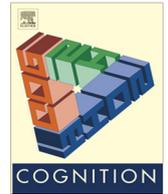




ELSEVIER

Contents lists available at ScienceDirect

Cognition

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

A hierarchy of cortical responses to sequence violations in three-month-old infants



Anahita Basirat^{a,b,c,*}, Stanislas Dehaene^{a,b,c,d}, Ghislaine Dehaene-Lambertz^{a,b,c}

^aINSERM, U992, Cognitive Neuroimaging Unit, F-91191 Gif/Yvette, France

^bCEA, DSV/I2BM, NeuroSpin Center, F-91191 Gif/Yvette, France

^cUniversity Paris-Sud, Cognitive Neuroimaging Unit, F-91191 Gif/Yvette, France

^dCollège de France, F-75005 Paris, France

ARTICLE INFO

Article history:

Received 3 August 2012

Revised 30 January 2014

Accepted 28 March 2014

Available online 5 May 2014

Keywords:

Mismatch response
Prediction
Brain development
Sequence learning
Rule extraction

ABSTRACT

The adult human brain quickly adapts to regular temporal sequences, and emits a sequence of novelty responses when these regularities are violated. These novelty responses have been interpreted as error signals that reflect the difference between the incoming signal and predictions generated at multiple cortical levels. Do infants already possess such a hierarchy of violation-detection mechanisms? Using high-density recordings of event-related potentials during an auditory local-global violation paradigm, we show that three-month-old infants process novelty in temporal sequences at two distinct levels. Violations of local expectancies, such as perceiving a deviant vowel “a” after repeated presentation of another vowel i-i-i, elicited an early auditory mismatch response. Conversely, violations of global expectancies, such as hearing the rare sequence a-a-a-a instead of the frequent sequence a-a-a-i, modulated this early mismatch response and led to a late frontal negative slow wave, whose cortical sources included the left inferior frontal region. These results suggest that the infant brain already possesses two dissociable systems for temporal sequence learning.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The classical constructivist perspective postulates that learning starts at an early sensory level and very slowly progresses towards increasingly abstract and logical levels (Piaget, 1954; Quartz & Sejnowski, 1997). The first functional MRI and NIRS studies in infants were thus surprising, revealing the involvement of high-level brain areas such as dorsolateral prefrontal cortex and Broca’s area at an early age (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002; Dehaene-Lambertz et al., 2010; Mahmoudzadeh et al.,

2013; Perani et al., 2010), supported by an efficiency long-range connectivity (Leroy et al., 2011). Learning in infants might thus not be limited to low-level processes, but might occur at all levels along the processing hierarchy, as proposed by recent Bayesian models of child development (Tenenbaum, Kemp, Griffiths, & Goodman, 2011; Téglás et al., 2011), with high-level regions generating top-down predictions modulating the down-stream computations (Friston, 2005; Rao & Ballard, 1999). Here, we test the hypothesis that the infant brain, at three months of age, already processes information about auditory sequences at two hierarchical levels. Using an auditory violation paradigm, we demonstrate that the infant brain contains a hierarchy of error signals that respond, respectively, to violations of local and global auditory sequences. We argue that the presence of these signals suggest that the infant

* Corresponding author at: INSERM, U992, Cognitive Neuroimaging Unit, F-91191 Gif/Yvette, France.

E-mail address: anahita@basirat.fr (A. Basirat).

brain, at three months, already generates top-down predictions about future incoming stimuli.

A simple and widely used paradigm to study infant auditory perception is the auditory oddball paradigm, in which a novel sound is introduced after a series of repeated sounds. This abrupt change generally elicits an early mismatch response (MMR), often consisting of a frontal positivity synchronous of a posterior negativity, around 200–400 ms after the deviant stimulus, generally followed, around 700 ms, by a late frontal Negative Slow Wave (NSW; Dehaene-Lambertz & Dehaene, 1994; Friederici, Friedrich, & Weber, 2002).

