

# Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants



C. Kabdebon<sup>a,b,c,\*</sup>, M. Pena<sup>d</sup>, M. Buiatti<sup>a,b,c</sup>, G. Dehaene-Lambertz<sup>a,b,c</sup>

<sup>a</sup> INSERM, U992, Cognitive Neuroimaging Unit, F-91191 Gif/Yvette, France

<sup>b</sup> CEA, DSV/I2BM, NeuroSpin Center, F-91191 Gif/Yvette, France

<sup>c</sup> University Paris 11, F-91191 Orsay, France

<sup>d</sup> Escuela de Psicología, Pontificia Universidad Católica de Chile, Chile

## ARTICLE INFO

### Article history:

Accepted 9 March 2015

Available online 9 April 2015

### Keywords:

Language  
Infant  
Statistics  
Learning  
Event-related potentials  
Frequency tagging  
Phase-locking  
Prematurity

## ABSTRACT

Using electroencephalography, we examined 8-month-old infants' ability to discover a systematic dependency between the first and third syllables of successive words, concatenated into a monotonous speech stream, and to subsequently generalize this regularity to new items presented in isolation. Full-term and preterm infants, while exposed to the stream, displayed a significant entrainment (phase-locking) to the syllabic and word frequencies, demonstrating that they were sensitive to the word unit. The acquisition of the systematic dependency defining words was confirmed by the significantly different neural responses to *rule-words* and *part-words* subsequently presented during the test phase. Finally, we observed a correlation between syllabic entrainment during learning and the difference in phase coherence between the test conditions (*rule-words* vs *part-words*) suggesting that temporal processing of the syllable unit might be crucial in linguistic learning. No group difference was observed suggesting that non-adjacent statistical computations are already robust at 8 months, even in preterm infants, and thus develop during the first year of life, earlier than expected from behavioral studies.

© 2015 Elsevier Inc. All rights reserved.

## 1. Introduction

Statistical learning is one of the most successful approaches to provide an efficient account of how infants learn the particular characteristics of their native language. From the distributional pattern of sounds in their environment that shapes their discrimination abilities (Maye, Weiss, & Aslin, 2008; Maye, Werker, & Gerken, 2002) to the co-occurrence of specific acoustic properties (low intensity, short duration, weak stress, particular distribution of phonemes) and reproducible positions in prosodic domains that help them to infer word category (open vs closed-class words) (Hochmann, Endress, & Mehler, 2010; Shi, Cutler, Werker, & Cruickshank, 2006), infants display particularly efficient capacities for analyzing the statistical properties of events in the speech signal from an early age on.

Infants are also rapidly able to keep track of the statistical dependency between two events. As transitional probabilities (TP) between adjacent syllables tend to be higher within words

than between words, it has been hypothesized that infants could segment words from fluent speech using this difference between word-internal and word-external TP. Several studies confirmed that infants can use this strategy (Goodsitt, Morgan, & Kuhl, 1993; Johnson & Tyler, 2010; Saffran, Aslin, & Newport, 1996). Tracking adjacent probabilities backwards was similarly proposed to infer a word category given the preceding word (e.g. noun vs verb depending on the preceding determiner vs personal pronoun), a capacity reported in 12-month-olds (Mintz, 2003) and 14-month-olds (Höhle, Weissenborn, Kiefer, Schulz, & Schmitz, 2004; van Heugten & Shi, 2010).

However, linguistic structures are not limited to adjacent syllables. The syntactic structure of a language in particular relies on building dependencies between distant parts of a sentence, as for example in subject-verb agreement. However, tracking long-distance statistics requires on the one hand more complex computations to bypass adjacent dependencies and on the other hand larger memory buffers to discover them. Yet, once long-distance dependencies are acquired, the coding of long utterances is simplified and the memory load reduced. The advantage of using this strategy to discover the higher levels of the syntactic hierarchy is obvious, and we may wonder at what age infants can rely on this capacity to process speech. Thereafter, we will refer to this

\* Corresponding author at: Laboratoire de Neuroimagerie Cognitive, INSERM, U992, CEA/SAC/DSV/DRM/NeuroSpin, Bat 145, point courrier 156, F-91191 Gif/Yvette, France. Fax: +33 1 69 08 79 73.

E-mail address: [claire.kabdebon@gmail.com](mailto:claire.kabdebon@gmail.com) (C. Kabdebon).

long-distance statistical dependency as AxC, i.e. syllable A predicts syllable C, independently of the intervening and variable x. AxC computations were first reported in 18-month-old infants who accepted sentences of the type “is running” but not of the type “can running”, whereas 15 month-olds were not able to differentiate the two types of sentences (Santelmann & Jusczyk, 1998). Using simpler artificial languages but brief exposure of a few minutes, Gómez and Maye (2005) reported success in non-adjacent statistical learning at the younger age of 15 months but failure at 12, and in a very recent experiment, Marchetto and Bonatti (2014) reported an even younger age for the onset of this capacity. They showed that 12- and 7-month-olds were able to discover the systematic dependency between the first and third syllable of trisyllabic words (AxC words) embedded in an artificial stream of syllables. However, only the 12 month-olds were able to generalize the structure (A predicts C) to new words comprising a novel syllable between A and C.

A few EEG studies have brought additional evidence for an early ability to track long-distance dependencies. For example, in French, long-distance computations are crucial to notice that the sentence “Je la donne” (I give it) is grammatical, while “Je la fraise” (I strawberry it) is not. In both cases, the local transitions are correct because the function word “la” can be either a determiner or a clitic (thus “je la”, “la donne” and “la fraise” are correct pairs of words). Yet, 24-month- French infants reacted to the ungrammatical sentences embedded in short stories by a positive electrical component over the parietal areas revealing that they were able to take into account long-distance dependencies during on-line processing of natural speech (Bernal, Dehaene-Lambertz, Millotte, & Christophe, 2010). At a much younger age, 4-months, Friederici, Mueller, and Oberecker (2011) also observed a similar positive component over centro-parietal regions in response to non-grammatical sequences. German infants who were exposed to 18 min of naturally produced Italian sentences (*Il fratello/La sorella sta x-ando or puo x-are*), progressively noticed the verb-inflexion exchange (*sta x-are or puo x-ando*) presented during short test periods regularly spaced during exposure. Prosodic cues might have here played a crucial role to help chunk the stream in smaller units in order to memorize the two edges of the second chunk (*sta x-are or puo x-ando*). This result at this age is remarkable given that the intervening x element was chosen among 32 possibilities, certainly overwhelming infant’s capacity to memorize each AxC triplet. The verb-inflexion exchange was thus only noticeable if infants had kept track of the exact relation between A and C. In a final study (Mueller, Friederici, & Mannel, 2012), 3-month-old infants were exposed to isolated trisyllabic non-words. The words belonged to two AxC families, the intervening x syllable being drawn from a set of 20 syllables. Two types of deviants were randomly interspersed: the last syllable of the deviant words was either exchanged between the two families (rule violation) or presented a change of pitch (control violation). The group of infants as a whole did not show a significant mismatch response (MMR) to the exchanged syllable (rule violation). However, once the infants were sorted as a function of gender and of polarity of their MMR to pitch, a complex interaction between gender and polarity of the mismatch response was reported, indicating that male and female infants with a positive MMR to pitch showed no MMR to rule violation, while within the negative MMR pitch group, males showed a positive MMR to rule violation and females showed a negative one. This interaction was interpreted by the authors as resulting from a maturational advantage of girls over boys in rule learning.

To sum up this rapid review, contrary to the computation of transition probabilities between adjacent syllables which is robust from, at least, the second semester of life, the age of success for long-distance computations appears to be highly variable across

studies. There is converging evidence between behavioral and EEG studies that toddlers are able to compute non-adjacent probabilities after 18 months, but the results in younger infants are less stable. However, whereas infants’ capacities to compute adjacent transitional probabilities have been tested with close paradigms across ages, the complexity of the learning utterances widely differs across the studies presented above because of the size (one or two syllables) and variability of the intervening non-pertinent element (i.e. the pool of x), the number of A\_C families to learn and thus the relative weight of the adjacent (Ax and xC) and non-adjacent (A\_C) dependencies, the length of the sentence to analyze, the length of the training, etc. Thus our goal was here to confirm whether infants in the first year of life were able to compute long-distance dependencies, or whether there is a sizeable age gap between adjacent and non-adjacent statistical computations which would suggest that they may depend on different neural bases.

