiar" on a six-step scale). Learning and test phases were repeated twice. Data of the two tests sessions were aggregated in the main analysis (see separated analysis of each session in Figure S5). Six conditions (three bi-syllabic as foils and three quadri-syllabic conditions) were used to avoid any bias based on the length of the test words, with four trials in each of the six conditions. To avoid phonological similarity effects that could bias toward one or the other condition, participants were assigned to one of two groups where conditions were reversed. Four different pairs of structured streams per group were also generated, and participants were randomly assigned to one pair to avoid any given particularity of a stream driving the results. Three conditions were studied: Words, PartWords, and ShuffleWords. Words corresponded to the words that were embedded in the structured streams (ABCD), while PartWords corresponded to the two last syllables of a word and the two first of another word (CDA'B'). Thus, although PartWords were heard during the structured stream, they violated chunking based on TP. ShuffleWords corresponded to words in which the second and third syllables were inverted, creating a null TP between all syllables (none of the transitions were heard during the structured stream). However, the first and last syllables were correct.

5.1.4 | Data processing

Each answer was converted to a numerical value from 1 (completely unfamiliar) to 6 (completely familiar). The responses to the bisyllabic trials were not considered. All data, from both test sessions, were aggregated together in each group to compute a linear mixed-effects model on items (y ~ condition + (1|subject)) to take the subject effect into account. The *p*-values were then FDR corrected. To compare subjects' segmenting performances for both streams, we computed the mean familiarity ranking for each condition in each subject and subtracted the PartWord ranking from the word ranking within each subject. We then performed a one-way unpaired t-test between the two groups.

5.2 | Infant EEG experiment

5.2.1 | Participants

Two groups of healthy full-term neonates were tested between days 1 and 3. There was no problem during pregnancy and delivery, birthweight > 2500 g, term > 38 wGA, APGAR ≥ 6 and 8 at 1′ and 5′, normal audition tested with otoacoustic emission. Parents provided informed consent, and the Ethical Committee (CPP Tours Region Centre Ouest 1) approved the study (EudraCT/ID RCB: 2017-A00513-50). In the first group (continuous), 34 neonates were tested. Among them, seven were excluded because they did not complete the experimental protocol or technical issues leaving 27 infants (14 males). In the second group, 34 infants were tested (with pauses). Nine were excluded because they did not complete the experimental protocol or technical issues, leaving 25 infants (13 males).

5.2.2 | Stimuli

We used the same 16 isolated syllables generated with MBROLA as in the adult experiment to construct four different streams (structured and random, with and without pause). The random stream consisted of 1600 pseudo-randomly concatenated syllables (6.7 mn). Each syllable could be followed by three others from the pool leading to a flat TP during the stream. This pseudo-random stream offers a more controlled stimulus than the random streams used previously because the TPs were fixed to 1/3 (instead of 1/15), a similar value than the TP between words in the structured stream. The structured stream was comprised of 3200 syllables (13.3 mn). All streams were ramped up and down during the first and last 5 s to prevent the beginning and the end of the streams from being used as anchors. We created only one syllabic order for each stream to obtain learning markers better comparable between infants. For the second group, a pause was added every four syllables in both the structured (duration: 13.7 mn) and the random streams (duration: 6.8 mn). Thus, the sequences were identical for all infants in both groups except for the 25-ms subliminal pauses every four syllables in the second group. Because all subjects had the same auditory materials, we carefully controlled for low-level acoustic-phonetic properties. We equilibrated the characteristics of consonants and vowels in the different words and at the different syllabic positions within words to avoid learning based on low-level acoustic cues (See Figure S1 for more details). As in adults, three types of test words were created: Word (ABCD), PartWord (CBA'B'), and ShuffleWords (ACBD) (Table 1 and Figure S1).

TABLE 1 Words used in the experiment

Words (W)	Part words (PW)	Shuffle words (SW)
RaFiBouNeu	BouNeuNonLo/BouNeuVouDon	RaBouFiNeu
GuMaReuZo	ReuZoVouDon/ReZoNonLo	GuReuMaZo
NonLoSanBi	SanBiGuMa/SanBiRaFi	NonSanLoBi
VouDonMuLan	MuLanRaFi/MuLanGuMa	VouMuDonLan

Note: During each small test block, 12 test words were presented in isolation, four words, four ShuffleWords and four PartWords out of the eight possible.