This two-stage response in infants is reminiscent of the MMN/P300 complex reported in adults, even if the latencies and topographies of these responses are different due to the immaturity of the infant's brain. In adults, an early and automatic mismatch response is recorded around 100 ms and consists of a frontal negativity with a polarity reversal above temporal regions (mismatch negativity or MMN, Näätänen, Gaillard, & Mäntysalo, 1978), and a late central positivity is recorded around 300 ms (P300 or P3b, Squires, Squires, & Hillyard, 1975). In adults, these two components differ in their functional properties: The MMN is present even if the subject does not pay attention to the stimuli, is asleep (Atienza, Cantero, & Gomez, 1997) or in coma (Fischer et al., 1999), but it disappears when the inter-stimulus-interval (ISI) is increased beyond a few seconds (Mäntysalo & Näätänen, 1987; Pegado et al., 2010). By contrast, the P300 is only present if the subject is conscious and attentive (Bekinschtein et al., 2009), and is not affected by long ISI (Wetter, Polich, & Murphy, 2004). The MMN has been associated with unconscious processing of auditory transition probabilities (Näätänen, Paavilainen, Rinne, & Alho, 2007; Wacongne, Changeux, & Dehaene, 2012; Winkler, 2007), and the P300 with conscious detection of novelty and “context updating” (Dehaene & Changeux, 2011; Donchin & Coles, 1988; Sergent, Baillet, & Dehaene, 2005).

Partially similar observations have been made in infants. The early MMR can be elicited in non-attentive or sleeping infants (Dehaene-Lambertz & Peña, 2001), and is reduced by long ISI (Cheour et al., 2002). Its brain sources are mainly located in the superior temporal regions (Bristow et al., 2009; Dehaene-Lambertz & Dehaene, 1994), congruent with the adults' description of the MMN sources (Celsis et al., 1999; Halgren, Sherfey, Irimia, Dale, & Marinkovic, 2011). The late response (NSW) is less often reported than the MMR, but this could simply be because it occurs too late to fit within the length of the studied ERP epoch. The NSW belongs to a set of late components observed in infants which have been linked to attention and novelty detection (Csibra, Kushnerenko, & Grossmann, 2008), and more recently to conscious perception (Kouider et al., 2013). Indeed, when comparing awake and asleep infants, Friederici et al. (2002) observed the NSW after a deviant sound only in awake infants.

The functional similarities of the MMR/NSW with the adult MMN/P300 components suggest a putative parallel with the adult functional architecture (i.e. an early automatic local response vs. a late context-dependent response). Yet at present, no study has tried to disentangle

whether the MMR and NSW are sensitive to different types of violations. A recently introduced hierarchical “local–global” paradigm epitomizes the two distinct processing stages behind the generation of a MMN and a P300 in adults (Bekinschtein et al., 2009). This paradigm measures brain responses to auditory novelty at two hierarchical levels. At the first level, a novel sound is introduced after a series of repeated sounds (e.g. xxxY, where x denotes the repeated sound and Y the novel sound), generating a “local” deviancy. At the second level, a series of sounds is selected as the frequent global sequence for a block of trials (e.g. xxxY), and then this sequence is violated on a rare subset of trials (e.g. by occasionally presenting the sequence xxxx). With this paradigm, Bekinschtein et al. (2009) disentangled two properties of the adult MMN and P300 responses. First, local deviants (the last sound Y in sequence xxxY) systematically elicit a MMN, even when the sequence itself is frequent and predictable; this response is automatic and remains present in inattentive or comatose subjects. It corresponds to an automatic error-signal generated when the incoming sound differed from what was expected given the previous sounds (Garrido, Kilner, Kiebel, & Friston, 2007; Garrido et al., 2008; Näätänen et al., 1978; Wacongne et al., 2011, 2012; Winkler, 2007). Second, global deviants (rare sequences) systematically elicit a P300 response, even when the rare deviating sequence is a monotonous sequence of repeated sounds (xxxx).