We tested healthy 8-month-old infants because, at this age, infants listening to long utterances display unquestionable sensitivity to its segmental properties and to the transitional probabilities between adjacent syllables (Saffran et al., 1996). Our participants belonged to a group of full-terms or to one of two possible groups of preterm infants, one evaluated at 8 months after birth and the other at 8 months after the age they should have been born (term age), to cover a wide range of maturation and exposure to the ex-utero environment and examine the influence of these factors on the emergence of this capacity. This study is part of a program of experiments designed to investigate the relative influence of neural maturation and exposure to a linguistic environment on the development of linguistic abilities. Some abilities may develop following a maturational calendar whatever the environment, while other abilities may depend on the duration of exposure to broadcast speech and social interaction. In that case, preterm infants should benefit from their longer ex-utero experience. Eventually, some abilities might be systematically delayed in preterm infants due to a non-optimal early environment, even when no brain lesion is observed. We already observed that neural maturation is crucial in the loss of electrophysiological responses to non-native linguistic features (Pena, Pittaluga, & Mehler, 2010; Pena, Werker, & Dehaene-Lambertz, 2012), whereas duration of ex-utero exposure is the key factor for the learning of the phonotactic rules of the native language (Gonzalez-Gomez & Nazzi, 2012), and for gaze-following (Pena, Arias, & Dehaene-Lambertz, 2014). Note, however, that if the capacity develops earlier than the tested age, results will be similar in all three groups.

To be as close as possible to the experimental paradigms used to test adjacent statistical computations (Saffran et al., 1996), we exposed infants to 2 min of an artificial monotonous stream of concatenated syllables in which trisyllabic words were embedded, separated by a 25 ms subliminal pause. The words belonged to three families, each one characterized by an A\_C dependency, with the intermediate syllable x coming from a pool of 3 syllables (Table 1). In the stream, each pair of adjacent syllables (Ax, xC, CA’, etc.) had a similar range of low transition probability, whereas transition probabilities between non-adjacent syllables structured the continuous stream into trisyllabic words. Trisyllabic items were subsequently presented in isolation during a test part, and were consistent, or not, with the structure of the stream.

This paradigm was successfully tested in adults (Pena, Bonatti, Nespor, & Mehler, 2002). Adults exposed to such AxC stream and subsequently presented with isolated trisyllabic items identified as words the triplets consistent with the A\_C dependency, although they never appeared in the stream (thereafter *rule-words*), but not the xca’ triplet (*part-words*) that did appear in the

**Table 1**

Material used to generate the artificial speech stream and the test items.

Learning speech stream	Test items isolated words	
'Words'	'Rule-words'	'Part-words'
kulEbi	kunabi	fibina
kufibi	kufebi	Lebina
kugobi	kudubi	gobina
nafidu	nalidu	Leduli
naLedu	nafedu	goduli
nagodu	nakudu	fiduli
lifife	linafe	gofeku
liLefe	libife	fifeku
ligofe	likufe	Lefeku

stream but with low statistical dependency between their syllables. However, this result was observed only when 25 ms subliminal pauses were inserted between each word in the stream. The authors hypothesized that long-distance computations can only be performed on small chunks of signal and that the subliminal pause was implicitly used by the participants as a prosodic cue to signal word ending. Using electrophysiology, we studied the electrophysiological correlates of this task in adults (Buiatti, Pena, & Dehaene-Lambertz, 2009). We exploited the temporal precision of the brain activity, and its property to respond to a rhythmic stimulus by steadily resonating at the same frequency (also denoted entrainment). These rhythmic responses can thereafter be detected as sharp peaks in the power spectrum of the cerebral activity. As the learning stream was artificially synthesized, all syllables had the exact same duration, and were presented at a steady rate. Not surprisingly, we reported a significant power increase in the brain oscillatory activity at the syllabic frequency. More interestingly, the brain activity also resonated at the word frequency, revealing that subjects were extracting the embedded AxC structure of the stream. Crucially, this was not the case for control streams containing randomly distributed sequences of syllables, even when a 25 ms pause was inserted every three syllables, confirming that this cue was not sufficient by itself to elicit a power increase at its frequency. The discovery of the long-distance dependency was crucial, as further demonstrated by the positive correlation between the power increase at the word frequency and the explicit recall of the detected words.

In the present study in infants, we adopted a similar approach, allowing us to investigate the neural mechanisms at play during learning, and to link them to the subsequent performances in test. First, we asked whether we could track on-line neural correlates of non-adjacent statistical learning in 8 month-old infants, and searched for an increased entrained activity at the word frequency. In a methodological breakthrough with respect to the previous study in adults (Buiatti et al., 2009), we quantified the frequency-tagged entrainment by measuring the phase-locking of the cerebral activity to the AxC stream at the syllabic and word frequencies, rather than the corresponding peaks of the power spectrum, for two reasons: (1) In a frequency-tagging design, phase-locking proved to be much more robust to the background low frequency fluctuations than power spectrum peak estimation (Forget, Buiatti, & Dehaene, 2009). Indeed, power peak estimation is typically a normalization of the power at a given frequency by the average power of the surrounding frequencies. The syllable (around 4 Hz) and word (around 1 Hz) frequencies being rather low, the normalization term captures the spontaneous background  $1/f$  noise (steeper in infants than in adults). (2) A growing body of research shows that phase-locking to speech rhythm (rather than power spectrum variations) reliably tracks speech comprehension (Ahissar et al., 2001; Kerlin, Shahin, & Miller, 2010; Luo & Poeppel, 2007; Peelle, Gross, & Davis, 2013).

Secondly, we assessed infants' ability to recognize the A\_C pattern in isolated trisyllabic items, and we compared the neural responses elicited by *rule-words* vs. *part-words*. We completed the classical ERPs by analyses of the event-related spectral perturbations up to 35 Hz, as speech processing elicits oscillatory activity beyond the one related to the speech envelope. Modulation in different frequencies-bands have been related to general processes such as working memory load (Palva, Monto, Kulashekhar, & Palva, 2010) and attentional processes (Ward, 2003) or to more specific speech processes, such as top-down processes in the beta-band (Fontolan, Morillon, Liegeois-Chauvel, & Giraud, 2014) and long-distance relation computation when adults listened to sentences, in the alpha-band (Meyer, Obleser, & Friederici, 2013). Finally, we related the neural responses recorded during the learning and the test phases to identify crucial factors for a successful learning in infants.

## 2. Materials and methods

### 2.1. Participants

Three groups of infants were included in this study: a group of 8-month-old full-term infants (FT8) and two groups of healthy preterm infants, one evaluated at 8 months after term (PT11) and the other at 8 months after birth (PT8) but thus at a younger neural age (i.e. 5 months post-term), all from a monolingual Spanish-speaking environment, for a total of 73 infants. Twenty-four infants were excluded because they did not complete the experimental protocol. Of the remaining 49 infants, some were excluded from the analyses of either the learning part (3 from FT8, and 1 from PT8 for a total of 4 infants) or the test part (3 from FT8, and 1 from PT11 for a total of 4 infants) of the experiment because of a too small number of trials after artifact detection in the EEG recordings (see data analysis section). We thus report on 45 infants for the learning part (15 FT8 (8 males), 11 PT11 (7 males), and 19 PT8 (10 males)), and 45 infants for the test part (15 FT8 (10 males), 12 PT11 (8 males) and 18 PT8 (10 males)). For the analyses concerning both the learning and test parts, 42 infants were included.

Preterm infants were born between 27 and 30.5 weeks gestational age (wGA) (mean =  $29.5 \pm 1.8$  wGA) and full-terms between 38 and 41 wGA (mean =  $39.5 \pm 0.8$  wGA). At birth, all infants had: (1) Apgar scores higher than 6 and 8 respectively at 1 and 5 min; (2) normal weight, size, and cranial perimeter for their gestational age; (3) normal otoacoustic emissions; and (4) neuropediatric scores corresponding to their gestational age. Auditory brainstem-evoked responses and brain ultrasonography were normal for gestational age in preterm infants. All infants came from lower-middle socioeconomic class families and presented normal clinical outcomes over 6 years. Ethical approval was obtained from the S6tero del R6o Hospital ethics committee and informed written consent from the parents.

Preterm infants received speech stimulation from birth on. In hospital, they were exposed to speech stimulation outside the incubator during a Kangaroo procedure from birth on. This medical protocol used in Neonatal Intensive Care Units involves skin-to-skin contact between the mother and the newborn (Feldman, Eidelman, Sirota, & Weller, 2002). Rapidly after birth, healthy preterm infants are placed in open cradles where they are largely exposed to broadcast speech from their mothers but also from other people talking to them or near them. Based on international recommendations (Pediatrics, 1998), preterm infants were discharged at home around 34 wGA where they received similar exposure to speech than full-terms. Thus, at the time of testing, the mean duration of exposure to ex-utero life for FT8, PT8, and PT11 was  $34.6 \pm 0.7$ ,  $34.9 \pm 0.6$  and  $46.5 \pm 0.7$  weeks, respectively.