Following a reviewer's requirement, we tested adults on the same material with the same exposure duration (~13mn). Their behavioral results are presented in Figure S5.

5.2.3 | Procedure

EEG was recorded with 128 electrodes (EGI geodesic sensor net), carefully placed on the neonates' heads by trained researchers to increase the consistency of the net placement. Three nets with different radii were used to fit infants' heads. For the continuous group, infants were tested while asleep in the experimenter or parent's arms. Due to COVID restrictions, the second group of babies was tested asleep in the crib. This slightly increased the noise level in the second group and might have marginally decreased the sensitivity of our analysis for this group.

Both groups followed the same procedure (Figure 1). A first control stream of a pseudorandom concatenation of 1600 syllables was followed by a structured stream composed of 3200 syllables grouped in words of four syllables. Infants then heard eight repetitions of short structured streams (160 syllables) followed by 12 test words presented in isolation (four in each condition: Word, PartWords and ShuffleWords, ISI 2-2.5s) for a total of 96 test-words (32 in each condition). The short streams were added to maintain learning because 2/3 of the test-words violated the learned structure. Each test word was preceded by a short click 200 ms before its onset. The click was added as a task unrelated auditory localizer and to reset the baseline with a neutral event to avoid long-range drifts following the words. Finally, a second control stream was presented. Thus, the two random-streams were sandwiching the structured stream to control for habituation to the auditory stimulation, change in sleep stage, and any confounding time effect.

5.2.4 | Data processing

EEG recordings were band-pass filtered between 0.2 and 15 Hz for all analyses. Artifact rejection was performed on the non-epoched recording session using APICE pipeline (Fló, et al., 2022) based on the EEGLAB toolbox (Delorme & Makeig, 2004). Artifacts were identified on continuous data, based on voltage amplitude, variance, first derivative, and running average. The variance algorithm was applied in sliding time windows of 500 ms with 100 ms steps. Adaptive thresholds were established for each subject and electrode as two interquartile ranges away from the 3rd quartile. This gave a logical matrix of the size of the recording, indicating bad data. Electrodes were definitely rejected if they were marked as bad more than 50% of the recording time, and time-samples were marked as bad if more than 35% of the electrodes were marked bad at this time-sample. For the ERP analysis, we then performed spatial interpolation of missing channels, and the data were mathematically referenced to the average of the 128 channels.

5.2.5 | Neural entrainment

The recordings from the structured and random streams were segmented into consecutive non-overlapping epochs of 15 words (corresponding to 15 s in the continuous group and 15.375 s in the pause group). All subjects having 10 good epochs or more in each condition were included in this analysis (25 neonates in the continuous group, 21 in the pause group). We averaged the activity over artifact-free epochs for each neonate and electrode and computed the Fourier Transform using the fast Fourier transform algorithm (FFT) as implemented in MATLAB. We then computed the power of the FFT. The PLV between trials was computed on the FFT of single trials. Those values were normalized with neighboring frequency bins [−8:1,1:8]. The frequencies of interest were selected as the inverse of the duration of a word (f = 1 Hz for the continuous group f = 0.975 for the second group with pauses) and one-quarter of a word (i.e., roughly a syllabic rate, f = 4 Hz for the first group, f = 3.9 Hz for the second). To assess the significance of the power/PLV at the two frequencies of interest, we computed a contrast between the power/PLV during the structured stream compared to the random streams for each electrode. As we expect learning during the structured stream to elicit a word rate oscillation, we computed a one-way (structured > random) paired t-test on each electrode. We corrected for multiple comparisons using a cluster corrected approach (alpha = 0.05). To look for a potential difference between groups, we computed an interaction between the previously described contrasts of both groups (difference of the structure minus random contrast in each group). Specifically, we ran a one-way unpaired t-test on each electrode and the clustering approach for the interaction.