The latter finding is particularly diagnostic of a second-order computation. While the first stage (MMN) simply weights the incoming sound against predictions based on past events, the second stage (P300) seems sensitive to the global rule governing the entire sequence. Especially, generating an error signal to a perfectly monotonic “xxxx” sequence (in a block where most trials are xxxY) can only be performed by a system that actively generates an expectation that the sequence should end with a different sound (xxxY). This second stage, in adults, requires attention to the sequence (Bekinschtein et al., 2009).

In the present study, we probed the existence of hierarchical novelty detection and predictive processes in three-month-old infants using high-density recordings of event-related potentials during a variant of Bekinschtein et al. (2009) auditory local–global paradigm. To maximize attention, although we were obviously unable to give instructions to our preverbal participants, we used audiovisual speech stimuli which infants spontaneously find strongly attractive (Fig. 1). Stimuli were presented in short series of four vowels, following an xxxY or xxxx pattern in distinct blocks (see Fig. 2 for experimental design). After a short training phase, which let infants learn the global sequence governing the present block, sequences violating this global pattern were randomly presented (i.e. xxxx trials in blocks with rule xxxY, and xxxY trials in blocks with rule xxxx).

Our predictions were simple. Based on the infant literature, we expected to record a mismatch response (MMR) around 200–400 ms after a deviant sound, followed by a late frontal negativity if infants direct their attention toward this novel event (Dehaene-Lambertz & Dehaene, 1994; Friederici et al., 2002). If the infant's MMR is

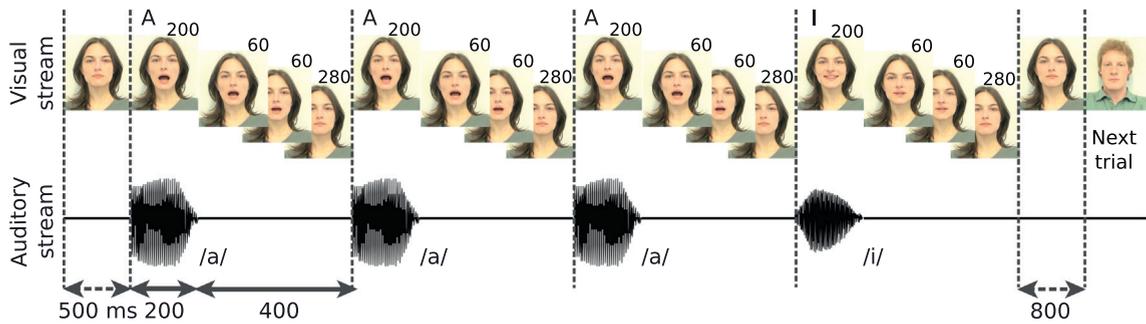


Fig. 1. Temporal structure of a trial. On each trial, infants were presented with a series of four successive vowels (/a/ or /i/). The first three sounds were always identical, and while the last sound could be identical or different. To maximize the infants' attention and minimize fussiness, stimuli were accompanied by a congruent view of a face speaking the same vowel. At the beginning of each trial, a face (and the corresponding voice) was randomly chosen among four different speakers. The mouth was initially closed, and opened suddenly in synchrony with sound onset. At the end of each sound, the mouth closed (two intermediate frames).

Learning phase	Test phase
xxxx rule (aaaa or iiii) [X X X X] Local standard 100% (= 15 items) Global standard	[X X X X] Local standard 75% (= 45 items) Global standard [X X X Y] Local deviant 25% (= 15 items) Global deviant
xxxY rule (aaai or iiii) [X X X Y] Local deviant 100% (= 15 items) Global standard	[X X X Y] Local deviant 75% (= 45 items) Global standard [X X X X] Local standard 25% (= 15 items) Global deviant