## 2.2. Stimuli

The learning part consisted of 129-seconds streams of continuous artificial monotonous speech, constructed by the pseudorandom concatenation of 9 nonsense trisyllabic items (thereafter words, see Table 1). Nine CV syllables (9 consonants and 5 vowels) were used to construct three families of words. Each family included 3 words, containing the same first and third syllable (Table 1). For instance, the family /ku\_bi/ (where \_ indicates the occurrence of the second syllable) consisted of the words /kuLebi/, /kufibi/ and /kugobi/. The words were concatenated with the restriction that either the same word, or two words of the same family or two words containing the same middle syllables could not be presented in succession. Thus infants could not infer the next three syllables from the three preceding syllables. Across all the stream, the average TP between adjacent syllables was 0.33 (0.25–0.42) within words and 0.5 (0.43–0.57) between words, resulting in two close word-internal and word-external TP distributions. By contrast, non-adjacent TP (with one intervening syllable) were very different within words and between words, 1 and 0.33 respectively. A 25-ms-long pause was included each three syllables (i.e. between words) as it has been shown that segmentation cues favor long-distance computations and generalization of the regularity to new items both in adults (Buiatti et al., 2009; Pena et al., 2002) and infants (Marchetto & Bonatti, 2014). Contrary to the 200 ms pause used by Marchetto and Bonatti (2014), this pause was subtle and not consciously perceived by adults (Pena et al., 2002).

If like the adults in Pena et al. (2002), infants are able to discover the word structure, they might generalize to new items. Thus in a test phase, we presented 72 isolated trisyllabic items (Table 1), 9 rule-words (repeated 4 times for a total of 36 trials) containing the first and third syllables of an A\_C family but combined with an intervening syllable, present in the learning stream but that never occurred in second position (i.e. the first or last syllable of another Ax\_C family, for example *kunabi*), and 9 part-words (repeated 4 times for a total of 36 trials) created by joining the last two syllables of a word to the first syllable of another word (i.e. coming from another family, for example *Lebina*).

All speech stimuli were generated with the MBROLA text-to-speech software (Dutoit, 1997), using French diphones (because French diphones in the MBROLA database are of better quality than many other romance language), digitalized in 16-bit mono files at 22050 Hz with no prosodic contour but with a 25 ms silence between each word to facilitate segmentation (Pena et al., 2002). All syllables only used common phonemes between French and Spanish, two close Romance languages. They were easily identified by MP, a native Chilean-Spanish speaker. The duration of all phonemes was equalized to 116 ms (syllable duration = 232 ms and word duration = 696 ms, i.e.  $6 * 116$  ms) the pitch at 200 Hz with no variation of intensity within and between words. For the continuous stream, several word orders were used to randomize the sequence of words across subjects.

## 2.3. Experimental procedure

Infants were tested in a soundproof Faraday booth. The infant sat on the parent's lap and the parent listened to music through earphones to mask the speech stimuli during testing. To avoid body movement, infants saw attention-grabbing images displayed in a monitor. Infants heard the 190-s long continuous stream, then after a 30-s break, they started the 72 test trials, each one consisting of an isolated rule-word or part-word, randomly selected and separated from the next item by 8 s of silence. EEG data were collected using a 64-electrode geodesic sensor net (EGI, Eugene, USA, amplifiers N200) referenced to the vertex with a sampling rate of

500 Hz. Maximal impedance was 40 k $\Omega$ . No behavioral measurement was carried out.

## 2.4. Data processing and statistical analyses

EEG recordings were band-pass filtered (0.2–40 Hz) for the frequency analyses of the learning and test parts, and (0.2–10 Hz) for the ERPs analyses in the test part, then segmented into epochs (see specific parts for their duration). Artifact rejection was performed using custom-made MATLAB scripts based on the EEGLAB toolbox (Delorme & Makeig, 2004). Specifically, for each channel epochs were rejected either when the fast average amplitude exceeded 250  $\mu$ V, or when deviation between fast and slow running averages exceeded 150  $\mu$ V. Electrodes were rejected if more than 50% of the epochs were marked bad, and trials were rejected if more than 70% of the epochs were marked bad. Specifically, in the learning part, on average 3 epochs, and 14 electrodes were rejected per subject, while in the test part, on average 2 epochs, and 4 electrodes were rejected per subject. Electrodes that were mostly rejected were those located on the peripheral ring over the neck and forehead. Non rejected epochs were mathematically referenced to the average of the 64 channels.

### 2.4.1. Learning part

**2.4.1.1. Frequency tagging.** In order to ensure a high frequency resolution, with one frequency bin centered on syllabic, bisyllabic and trisyllabic frequencies, we segmented the EEG recordings from the learning part into consecutive non-overlapping epochs of exactly 12 trisyllabic words (8.652 s). For each electrode, we computed the Fourier Transform of the activity averaged over artifact-free epoch, using the fast Fourier transform algorithm as implemented in MATLAB

$$FT(f) = \sum_{n=0}^{N-1} c_n \cdot e^{2\pi i n f / N}$$

with  $N$  the length of the epoch, in samples.

From the Fourier transform, a phase locking value (PLV) was computed at each frequency of interest and for each subject over the whole learning phase (Tallon-Baudry, Bertrand, Delpuech, & Pernier, 1996). The frequencies of interest were selected as the inverse of the duration of a word ( $f = 1.39$  Hz) and one-third of a word (i.e. roughly a syllabic rate,  $f = 4.17$  Hz). As a control, we also computed the frequency corresponding to two-thirds of a word (i.e. roughly a bi-syllabic rate,  $f = 2.08$  Hz) for which we expected no entrainment, and thus a low phase locking value.

$$PLV(f) = \frac{1}{M} \sum_{m=1}^M \frac{FT_m(f)}{|FT_m(f)|}$$

with  $M$  the number of epochs. The PLV ranges from 0 (purely non-phase-locked activity) to 1 (strictly phase-locked activity).

To evaluate the statistical significance of the PLV, we generated surrogate data to assess the PLV distribution under the null hypothesis of non-entrained oscillatory activity. Only values departing significantly from this null distribution can then be considered as revealing the presence of phase locking. The null hypothesis being “no cerebral entrainment by the auditory stimulation”, we built a surrogate PLV by breaking up this entrainment: epochs from the continuous EEG signal of the learning part were extracted with random time onsets and lasting the same duration as for the previous data analysis; PLVs were then computed as on the real data, resulting in a surrogate PLV for each electrode and each frequency of interest. 10 surrogate PLVs were generated using



this procedure, and averaged to build a stable estimate of the surrogate PLV distribution.

The statistically significant differences between surrogate and real PLVs were then assessed at each frequency of interest (syllabic:  $f = 4.17$  Hz, word:  $f = 1.39$  Hz, and the control bi-syllabic frequency  $f = 2.08$  Hz) using customized cluster analysis coupled with a randomization procedure (Maris & Oostenveld, 2007). The two conditions (real PLV vs. surrogate PLV), were first compared separately using a two-tailed paired  $t$ -test for each electrode. Levels of probability obtained were then thresholded at  $p = 0.05$ , in order to define spatial clusters. Our procedure pooled above-threshold samples corresponding to contiguous electrodes, separately for positive and negative  $t$ -values. Each positive and negative cluster was then attributed a weight equal to the sum of the  $t$ -values of all its constitutive electrodes. To evaluate the significance of the effects, we then recomputed the same analysis on 5000 sets of randomly permuted data, for which no significant effect was expected. A permutation was defined by randomly attributing the label “surrogate” or “real” to the two sets of PLV for each subject. For each permutation, we extracted distribution of maximal effect strengths. The corrected  $p$ -values corresponded to the rank of the experimental data within the distribution of maximal effect strength, divided by the total number of permutations. This method allowed selecting the clusters of electrodes that were significantly affected by our experimental conditions.

When a significant effect was observed in the previous analyses, we investigated whether all three groups were similarly reacting through an ANOVA with group as between-subject factor. The studied variable was the phase-locking value in the frequency of interest averaged in each subject across all significant electrodes of the pertinent previous analysis.