Additionally, we also replicated the neural entrainment effects with a slightly different method as proposed in Fló, et al. (2022). With this approach, the signal is decomposed on 1s long epochs and reconstructed in longer meta-epochs composed of several non-necessarily consecutive segment. It allows to save more data for

shorter experiments at the expense of more data manipulations. Both approaches were quite similar, confirming the validity of both that can be better adapted depending on the amount of available data.

5.2.6 | Correlation analysis

In both experiments, all subjects heard the exact same auditory material avoiding differences in stimulation between participants. We could thus compute the instantaneous correlation between each participant and the others. For each subject at each time during the streams, we computed the correlation at the topographical level between the topography of subject i at time t (a vector of 128 voltage values at time t corresponding to the 128 electrodes) and the topography of the grand average excluding subject i at time t (a vector of 128 values corresponding to the average across the other subjects at time t for each of the 128 electrodes). It gave, for each subject, a vector of correlation between its own topography and the mean topography of all other subjects throughout time. Bad data were replaced by zeros and not taken into account for the average topographies across subjects. Time points with only bad data gave NaN correlation results. We hypothesized that learning should lead to an increase with time in the correlation between neonates as they learn the same material. To test it, we used two different methods. In the first one, we smoothed the correlation signal using a 400s-sliding-average-window in each neonate and stream, then computed a cluster-based analysis to reveal a significant cluster of time during which one stream showed a greater correlation than the other one. In the second one, we computed the slope of the linear regression with time in each subject and then considered the slope as a variable for the structured and random conditions in *t*-test comparing both groups.

5.2.7 | Pattern similarity analysis

To compute pattern similarity between syllables, we epoched each syllable from the structured stream from −100 ms to 350 ms. We removed the 100 first syllables to give enough time for participants to learn the task. The remaining epochs were averaged by syllables for each subject and a correlation matrix between each pair of syllables was computed with all the electrodes between 0 and 350 ms. We then separated the pairs of syllables into five conditions: First syllable (AA'), Ordinal position (BB' or CC' or DD'), Word and TP (AB or BC or CD), Word only (AC or AD or BD) and Low TP (DA'). We then averaged the similarity per condition and subtracted the correlation between all the other pairs. We then compared if pattern similarity between groups of syllables was increased differently across groups (One-way *t*-test with pauses > continuous).

5.2.8 | ERP analysis

Data were segmented in 2850 ms long epochs ([−750 +2100]ms relative to word onset), averaged in the three conditions (Words, PartWords, and ShuffleWords), and baseline-corrected with the mean voltage value in the interval [−750 to 0] in each neonate. Neonates with less than 20 remaining trials in total were excluded from analysis (none in the continuous group, 1 in the pause group).

To extract ROI corresponding to the functional auditory localizer of each group, we measured the auditory event-related potential associated with the click presentation at the beginning of each trial by running a cluster-based analysis against zero to extract auditory ERP (5000 randomizations, two-tailed *t*-test, alpha < 0.01, cluster-alpha < 0.01, between −200 and 0 ms). This procedure identified a positive frontal and a negative occipital cluster in each group, on which we restricted the ERP analyses. Therefore, the voltage was averaged across electrodes in each of the two clusters in each neonate and condition.

A cluster-based analysis was performed on the obtained time-series (10000 randomizations two-tailed *t*-test alpha < 0.05, cluster alpha < 0.05) between 250 ms (end of the first syllable) and 2000 ms to compare all pairs of conditions. Because of the adults' behavioral results, we added the contrast "heard" (average of Word and PartWord) vs. "non heard" (ShuffleWord) in the continuous group. Finally, we computed the interaction between groups and conditions (Word-PartWord) during the time window in which the previous analysis revealed a significant effect.

ACKNOWLEDGMENTS

We thank Simon Henin for help and comments about pattern similarity analysis. We also thank all the families who participated in our study as well as the maternity hospitals of Port-Royal and Orsay. This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 695710).

COMPETING INTERESTS

All other authors declare they have no competing interests.