Fig. 2. Experimental design. Two types of trials were presented. Local standards, denoted xxxx, comprised four identical vowels (/aaaa/ or /iiii/). Local deviants, denoted xxxY, comprised three identical vowels followed by a different one (/aaai/ or /iiia/). In a given block of trials, a rule (xxxx or xxxY) was selected. During an initial learning phase (15 trials), 100% of trials followed this rule. Then during the test phase, 75% of the trials followed the rule (global standard) while 25% violated it (global deviant). The rule was alternated across blocks in order to expose each infant to both rules. The order was counterbalanced across infants.

equivalent to the adults' MMN, it would occur after any local change (i.e. any xxxY sequence). Because the size of the mismatch depends on the frequency of the standard and deviant sounds in the preceding seconds (King et al., 2013; Sato et al., 2000; Winkler, Cowan, Csépe, Czigler, & Näätänen, 1996), the MMR might be modulated by block structure and be smaller on xxxY blocks (where local violations are frequent) than on xxxx blocks (where they are rare). Crucially however, if infants and adults share a similar hierarchical functional architecture, the late negative slow wave (NSW) should depend on the presence of a global rule violation. Thus, it should be elicited by any rare sequence, even when it merely consists in the repetition of the same sound (xxxx). If such a late response was observed, it would suggest that three-month-old infants are able to represent local and global auditory regularities and detect their violations at two hierarchical levels, as in adults.

2. Methods

2.1. Subjects

Twenty-nine healthy full-term infants (twelve females) were tested between 11 and 15 weeks after birth (mean

age = 13.6 weeks, SD = 1.1 weeks). Twenty additional infants were tested but rejected for fussiness, excessive movement, bad recording or lack of enough data (<150 trials in total or <20 artifact-free trials in the test phase of each rule-type condition [xxxx and xxxY]). The study was approved by the regional ethical committee for biomedical research, and parents gave their written informed consent.

2.2. Stimuli

To maximize infants' attention, we used audio-visual speech stimuli. Two males and two females were filmed articulating /a/ and /i/. Four frames were extracted from each clip: mouth fully opened (frame 1), mouth gradually closed (frames 2 and 3), mouth closed (frame 4). Auditory stimuli matched in duration (200 ms) and subjective intensity and produced by the same speakers were extracted from the audio track.

2.3. Procedure

Infants were seated on their parent's lap in front of a computer screen inside a shielded room (Faraday cage), with the loudspeakers located in the roof of the room. Each

trial consisted of the audio–visual presentation of four vowels with a 600 ms stimulus onset asynchrony (SOA). The onset of the auditory vowel always coincided with a visual presentation of the mouth fully opened as if pronouncing the corresponding vowel. At the end of the sound (after 200 ms), the mouth was gradually closed (two frames of 60 ms duration each), then the face with a closed mouth was presented for 280 ms until the onset of the next vowel (Fig. 1). This design ensured a sufficiently natural mouth movement while perfectly controlling the onset of the audio–visual stimulation. Note that, with this design, the novelty of the fourth event was simultaneously conveyed by audition (vowel) and by vision (mouth shape), a feature which should only facilitate novelty detection and global rule extraction. 1400 ms after the onset of the last syllable of the trial, a new face was presented, randomly chosen amongst the three remaining faces. 500 ms later, the next audio–visual presentation began. Thus, the inter-trial interval was 1900 ms long. Two types of trials were presented: local standard trials, denoted xxxx, in which the same vowel was repeated (/aaaa/ or /iiii/), and local deviant trials, denoted xxxY, in which the last syllable was changed (/aaai/ or /iiia/).