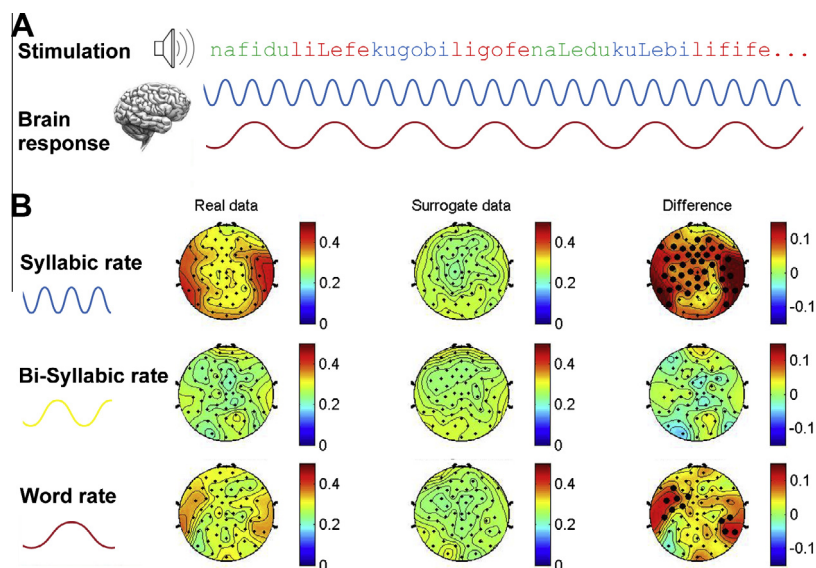
**2.4.1.2. Time course of the entrainment.** The sensitivity of PLV measurements allowed us to identify the cluster of electrodes specifically phase-locked to the stimulation at syllabic and word rhythms; however, it does not allow us to track the time-course of the phase locking. Indeed, phase locking is not a dynamic measure as it integrates over several epochs (see section *Frequency*

*tagging analysis*). Therefore, due to the relatively short stimulation period (180 trisyllabic words, that is 15 epochs), and poor SNR, even splitting the data into two time periods (two datasets of 60 words, that is 8 epochs) would have not provided enough epochs to compute a reliable PLV. We thus used the power of the oscillatory activities to examine a putative effect of learning. In each infant, we extracted the power of the oscillatory activity at the frequency of interest in epochs of 12 words, overlapping for 11/12 of their length. We thereafter averaged these values across the electrodes showing a significant steady state response in the above analysis. These values averaged across participants were fed into a linear model with epochs as a regressor. We thereafter tested the null hypothesis that the coefficient of the fitted line was equal to zero (no effect).

#### 2.4.2. Test part

**2.4.2.1. ERP analysis.** Because of the weak myelination of long fiber tracks at this age and probably also because of less efficient computations, infants' ERP are generally delayed compared to adults and late responses around the second and later are common in infants. These late waves have been generally related to attentional orientation and compared to the P300 in adults (Kouider et al., 2013; Nelson & deRegnier, 1992). We therefore chose to inspect neural responses up to more than one second after the word offset (word duration: 696 ms). Epochs were 2000 ms long (−200 + 1800 ms after word onset). They were baseline-corrected (from 200 ms before the onset of the trial), and averaged in the two conditions (*rule-word* and *part-word*). Subjects with less than 10 remaining trials in one of the two conditions after artifact rejection were rejected (see Section 2.2). On average, we report on 26 trials per subject in the *rule-word* condition, and 24 trials per subject in the *part-word* condition. More specifically, in the 8FT group, the analyses are based on 29 trials in the *rule-word* condition and 27 trials in the *part-word* condition, on average. In the PT11 group we report on 24 *rule-word* trials and 21 *part-word* trials, and in the PT8 group we have, 26 *rule-word* trials and 23 *part-word* trials, on average.

To identify the latencies and topographies of the auditory ERPs in our subjects for this type of stimuli, we first averaged



**Fig. 1.** Frequency tagging during the learning stream. A: Schematic representation of the expected brain activity in response to the stimulation, at syllabic and word frequencies. The oscillatory activity is phase locked to the onset of syllables (blue) and words (red). B: Phase-locking values at syllabic, bi-syllabic and word rates in the real data (first column), and surrogate data (second column). The difference between the two is presented as a third column. The electrodes showing a significant difference are highlighted.

the response of all participants, regardless of the group, across both *rule-* and *part-words* and identified the different ERP components by visual inspection. Because voltages are reference-average, the dipole configuration of the auditory response was easily seen (Fig. 2) and clusters of electrodes were defined at the negative and positive maxima of the response. Voltages were then averaged across these clusters of electrodes and we visually selected time-window to enter ANOVAs. A first time-window ( $t_1$ ) encompassed the responses to the last two syllables of the word and a second ( $t_2$ ) encompassed the late slow wave (Figs. 2 and 3). We declared Condition (Rule-word & Part-word) and Polarity (Positive & Negative), as within-subject factors and Group (FT8, PT11 & PT8) as a between-subjects factor. We added a factor Hemisphere (Right & Left) for the first analysis because of the wide and bilateral topography of the auditory response.

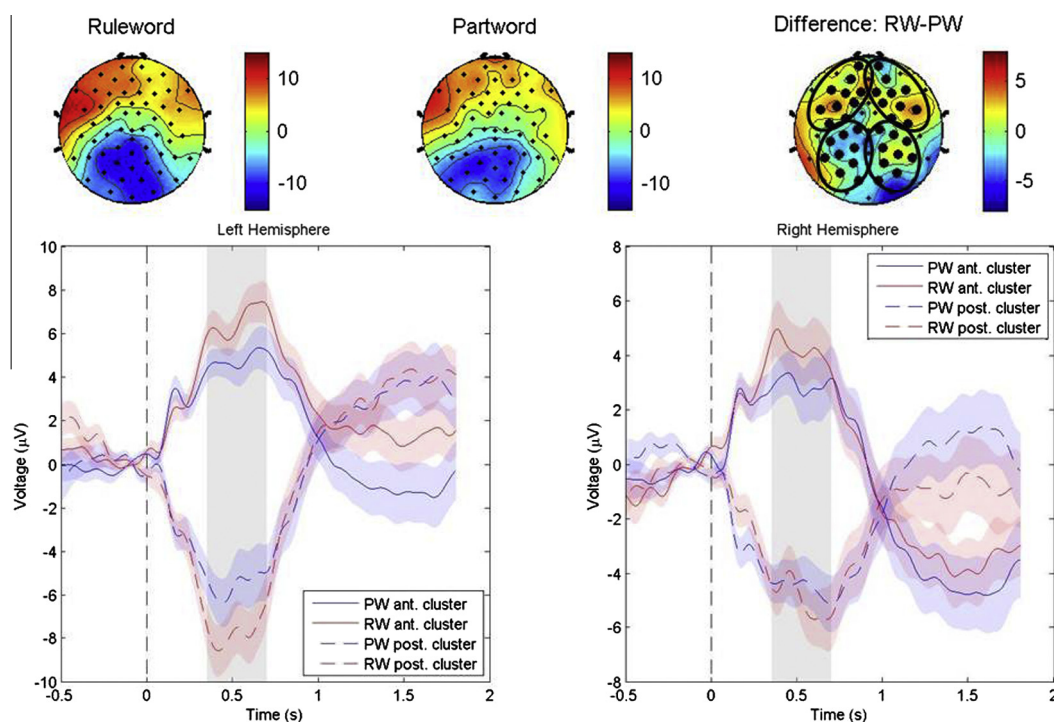
**2.4.2.2. Phase-locking value analysis.** For the frequency analysis of the test part, as there is a tradeoff between frequency resolution and epoch length, we segmented the dataset into epochs including twice the duration of a word preceding the test item onset, and three times the duration of a word after test item onset, resulting in 3605 ms-long epochs. They were baseline corrected to the 200 ms before the onset of the trial.

We capitalized on the perfect temporal regularity of the stimuli to examine differences in phase-locking value between *rule-words* and *part-words*, at the precise frequencies of the syllables ( $f=4.17$  Hz) and of the words ( $f=1.39$  Hz). We wondered whether attention to the precise structure of the items would induce resetting of the phase at the syllable frequency, or even at the word frequency. PLV in the test part was computed as described above for the learning part with the only difference that epochs were associated with isolated presentations of the test stimuli (trials), rather than with consecutive segments of