DATA AND MATERIALS AVAILABILITY

All data and analysis are available upon request.

ETHIC APPROVAL

Ethical research committee of Paris Saclay University under the reference CER-Paris-Saclay-2019-063 & Ethical Committee CPP Tours Region Centre Ouest 1 (EudraCT/ID RCB: 2017- A00513-50)

ORCID

Lucas Benjamin https://orcid.org/0000-0002-9578-6039
Ana Fló https://orcid.org/0000-0002-3260-0559

REFERENCES

Andrillon, T., & Kouider, S. (2016). Implicit memory for words heard during sleep. *Neuroscience of Consciousness*, *2016*(1), niw014.

Andrillon, T., Poulsen, A. T., Hansen, L. K., Léger, D. L., & Kouider, S. (2016). Neural markers of responsiveness to the environment in human sleep. *Journal of Neuroscience*, *36*(24), 6583–6596.

Bagou, O., & Frauenfelder, U. H. (2018). Lexical segmentation in artificial word learning: The effects of converging sublexical cues. *Language and Speech*, *61*(1), 3–30.

Batterink, L. J., & Choi, D. (2021). Optimizing steady-state responses to index statistical learning: Response to Benjamin and colleagues. *Cortex*, *142*, 379–388.

Batterink, L. J., & Zhang, S. (2022). Simple statistical regularities presented during sleep are detected but not retained. *Neuropsychologia*, *164*, 108106. https://doi.org/10.1016/j.neuropsychologia.2021.108106

Benjamin, L., Dehaene-Lambertz, G., & Fló, A. (2021). Remarks on the analysis of steady-state responses: Spurious artifacts introduced by overlapping epochs. *Cortex*, *142*(2), 370–378.

Black, A., & Bergmann, C. (2017). Quantifying infants' statistical word segmentation: A meta-analysis. *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, (3), 124–129.

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.

Choi, D., Batterink, L. J., Black, A. K., Paller, K. A., & Werker, J. F. (2020). Preverbal infants discover statistical word patterns at similar rates as adults: Evidence from neural entrainment. *Psychological Science*, *31*(9), 1161–1173.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America*, *95*(3), 1570–1580.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114.

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, *298*(5600), 2013–2015.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behav Res*, *47*, 1–12. https://doi.org/10.3758/s13428-014-0458-y

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vrecken, O. (1996). The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, *3*, 1393–1396. https://doi.org/10.1109/ICSLP.1996.607874

Ellis, C. T., Skalaban, L. J., Yates, T. S., Bejjanki, V. R., Córdova, N. I., & Turk-Browne, N. B.. Evidence of hippocampal learning in human infants. *Current Biology*, *31*(15), 3358–3364.e4.

Farthouat, J., Atas, A., Wens, V., De Tiege, X., & Peigneux, P. (2018). Lack of frequency-tagged magnetic responses suggests statistical regularities remain undetected during NREM sleep. *Science Reports*, *8*(1), 11719.

Fernandes, T., Kolinsky, R., & Ventura, P. (2010). The impact of attention load on the use of statistical information and coarticulation as speech segmentation cues. *Atten Percept Psychophys*, *72*(6), 1522–1532.

Fernandes, T., Ventura, P., & Kolinsky, R. (2007). Statistical information and coarticulation as cues to word boundaries: A matter of signal quality. *Perception and Psychophysics*, *69*(6), 856–864.

Ferry, A. L., Fló, A., Brusini, P., Cattarossi, L., Macagno, F., Nespor, M., & Mehler, J. (2016). On the edge of language acquisition: Inherent constraints on encoding multisyllabic sequences in the neonate brain. *Developmental Science*, *19*(3), 488–503.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15822–15826.

Fló, A., Benjamin, L., Palu, M., & Dehaene-Lambertz, G. (2022). Sleeping neonates track transitional probabilities in speech but only retain the first syllable of words. *Scientific Reports*, *12*(1), 4391. https://doi.org/10.1038/s41598-022-08411-w

Fló, A., Brusini, P., Macagno, F., Nespor, M., Mehler, J., & Ferry, A. L. (2019). Newborns are sensitive to multiple cues for word segmentation in continuous speech. *Developmental Science*, *22*(4), e12802.