The experiment was separated into four blocks of 75 trials, each lasting ~5 min, following two different types of rules (Fig. 2). In xxxx blocks (aaaa or iiii), infants were expected to learn that all trials were of the xxxx form, whereas in xxxY blocks (aaai or iiia), they were expected to learn that all trials followed the xxxY form. In order to establish the global sequence, during an initial learning phase, the first 15 trials always respected it. This phase was followed by a test phase of 60 trials, in which 75% of the trials were “global standards” which respected the sequence chosen for that block (45 trials) but 25% were “global deviants” which violated it (15 trials). These deviant trials were always followed by at least one standard trial, which was not included in the analyses. The repeated vowel was kept constant during the block, but its gender could change from trial to trial in congruence with the visual stimuli. The rule order was counterbalanced across infants and alternated from one block to the next (e.g. /aaaa/ /iiia/ /iiii/ /aaai/). Stimuli were presented using E-prime v1.2 (Psychology Software Tools, Inc.). The experiment was stopped if the infant became fussy or after the four blocks. All infants were exposed to both rules (see exclusion criteria).

2.4. EEG recordings

Scalp voltages were collected using a 128-channel Geodesic Sensor Net (EGI) referenced to the vertex. They were amplified, sampled at 250 Hz, and filtered between 0.5 and 20 Hz. Trials were segmented relative to the onset of the fourth (or test) syllable (–2100 to 1800 ms). Channels contaminated by eye or motion artifacts were automatically rejected and trials with more than 50% bad channels were excluded. Due to an initial problem in the stimulus presentation program, six infants did not receive the learning phase for the first block. The entire first block was rejected for these infants. Artifact-free trials were averaged per infant and per condition. Averages were baseline corrected

(using 300 ms before the onset of the first sound of the trial) and average referenced. An average number of 97.1 trials were kept in each infant, corresponding to 34.5 xxxx and 17.2 xxxY trials on xxxx blocks, and 30.1 xxxY and 15.3 xxxx trials on xxxY blocks.

2.5. Data analysis

Our 2 × 2 experimental design combined two factors: local deviance (whether or not the last item of the sequence changed, i.e. xxxx vs. xxxY) and global deviance (whether or not the trial sequence violated the block rule). We analyzed the local and global effect by merging the 4 types of trials accordingly: for the local effect (xxxY vs. xxxx), frequent xxxY in the xxxY blocks and rare xxxY in the xxxx blocks were compared to frequent xxxx in the xxxx blocks and rare xxxx in the xxxY blocks. For the global effect, rare xxxY and rare xxxx were compared to frequent xxxx and frequent xxxY.

Given that we used a 128-channels recording system and a time resolution of 4 ms (250 Hz), performing a systematic comparison of all samples would run into the severe risk of false positives. We thus used two orthogonal approaches. We first tested whether, and at which time, the expected two stages of novelty processing were found. We did this computing at each time point the mean across the 129 electrodes of the absolute value of the voltage difference related to the studied effect (local or global deviance). This analysis therefore discarded topographic information and reduced the data to one point per time sample (Fig. 3). We compared the grand-average computed across the infants, to surrogate data obtained by permuting the labels of the different conditions within-subjects. 1000 permutations were done in each infant and 1000 grand-averages were thus computed to which the real grand-average was compared. This evaluated the chance probability of obtaining a difference at least as large as the observed one at each time-point.

Second, to evaluate the topography of the effects while avoiding a combinatorial explosion in the number of statistical tests, we based our statistical analyses on the infant literature and averaged the voltage across pre-selected clusters of electrodes and time-windows, which were determined a priori using previous studies of the early mismatch response and the late slow wave in infants (Bristow et al., 2009; Dehaene-Lambertz & Gliga, 2004; Friederici et al., 2002). Our goal was to understand how these known components were affected by the two types of deviance.

For the MMR, Bristow et al. (2009) performed an audio–visual speech perception experiment in two-month-olds using the same stimuli as here. They reported a dipole configuration with a right positive anterior (around C4, Fz, F4, and F8) and a left posterior negative pole (around O1, P3, T5, mastoid) at about 300 ms after the onset of the deviant sound (see also Dehaene-Lambertz & Dehaene, 1994). Thus, in our analysis, we selected a time window centered on the peak response (270–370 ms) congruent with this study and using clusters of electrodes placed over the same regions (15 anterior channels and 12 posterior electrodes on each side, Fig. 4). We also selected their homologues on the contralateral hemisphere to enter hemisphere as a