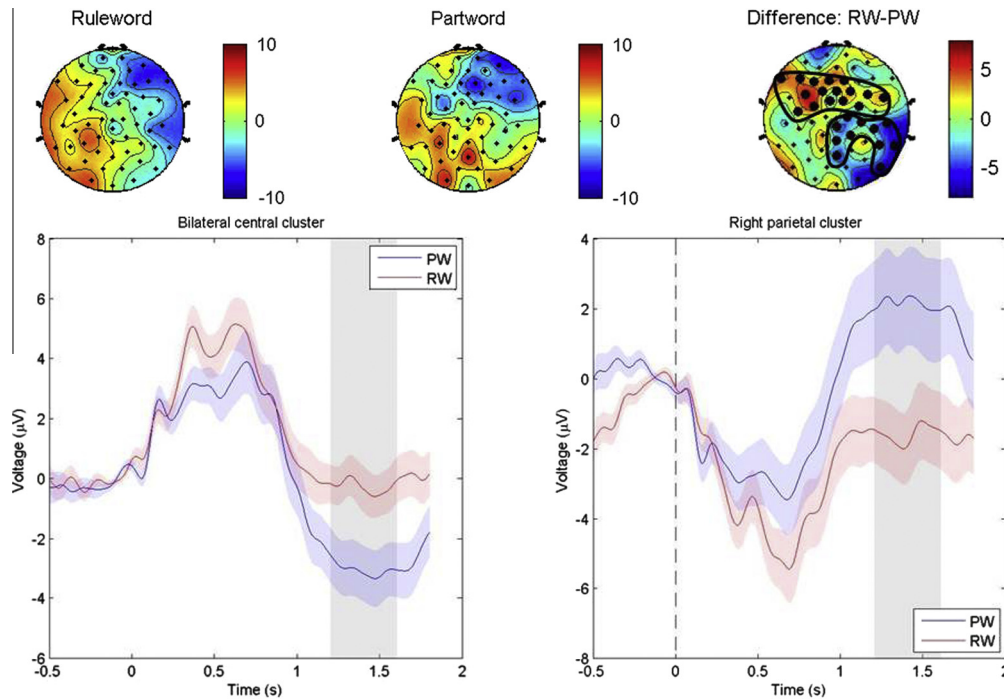
the speech. We assessed statistical differences between conditions using the same customized cluster analysis coupled with a randomization procedure applied to the PLV in the learning part. 5000 permutations were built by randomly attributing the labels *rule-word* and *part-word* to the two conditions in each subject. To boost statistical sensitivity, we set the first thresholding level to 0.2 (a parameterization that affects the sensitivity of the test but not its validity as it does not increase the level of false alarms (Maris & Oostenveld, 2007)). For the significant clusters ( $p < 0.05$ ), we investigated potential group effects by entering in an ANOVA the PLV averaged over the clusters, with group as between-subjects factor.

We also examined the whole range of frequencies up to 35 Hz: the signal was convoluted by complex Morlet wavelets. We set the number of cycles in each Morlet wavelet as increasing linearly with frequency, from 4 at the lowest frequency to 17.5 at highest. The analysis resulted in 200 time points, from  $-885$  to 1605 ms. To get a reasonably good compromise between time and frequency resolution, we started our analysis at 4 Hz (The previous analysis already examined low frequencies at the syllable and word frequency (1.39 and 4.17 Hz) which would be the main components of a general analysis under 4 Hz), and we investigated phase-locking value up to 35 Hz. The phase-locking value of the oscillatory burst was then computed for each time frequency point across trials.

We used again a cluster analysis coupled with a randomization procedure to identify significant differences between *rule-words* and *part-words* over the time period [0–1360] ms. After having computed a two-tailed paired  $t$ -test for each point, spatio-spectro-temporal clusters statistics were defined as the sum of the  $t$  values above a threshold ( $p$  value  $< 0.05$ ) on contiguous points, in terms of space, time and frequency, separately for positive and negative  $t$ -values. Permutations and cluster-level  $p$ -values were computed as above.



**Fig. 2.** Event-related potentials in response to *rule-words* (red) and *part-words* (blue). Early response: Top panel shows the topography of the response to *rule-words* and *part-words*, and the difference (RW – PW). Lower graphs show the time course of the voltage averaged over the circled groups of electrodes, for the left and right hemispheres (dotted lines: posterior clusters, plain lines: anterior clusters).



**Fig. 3.** Event-related potentials in response to *rule-words* (red) and *part-words* (blue). Late response: Top panel shows the topography of the response to *rule-words* and *part-words*, and the difference (RW – PW). Lower graphs show the time course of the voltage averaged over the circled groups of electrodes from the bilateral central cluster, and the negative right parietal cluster.

#### 2.4.3. Correlation between learning and test

Finally, we explored the correlations between the subjects' responses during the test and the learning part, by submitting PLV from learning phase and from test phase to bivariate correlation tests. On one hand, we used the normalized difference of phase-locking value between *rule-word* and *part-word*  $((RW - PW)/(RW + PW))$  averaged over the significant clusters identified during the test analyses, and on the other hand the phase locking values at word and syllable frequencies, averaged over the significant clusters, identified during the learning phase statistical analyses.

### 3. Results

#### 3.1. Learning part: Neural correlates of online statistical computations

A large cluster of electrodes showed significant entrainment at the syllable frequency ( $f_3 = 4.17$  Hz;  $p_{\text{cor}} < 0.001$ ). Crucially, entrainment was also significant at the word frequency ( $f_1 = 1.39$  Hz) in two clusters over temporal regions: a left cluster ( $p_{\text{cor}} = 0.001$ ) spreading over a temporo-frontal region, in between C3, F7 and T7 and a right cluster ( $p_{\text{cor}} = 0.005$ ) located on T8 and posterior, toward P8. No significant entrainment was observed at a two-syllable frequency (no significant cluster was found) (Fig. 1).

The phase-locking values were averaged over these significant clusters separately at the word and syllabic frequencies. No group effect was observed at the syllable presentation rate but a marginal group effect at the word frequency ( $F(2,42) = 2.61$ ,  $p = 0.086$ ), with no significant interaction with hemisphere.

To track the temporal dynamics of learning, we analyzed the power at the word frequency in these clusters across epochs (the PLV is not pertinent here as it needs several epochs to be computed). It significantly increased with the number of epochs ( $p = 0.02$ ,  $r = 0.03$ ). Post-hoc analyses over the two clusters separately revealed a significant increase in the left hemisphere

( $p = 0.01$ ,  $r = 0.08$ ), but not in the right ( $p > 0.1$ ). The power also increased at the syllabic frequency ( $p < 10^{-15}$ ,  $r = 0.32$ ), but not at a bisyllabic rhythm computed over the same clusters.

#### 3.2. Test part: Auditory evoked potentials to rule-words and part-words

Visual inspection of the grand average merged over the two conditions showed 3 components regularly interspaced at 150, 382 and 614 ms (that is each 232 ms which corresponded to the duration of a syllable), congruent with a response evoked by the successive syllables. All three peaks shared a roughly similar topography slightly asymmetric toward the left side and consisting of a bilateral fronto-lateral positivity synchronous of a bilateral posterior negativity. Then, a late response was observed, 500–900 ms after the end of the word, (that is between 1200 and 1600 ms after word-onset). It consisted of a right frontal negativity and a left positivity which progressively extended from anterior temporal areas until posterior temporo-parietal areas.

We selected two time-windows (t1: [350–700] ms encompassing the two peaks of response to the second and third syllable of the word and t2: [1200–1600] ms corresponding to the late response following the word). For t1 ([350–700] ms), we considered symmetrical clusters of electrodes on each hemisphere, located on the frontal positive (10 channels) and posterior negative (9 channels) maxima of the general auditory response. Voltage was averaged across electrodes and time-windows in each infant and entered into an ANOVA (Condition  $\times$  Polarity  $\times$  Hemisphere  $\times$  Group). *Rule-words* evoked a larger auditory response, notably over the left hemisphere, yielding a marginal interaction Condition  $\times$  Polarity ( $F(1,42) = 3.03$ ,  $p = 0.08$ , Fig. 2). Because of the significant interaction Hemisphere  $\times$  Polarity ( $F(1,42) = 15.98$ ,  $p < .001$ ), we computed post hoc analyses on each hemisphere. A significant Condition  $\times$  Polarity interaction was observed on the left hemisphere ( $F(1,42) = 5.31$ ,  $p = 0.03$ ), but not on the right



( $F(1,42) < 1$ ). Post-hoc analyses on each cluster of the left hemisphere revealed marginally significant Condition effects ( $F(1,42) = 4.04$ ,  $p = 0.05$  on the left positive cluster and  $F(1,42) = 3.31$ ,  $p = 0.08$  on the left negative cluster).

For  $t_2$  ([1200–1600] ms), we defined two clusters over the positive and negative maxima of the difference between the two conditions, i.e. sixteen electrodes over the fronto-central region and eleven electrodes over the right posterior temporal parietal region. Voltages averaged across the time window and over each cluster in each infant were entered in an ANOVA (Condition  $\times$  Polarity  $\times$  Group). The Condition  $\times$  Polarity interaction was significant ( $F(1,42) = 7.41$ ,  $p = 0.009$ , Fig. 3). The Condition effect was significant on each cluster ( $F(1,42) = 8.34$ ,  $p = 0.006$  for the positive cluster, and ( $F(1,42) = 4.87$ ,  $p = 0.033$  for the negative cluster).

In none of the analyses, there was a significant group effect or interaction of this factor with the other factors.