Fló, A., Gennari, G., Benjamin, L., & Dehaene-Lambertz, G. (2022). Automated Pipeline for Infants Continuous EEG (APICE): A flexible pipeline for developmental cognitive studies. *Developmental Cognitive Neuroscience*, *54*, 101077. https://doi.org/10.1016/j.dcn.2022.101077

Green, C. (2017). Usage-based linguistics and the magic number four. *Cognitive Linguistics*, *28*(2), 209–237.

Hassin, R. R., Bargh, J. A., Engell, A. D., & McCulloch, K. C. (2009). Implicit working memory. *Consciousness and Cognition*, *18*(3), 665–678.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*(3), 53–64.

Henin, S., Turk-Browne, N. B., Friedman, D., Liu, A., Dugan, P., Flinker, A., Doyle, W., Devinsky, O., & Melloni, L. (2021). Learning hierarchical sequence representations across human cortex and hippocampus. *Science Advances*, *7*(8), 1–13.

Hochmann, J. R., Endress, A. D., & Mehler, J. (2010). Word frequency as a cue for identifying function words in infancy. *Cognition*, *115*(3), 444–457.

James, L. S., Sun, H., Wada, K., & Sakata, J. T. (2020). Statistical learning for vocal sequence acquisition in a songbird. *Scientific Reports*, *10*(1), 1–18.

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*(2), 339–345.

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 and Language*, *148*, 25–36.

Mandel, D. R., Jusczyk, P. W., & Kemler Nelson, D. G. (1994). Does sentential prosody help infants organize and remember speech information?, *Cognition*, *53*(2), 155–180.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology General*, *134*(4), 477–500.

Nespor, M., & Vogel, I. (2006). Prosodic phonology.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*(2011), Article ID 156869, https://doi.org/10.1155/2011/156869

Ordin, M., Polyanskaya, L., Laka, I., & Nespor, M. (2017). Cross-linguistic differences in the use of durational cues for the segmentation of a novel language. *Memory and Cognition*, *45*(5), 863–876.

Palmer, S. D., & Mattys, S. L. (2016). Speech segmentation by statistical learning is supported by domain-general processes within working memory. *Quarterly Journal of Experimental Psychology (Hove)*, *69*(12), 2390–2401.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*(5593), 604–607.

Pothos, E. M., & Juola, P. (2007). Characterizing linguistic structure with mutual information. *British Journal of Psychology*, *98*(2), 291–304.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606–621.

Scher, M. S. (2008). Ontogeny of EEG-sleep from neonatal through infancy periods. *Sleep Medicine*, *9*, 615–636.
Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition*, *106*(2), 975–983.
Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, *54*(1), 1–32.
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. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(15), 6038–6043.
Sigurd, B., Eeg-Olofsson, M., & van de Weijer, J. (2004). Word length, sentence length and frequency – Zipf revisited, *Studia Linguistica*, *58*(1), 37–52.
Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*(1), 86–132.
Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*, 21.
Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), 25–34.
Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception and Psychophysics*, *67*(5), 867–875.
Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *The Journal of the Acoustical Society of America*, *126*(1), 367–376.
Wakefield, J. A., Doughtie, E. B., & Lee Yom, B. H. (1974). The identification of structural components of an unknown language. *Journal of Psycholinguistic Research*, *3*(3), 261–269.
Wang, F. H., Zevin, J. D., Trueswell, J. C., & Mintz, T. H. (2020). Top-down grouping affects adjacent dependency learning. *Psychonomic Bulletin & Review*, *27*(5), 1052–1058.
Zipf, G. K. (reprinted 1965). (1935). *The psycho-biology of language*. MIT Press.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Benjamin, L., Fló, A., Palu, M., Naik, S., Melloni, L., & Dehane-Lambertz, G. (2022). Tracking transitional probabilities and segmenting auditory sequences are dissociable processes in adults and neonates. *Developmental Science*, e13300. https://doi.org/10.1111/desc.13300