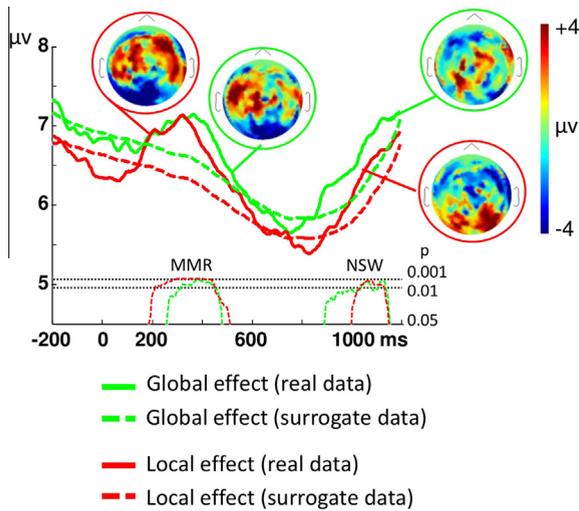


Fig. 3. Auditory deviance affected two distinct time windows (MMR = mismatch response, and NSW = negative slow wave). Grand-averages of the local and the global effects are compared to surrogate data obtained through 1000 permutations of the condition labels within each subject. *P*-values are presented at the bottom of the plot (scale on the right y axis). The *p*-value reached .05 from 252 to 452 ms and from 884 to 1140 ms for the global effect (green lines) and from 180 to 500 ms and from 992 to 1140 ms for the local effect (red lines). Topographies of the voltage averaged across each significant time-window are presented above the plot (green circles for the global effect and red circles for the local effect). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

factor in our analyses and test for hemispheric asymmetries in novelty detection. Indeed, the human brain displays structural and functional asymmetries from the preterm period on. Markers of maturational development show a differential calendar in the left and right hemispheres (Chiron et al., 1997; Dubois et al., 2008; Leroy et al., 2011; Lin et al., 2012) and several studies using near infra-red spectroscopy (Mahmoudzadeh et al., 2013; Telkemeyer et al., 2011), fMRI (Dehaene-Lambertz et al., 2010; Perani et al., 2010), and ERPs (Bristow et al., 2009) describe functional asymmetries which suggest a functional counterpart of the structural asymmetries.

As an additional control for the possibility that the observed differences in response to the last sound were due to chance, we introduced in our analyses the factor “stimulus number”, and tested whether the difference between conditions was significantly larger for the last vowel than during the first three vowels. We thus performed a repeated-measure analysis of variance (ANOVA) on the voltage averaged over the selected time window and cluster of electrodes with stimulus number (mean of first three vowels vs. fourth vowel), local deviance (local deviant vs. local standard), global deviance (global deviant vs. global standard), electrode location (anterior positive vs. posterior negative pole) and hemisphere (left vs. right) as within-subject factors. In this analysis, the only pertinent comparisons were the interactions of local/global deviance with stimulus number and electrode location (\pm hemisphere). We considered only interactions with electrode location (and not the main effect over all locations)

because the electrode clusters were located on the positive and negative poles of the dipolar fields evoked by the deviant stimuli, and thus deviancy was expected to yield more positive voltages at one location and more negative voltages at the other, i.e. a deviancy \times location interaction.

The late slow wave belongs to a set of late responses observed after 600 ms in infants. In previous auditory studies a bilateral frontal negative wave, without a clear dipolar configuration, was reported from 700 to more than 1000 ms after the deviant stimulus (e.g., Dehaene-Lambertz & Dehaene, 1994; Friederici et al., 2002). We thus selected a cluster of 10 frontal electrodes for this analysis (Fig. 5C) on the 900–1200 ms time window, where the maximum of the response was recorded. An ANOVA was performed on the mean ERPs across this time window with local deviance (local deviant vs. local standard) and global deviance (global deviant vs. global standard) as within-subject factors. Note that, since the duration between sounds (SOA) within a trial sequence was 600 ms, the stimulus number could not be included as a factor in this ANOVA. We thus examined for this analysis only the main effects of local and global deviance.