### 3.3. Test part: Phase-locking value

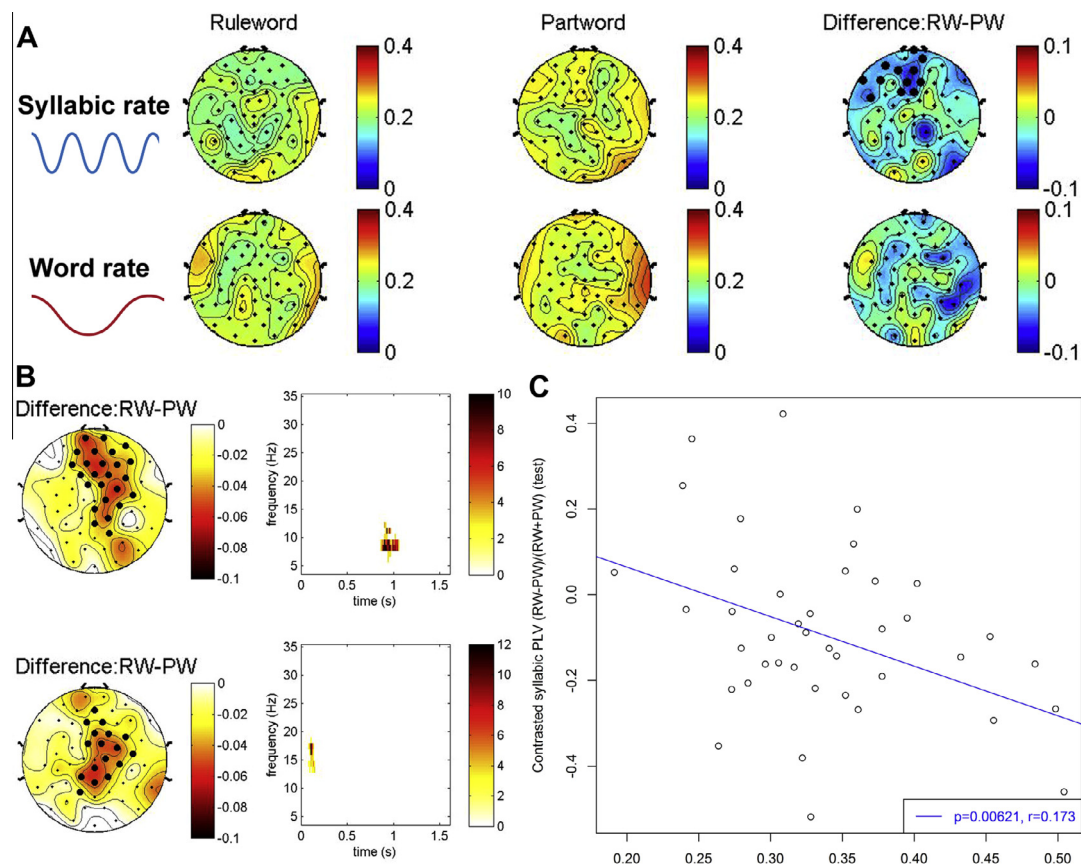
The analysis of the phase-locking value revealed no difference between *rule-* and *part-words* at the word frequency ( $f = 1.39$  Hz). By contrast, a cluster of 12 frontal electrodes around electrode F3 displayed significantly higher PLV for *part-word* than for *rule-word* ( $p_{\text{cor}} = 0.017$ ) at the syllabic frequency ( $f = 4.17$  Hz, Fig. 4A) with no significant group effect ( $F(2,42) < 1$ ).

Inspecting a broader range of frequencies, the analysis revealed two significant negative spectro-temporal clusters and

no positive cluster for the contrast *rule-word* minus *part-word* (Fig. 4B). In other words, there was a higher PLV for *part-words* relative to *rule-words* in these clusters: first in the beta band (13–18 Hz,  $p_{\text{cor}} = 0.018$ ) between 70 and 150 ms after stimulus onset (i.e. during the first syllable) in a cluster surrounding Cz, and second in the alpha band (8–12 Hz,  $p_{\text{cor}} = 0.004$ ) between 850 ms and 1050 ms (i.e. after the word), in a cluster spreading over frontal areas, around electrode Fz, and slightly to the right toward electrode F4.

### 3.4. Correlation between learning and test parts

As significant phase coherences (PLV) were recorded both in the learning part and in the test part, we examined whether these two effects were linked. We thus averaged the PLV difference between *rule-words* and *part-words* over the significant spatial clusters identified in the PLV analyses above and entered these values in two separate linear regressions with the phase-locking values of the learning part at the word and syllable frequencies. No correlation was significant at the word frequency but at the syllable frequency, there was a negative correlation between PLV in learning and test ( $p = 0.006$ ,  $r = 0.17$ ). In other words, the stronger the syllabic entrainment during training (PLV closer to one), the stronger the PLV for *part-words* at the syllable frequency on frontal electrodes (PLV away from zero, corresponding to a stronger PLV for *part-words* relative to *rule-words*, Fig. 4C).



**Fig. 4.** Phase locking value (PLV) analysis in response to *rule-words* and *part-words*. A: PLV at syllabic and word rates in response to *rule-words* (first column) and *part-words* (middle column), and the difference (RW – PW, third column). Electrodes showing a significant difference between the two conditions are highlighted. B: PLV differences between *rule-words* and *part-words* computed over a broader range of frequencies. The left panels show the significant spatial clusters for the PLV difference (RW – PW), while the right panels show the corresponding spectro-temporal significant clusters. The intensity of the time–frequency representations represents the number of electrodes showing a significant effect for the considered time–frequency point. C: Significant learning–test correlation between the syllabic phase-locking value recorded during learning and the normalized phase-locking value difference between *rule-words* and *part-words* recorded during test.



## 4. Discussion

In this study, we investigated whether 8 month-old infants were able to extract word structure from a stream of concatenated syllables with no prosodic information (except a subliminal pause between words) by means of long-distance statistics. We observed several effects related to a successful learning. First, during the stream, we recorded an increase in the phase-locking value at the word frequency over the left and right temporal areas, demonstrating that infants were sensitive to a tri-syllabic unit. Second, we observed a significant difference between *rule-words* and *part-words* in the test part that could only be explained by expectations created during the learning phase on the word structure, given that all syllables of the test items had been equally presented in the learning stream – and were thus similarly familiar. Finally, a stronger syllabic entrainment during learning was correlated with a larger difference between conditions during test establishing a relation between the learning and test responses.

### 4.1. EEG evidence of on-line continuous stream segmentation

The advantage of EEG relative to behavioral studies is that it is possible to examine the responses while infants are exposed to a speech stream and to follow their learning on-line. We recorded an increased phase-locking to the speech stream at the syllable frequency, but above all at the word frequency over the left and right temporal areas. In adults, a syllabic entrainment has already been related to on-line auditory processing of the slow fluctuations of the speech envelope (Ahissar et al., 2001; Luo & Poeppel, 2007; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). In a similar paradigm than here but in adults, Buiatti et al. (2009) reported a cortical entrainment at the syllable frequency for streams of randomly concatenated syllables, and more importantly at the word frequency, only when the stream had an AxC structure, with subliminal pauses at word edges. A subliminal pause inserted every three syllables was not sufficient by itself to induce a trisyllabic entrainment when syllables were randomly concatenated, nor an AxC structure when there was no pause between words. It was hypothesized that the pause, although not explicitly detected, was nevertheless converted in a prosodic cue equivalent to a lengthening of the previous vowel, a common signal for word ending (Pena et al., 2002).

This cue was used to limit adjacent and non-adjacent TP computations and thus favored the discovery of the AxC structure and the segmentation of the words (Buiatti et al., 2009; Pena et al., 2002). Several experiments in adults (Shukla, Nespors, & Mehler, 2007) and infants (Johnson & Jusczyk, 2001; Shukla, White, & Aslin, 2011) have shown that prosodic cues reduce the influence of TP computations in speech segmentation. Morgan, Meier, and Newport (1987) have postulated that prosodic cues are essential to bracket the speech stream and limit distributional analyses to small chunks of relevant co-occurrences. Shukla et al. (2011) reported that mapping a word (a pair of syllables with a TP of 1) on a visual form was considerably facilitated in 6-month-old infants if this word-form was aligned on a phrase boundary. The pause could have played a similar role here. Because neonates are already sensitive to subtle prosodic cues such as the lengthening of the last syllable of a word (Christophe, Mehler, & Sebastián-Gallés, 2001), our 8-month-old infants were certainly using this cue as word ending cue, similarly to adults. Thus, as adults, infants may use both the A–C systematic dependency and the subliminal pause as cues to segment the continuous stream in tri-syllabic units, this hypothesis might explain the recorded entrainment at the word frequency.

### 4.2. Infants differently respond to rule-words and part-words during the test

In behavioral studies, it is expected that when isolated words are presented after the familiarization stream, infants respond differently to words consistent or not with the learned template, because they are primed to expect similar structures to those detected in the stream (Saffran et al., 1996). It is what we observed here, confirming that infants have learned a structure during the short 2 min of exposure.

The observed differences can be grouped in two stages: early effects during the presentation of the word mainly visible over left electrodes, which may correspond to linguistic analyses of each syllable relative to a learned template, and late effects after the end of the word which may correspond to the orientation of attention to unexpected events.

Firstly, a significant difference in the beta range was observed during the presentation of the first syllable, around 100 ms. This effect can be related to the infants' expectation of putative first syllables. Secondly, a significant evoked response was recorded in response to the second and third syllables of the word (350–700 ms) with significantly larger voltage for *rule-words* than for *part-words*. Thirdly, the significant increase of the phase-locking value at the syllable frequency for *part-words* relatively to *rule-words* might be related to a phase resetting for each syllable, as *part-words* were deviating from the learned template at each syllable position. These effects which appeared to be related to syllable analysis were predominantly observed over left electrodes (Figs. 3 and 4). Although it is always tricky to infer cortical sources from surface voltage, the bilateral front-back polarity reversal of the ERPs suggests bilateral sources in temporal areas with stronger responses originating from the left hemisphere producing stronger differences over the frontal and posterior poles during word presentation. This stronger involvement of the left hemisphere is in line with several studies showing an early left-hemispheric bias in speech processing (Dehaene-Lambertz & Baillet, 1998; Dehaene-Lambertz et al., 2010).

After the ending of the test item (1200–1600 ms), a late difference in ERPs was observed, mainly driven by *part-words* (Fig. 3). Late responses are often recorded in attentive infants when unexpected events occurred (Csibra, Kushnarenko, & Grossmann, 2008), for example after a rare event in oddball designs (Basirat, Dehaene, & Dehaene-Lambertz, 2014). This effect was preceded by a phase-locking value (PLV) difference between conditions in the alpha range between 850 and 1050 ms due to a PLV increase for *part-words*. Alpha oscillations have been related to the inhibition of irrelevant information (i.e. *part-words* were unexpected) but also to working memory load (i.e. attempt to recover statistics of the stream). In particular, increase in alpha power was reported during the computation of long-distance relation between argument and verb in sentence listening in adults (Meyer et al., 2013). These late differences might thus reflect infant's orientation of attention to the unexpected structure of *part-words* and attempts of corrections. They may represent the cerebral markers of the longer looking-time reported in behavioral studies.

### 4.3. What did infants learn?

Thus, after a brief familiarization to a continuous stream of AxC word, eight-month-old infants differentially reacted to *rule-words* and *part-words* during the test-part. Note first that differences between *part-words* and *rule-words* cannot be explained by the absolute frequency of syllable presentation in the stream. Indeed, all test syllables were experienced during learning phase, and were thus equally familiar to the infants.

It is also improbable that infants have learned adjacent transitional probabilities. Although the average adjacent TPs were slightly larger between words than within words (0.5 vs 0.33) in the stream, the range of variation was close (0.43 to 0.57 vs. 0.25 to 0.42). Previous studies have shown that infants could use adjacent TP as a cue to group syllables into units but in the systematic case of a large difference between word-internal and word-external TP (e.g. 0.33 between words vs. 1 within words in Saffran et al. (1996)). Moreover the larger, and a priori more memorable, word-external adjacent TP had to be computed across the pause which usually acts as a prosodic boundary in adults (Buiatti et al., 2009; Pena et al., 2002), blocking TP computations (Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; Mintz, Newport, & Bever, 2002; Shukla et al., 2007). Another explanation could be that instead of computing statistics, infants have simply memorized all pairs of syllables they have been exposed to during the stream, and therefore processed differently *part-words*, containing a familiar pair, and *rule-words* containing novel pairs. The memory effort is certainly too important at this age especially given the short 2 min duration of the stream and given the fact that we constrained each syllable not to be repeated within the next three after its presentation, increasing the delay between repetition of the same pair. Finally, attention to pairs of syllables should have elicited responses at the bisyllable frequency during the stream. This was not observed.

Thus, the only systematic dependency in the stream was between the first and third syllable and the PLV at word frequency during the stream establishes that infants were indeed sensitive to this unit. Once this dependency acquired, infants might rely on different strategies when presented with isolated words. They might have expected the first syllable to be part of the pool of A syllables (ku, na or li), and/or the third syllable to be part of the pool of C (bi du or fe), and/or that a syllable coming from the C pool should always be at the end of a word. Thus the A\_C dependency might be less important during the test than the position of the syllables but note that their position can have only been learned because of the A\_C association in the stream. The differences between *rule-words* and *part-words* observed as early as the presentation of the first syllable, suggest that infants had an expectation about the identity of the first syllable. Whether they use the A\_C dependency in test should be further studied in follow-up experiments by exchanging the last syllable between families. However, younger infants, at 4-months of age, noticed this exchange (Friederici et al., 2011). It was also the case in the 7-month-olds in Marchetto and Bonatti's study (2014). A second unresolved question concerns the generalization process and whether infants similarly consider a word heard in the stream and a rule-word. This question is ambiguous and depends on the task. In a forced-choice task, adults preferred rule-words to part-words (Pena et al., 2002), but when items were presented one at a time, they accepted words, rejected part-words but were at random for rule-words (Buiatti et al., 2009). This ambiguity is also present in behavioral experiments in infants, such as in Marchetto and Bonatti's study (2014), in which it is not possible to disambiguate if infants rejected part-words or accepted rule-words, and thus were truly generalizing the rule to new items. These questions should be further pursued in future experiments.

#### 4.4. Steady-state response indexes subjects' performances

Finally, we investigated whether there was a link between the brain responses in the learning and test parts, and we reported that a stronger syllabic entrainment during learning was correlated with a larger difference between conditions, due to a stronger PLV at the syllable frequency for *part-words*. First, this correlation established that previous implicit learning during concatenated

speech presentation was re-used by infants when listening to isolated trisyllabic items. Second, this result echoes previous findings in adults by Ahissar et al. (2001), who found that cortical response locking to the temporal structure of speech envelope is an on-line requirement of speech comprehension; and by Henry and Obleser (2012) who demonstrated that phase realignment while listening to a rhythmic stream optimizes perceptual processing. The steady-state response to syllables may thus provide an index of infants' speech processing efficiency. The most efficient and/or attentive infants might both compute more easily the statistical regularities in the stream and better detect their violation in the case of the *part-words*. Surprisingly, we did not observe any correlation between test responses and what we interpreted as a direct measure of segmentation: trisyllabic entrainment. This might be due to the weaker size of the trisyllabic effect, spatially and/or temporally, which might have reduced the statistical power of the correlation. Alternatively, this result might indicate that syllabic stimulus–response locking plays a crucial role in speech processing at this age and that temporal reliability at the segmental level might predict individual differences in language acquisition.

#### 4.5. Discrepancies between behavioral and EEG studies

We observed long-distance statistical learning at an earlier age than many behavioral studies. Furthermore, we were not able to identify any significant differences between our groups, suggesting that 8 months of exposure to ex-utero speech were sufficient to achieve the task for both full- and preterms who had different degree of brain maturation, or that this capacity is acquired before 5 months post-term (neural age of one of our preterm group). This age is in line with Friederici et al.'s (2011) study reporting that 4-month-old German infants after being exposed to naturally produced Italian sentences (*sta-x-ando* or *puo-x-are*), detect ungrammatical sentences in which verbs inflexions have been exchanged. By contrast, Gómez and Maye (2005) and Marchetto and Bonatti (2014) failed to observe generalization of long-distance dependency learning before 12 months of age. The age discrepancy between EEG and behavioral studies might be due to several paradigm differences. With EEG, the number of test trials is more important and the different types of words are randomly interspaced and not presented in lists, decreasing the possibility of learning during test. The number of families, the number of words per family and the duration of the pauses are also crucial differences that may favor different strategies in infants. For example, in a close design to ours, Marchetto and Bonatti (2014) used two words per family, separated by 200 ms-long silence. Silence duration is an important factor to grade prosodic units and recover speech hierarchical structure. Long pauses might favor a sentence-level analysis whereas subliminal pauses may restrict analyses to the word-level. These different levels of linguistic computations relies on different brain regions that may be at different stages of maturation and thus of efficiency (Dehaene-Lambertz et al., 2006).