2.6. Cortical sources modeling

Although ERPs have a coarse spatial resolution, it is possible to approximately infer cortical sources from the scalp voltage even in infants (Dehaene-Lambertz & Dehaene, 1994; Richards, 2005). Using a two-month-old infant template (Kabdebon et al., submitted for publication), we modeled the sources of the grand-averages of the local and global effects using Brainstorm, a matlab software package (Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011). We used a distributed model of 14,000 current dipoles, whose locations were constrained to the cortex of the infant template. The EEG forward model was computed using overlapping spheres whose radiuses were adapted to the infant head, skull and brain size. Conductivities were also modified to take into account infants' specific tissue properties. Cortical current maps were computed from the EEG time series using a linear inverse estimator (weighted minimum-norm current estimate).

3. Results

3.1. A two-stage response to auditory deviance

Our first analysis of the local and global deviance, based on permutations of these labels within-subjects, isolated two significant time periods for each effect (Fig. 3). For the local deviance (xxxY vs. xxxX), a large significant difference was seen at the classical latency of the infants' MMR ($p < .05$ from 180 to 500 ms), followed by a later difference ($p < .05$ from 992 to 1140 ms). The topography of the local difference during the early time-window was characterized by a frontal positivity synchronous of a posterior negativity whereas the late response corresponded to a frontal negativity with two posterior positive poles. These responses are congruent with the MMR/late frontal negative response described in attentive infants when a novel