## 5. Conclusion

In conclusion, we suggest here that infants, from at least 8-month-old (and probably younger infants given the lack of differences between our groups), are able to compute long-distance dependencies to segment a continuous stream of concatenated speech. However, given the very specific artificial and impoverished experimental conditions, our result is more a proof of concept rather than a claim that infants actively use this strategy in the complex natural speech condition. The words embedded in the stream had for example the same number of syllables, and it has been shown that when this number varied, 8-month-olds were

no longer able to show evidence of word recognition based on adjacent transitional probabilities (Johnson & Tyler, 2010). We also raised several questions during discussion that should be answered by future studies. This result nevertheless enriched the list of young infants' abilities to process speech, and notably revealed that long-distance dependencies can be computed at a similar age than adjacent dependencies, and may thus rely on similar computational resources. Finally, EEG and in particular frequency tagging may represent a more sensitive method than looking behavior to target implicit computations as those involved in statistical learning and to highlight individual differences in speech acquisition. We recorded here a correlation between the phase-locking value, that is the temporal precision of the coding, and the latter discriminative responses during test. Temporal imprecision has been proposed as one of the mechanism at the origin language impairment (Abrams, Nicol, Zecker, & Kraus, 2006).

## Acknowledgments

CK and GDL were supported by the Bettencourt Foundation, Fondation de France and MP by the grant Fonddecy 1141040.

## References

- Abrams, D. A., Nicol, T., Zecker, S. G., & Kraus, N. (2006). Auditory brainstem timing predicts cerebral asymmetry for speech. *Journal of Neuroscience*, 26(43), 11131–11137.
- 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. *Proceedings of the National Academy of Sciences*, 98(23), 13367–13372.
- Basirat, A., Dehaene, S., & Dehaene-Lambertz, G. (2014). A hierarchy of cortical responses to sequence violations in three-month-old infants. *Cognition*, 132(2), 137–150. <http://dx.doi.org/10.1016/j.cognition.2014.03.013>.
- Bernal, S., Dehaene-Lambertz, G., Millotte, S., & Christophe, A. (2010). Two-year-olds compute syntactic structure on-line. *Developmental Science*, 13(1), 69–76.
- Buiatti, M., Pena, M., & Dehaene-Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *NeuroImage*, 44(2), 509–519.
- Christophe, A., Mehler, J., & Sebastián-Gallés, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy*, 2, 385–394.
- Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language*, 51(4), 523–547.
- Csibra, G., Kushnerenko, E., & Grossmann, T. (2008). Electrophysiological methods in studying infant cognitive development. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (2nd ed., pp. 247–262). Cambridge, Massachusetts: MIT Press.
- Dehaene-Lambertz, G., & Baillet, S. (1998). A phonological representation in the infant brain. *NeuroReport*, 9(8), 1885–1888.
- Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Meriaux, S., Roche, A., & Sigman, M. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proceedings of the National Academy of Sciences of the United States of America*, 103(38), 14240–14245. doi: 0606302103 [pii] 10.1073/pnas.0606302103.
- Dehaene-Lambertz, G., Montavont, A., Jobert, A., Alliro, L., Dubois, J., & Hertz-Pannier, L. (2010). Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain and Language*, 114(2), 53–65.
- 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. <http://dx.doi.org/10.1016/j.jneumeth.2003.10.009>.
- Dutoit, T. (1997). *An introduction to text-to-speech synthesis* (Vol. 3). Springer.
- Feldman, R., Eidelman, A. I., Sirota, L., & Weller, A. (2002). Comparison of skin-to-skin (kangaroo) and traditional care: Parenting outcomes and preterm infant development. *Pediatrics*, 110(1), 16–26.
- Fontolan, L., Morillon, B., Liegeois-Chauvel, C., & Giraud, A.-L. (2014). The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nature Communications*, 5.
- Forget, J., Buiatti, M., & Dehaene, S. (2009). Temporal integration in visual word recognition. *Journal of Cognitive Neuroscience*, 22(5), 1054–1068. <http://dx.doi.org/10.1162/jocn.2009.21300>.
- Friederici, A. D., Mueller, J. L., & Oberecker, R. (2011). Precursors to natural grammar learning: Preliminary evidence from 4-month-old infants. *PLoS One*, 6(3), e17920.
- Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, 7(2), 183–206.
- Gonzalez-Gomez, N., & Nazzi, T. (2012). Phonotactic acquisition in healthy preterm infants. *Developmental Science*, 15(6), 885–894. <http://dx.doi.org/10.1111/j.1467-7687.2012.01186.x>.
- Goodsitt, J. V., Morgan, J. L., & Kuhl, P. K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language*, 20(2), 229–252.
- Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 109(49), 20095–20100. doi: 1213390109 [pii] 10.1073/pnas.1213390109.
- 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. <http://dx.doi.org/10.1016/j.cognition.2010.03.006>.
- Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., & Schmitz, M. (2004). Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy*, 5(3), 341–353.
- Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, 44, 548–567.
- Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, 13(2), 339–345. <http://dx.doi.org/10.1111/j.1467-7687.2009.00886.x>.
- Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party". *Journal of Neuroscience*, 30(2), 620–628. <http://dx.doi.org/10.1523/JNEUROSCI.3631-09.2010>.
- Kouider, S., Stahlhut, C., Gelskov, S. V., Barbosa, L. S., Dutat, M., & De Gardelle, V. (2013). A neural marker of perceptual consciousness in infants. *Science*, 340(6130), 376–380.
- Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, 54(6), 1001–1010.
- Marchetto, E., & Bonatti, L. L. (2014). Finding words and word structure in artificial speech: The development of infants' sensitivity to morphosyntactic regularities. *Journal of Child Language*.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. <http://dx.doi.org/10.1016/j.jneumeth.2007.03.024>.
- Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, 11(1), 122–134. <http://dx.doi.org/10.1111/j.1467-7687.2007.00653.x>.
- Maye, J., Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, B101–B111.
- Meyer, L., Obleser, J., & Friederici, A. D. (2013). Left parietal alpha enhancement during working memory-intensive sentence processing. *Cortex*, 49(3), 711–721. <http://dx.doi.org/10.1016/j.cortex.2012.03.006>.
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90(1), 91–117. [http://dx.doi.org/10.1016/S0010-0277\(03\)00140-9](http://dx.doi.org/10.1016/S0010-0277(03)00140-9).
- Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, 26(4), 393–424.
- Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, 19, 498–550.
- Mueller, J. L., Friederici, A. D., & Mannel, C. (2012). Auditory perception at the root of language learning. *Proceedings of the National Academy of Sciences of the United States of America*, 109(39), 15953–15958. <http://dx.doi.org/10.1073/pnas.1204319109>.
- Nelson, C. A., & deRegnier, R. A. (1992). Neural correlates of attention and memory in the first year of life. *Developmental Neuropsychology*, 8(2–3), 119–134.
- Palva, J. M., Monto, S., Kulashekhar, S., & Palva, S. (2010). Neuronal synchrony reveals working memory networks and predicts individual memory capacity. *Proceedings of the National Academy of Sciences of the United States of America*, 107(16), 7580–7585. doi: 0913113107 [pii] 10.1073/pnas.0913113107.
- Pediatrics, A. A. o. (1998). American academy of pediatrics. Committee on nutrition. Cholesterol in childhood. *Pediatrics*, 101(1 Pt 1), 141.
- Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, 23(6), 1378–1387. <http://dx.doi.org/10.1093/cercor/bhs118>.
- Pena, M., Arias, D., & Dehaene-Lambertz, G. (2014). Gaze following is accelerated in healthy preterm infants. *Psychological Science*. <http://dx.doi.org/10.1177/0956797614544307>.
- Pena, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298(5593), 604–607.
- Pena, M., Pittaluga, E., & Mehler, J. (2010). Language acquisition in premature and full-term infants. *Proceedings of the National Academy of Sciences of the United States of America*, 107(8), 3823–3828. doi: 0914326107 [pii] 10.1073/pnas.0914326107.
- Pena, M., Werker, J. F., & Dehaene-Lambertz, G. (2012). Earlier speech exposure does not accelerate speech acquisition. *The Journal of Neuroscience*, 32(33), 11159–11163.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928.
- Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition*, 69(2), 105–134. [http://dx.doi.org/10.1016/S0010-0277\(98\)00060-2](http://dx.doi.org/10.1016/S0010-0277(98)00060-2).
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234), 303–304.



- Shi, R., Cutler, A., Werker, J., & Cruickshank, M. (2006). Frequency and form as determinants of functor sensitivity in English-acquiring infants. *The Journal of the Acoustical Society of America*, 119(6), EL61–EL67. <http://dx.doi.org/10.1121/1.2198947>.
- Shukla, M., Nespors, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, 54(1), 1–32. <http://dx.doi.org/10.1016/j.cogpsych.2006.04.002>.
- Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, 108(15), 6038–6043. <http://dx.doi.org/10.1073/pnas.1017617108>.
- Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *The Journal of Neuroscience*, 16(13), 4240–4249.
- van Heugten, M., & Shi, R. (2010). Infants' sensitivity to non-adjacent dependencies across phonological phrase boundaries. *The Journal of the Acoustical Society of America*, 128(5), EL223–EL228. <http://dx.doi.org/10.1121/1.3486197>.
- Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, 7(12), 553–559.