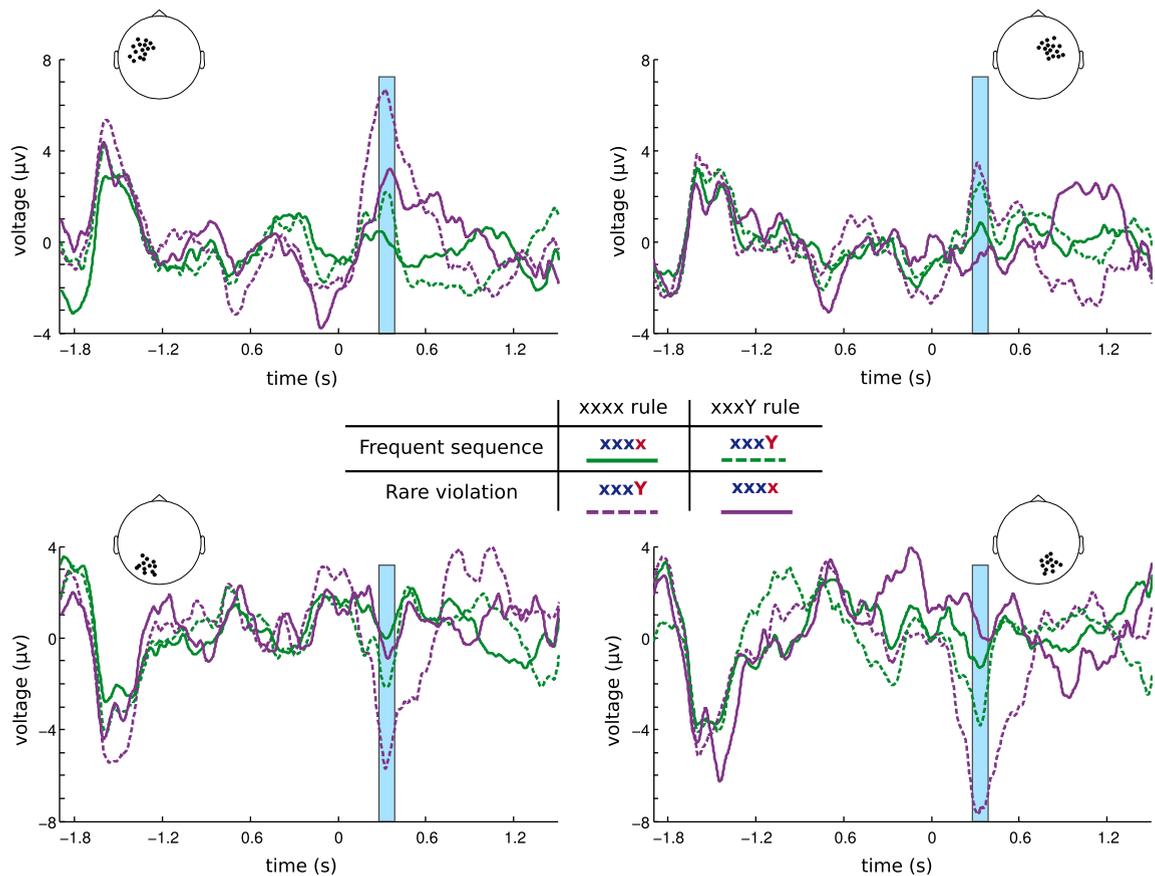


Fig. 4. Grand-average event-related potentials averaged across the electrodes displayed in each panel. Time 0 marks the onset of the fourth sound. Dashed vs. solid lines indicate ERPs to local deviants vs. local standards, while purple vs. green lines indicate ERPs to global deviants (rare rule violations) vs. global standards (frequent rule-governed stimuli). Within each condition, a baseline corresponding to the mean ERP during the presentation of the first three sounds (i.e. time window from -1800 to 0 ms) was subtracted from the corresponding ERPs. A mismatch response is visible on all selected clusters around 300 ms after the onset of the local deviant sound (dashed vs. solid lines), whether it is rare (in *xxxx* rule blocks) or frequent (in *xxxY* rule blocks). However, the amplitude of this response is reduced in *xxxY* rule blocks (green vs. purple dashed lines). In the same time window, a response to rare violations is mainly seen over the left anterior electrodes (purple vs. green lines in the top left panel). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

auditory stimulus is introduced after repeated stimuli (Dehaene-Lambertz & Dehaene, 1994; Friederici et al., 2002). For the global deviance (violation of the block structure), a two-stage response was also observed with a shorter early mismatch response (significant from 252 to 452 ms) and a longer late response (884 – 1140 ms) relative to the local effect. The topographies appeared more left lateralized than for the local deviance.

To analyze these effects topographically while avoiding double dipping problems (Vul & Pashler, 2012), we based our following analyses on the time-windows and sets of electrodes previously used in the infants' literature.

3.2. Early time-window: mismatch response

A bilateral mismatch response was observed when the trial was locally deviant (*xxxY*). Four clusters of electrodes selected from the literature captured quite well the topography of the response. Fig. 4 shows the time course of ERPs for each of the four types of trials over these clusters of electrodes, while Fig. 5A shows their

match to the full topography of the local effect. The ANOVA computed across the time window 270 – 370 ms reported a significant local effect (local deviance \times stimulus number \times electrode location: $F(1,28) = 21.13$, $p < 0.001$) which did not interact with hemisphere (local deviance \times hemisphere \times stimulus number \times electrode location: $F(1,28) = 1.02$). This response was considerably larger when the *xxxY* trials were rare (in *xxxx* blocks) than when they were frequent (in *xxxY* blocks) as shown by a significant interaction of local and global deviance (Figs. 5 and 6, global deviance \times local deviance \times stimulus number \times electrode location: $F(1,28) = 4.47$, $p < 0.05$).

We also observed a global effect in this early time-window (global deviance \times stimulus number \times electrode location: $F(1,28) = 8.81$, $p < 0.01$, Fig. 5B). This response was significantly different over the left and right hemisphere (4-way interaction of global deviance \times hemisphere \times stimulus number \times electrode location, $F(1,28) = 4.36$, $p < 0.05$). Post hoc analyses restricted to anterior electrodes revealed a significant hemispheric difference for the global effect (global deviance \times hemisphere

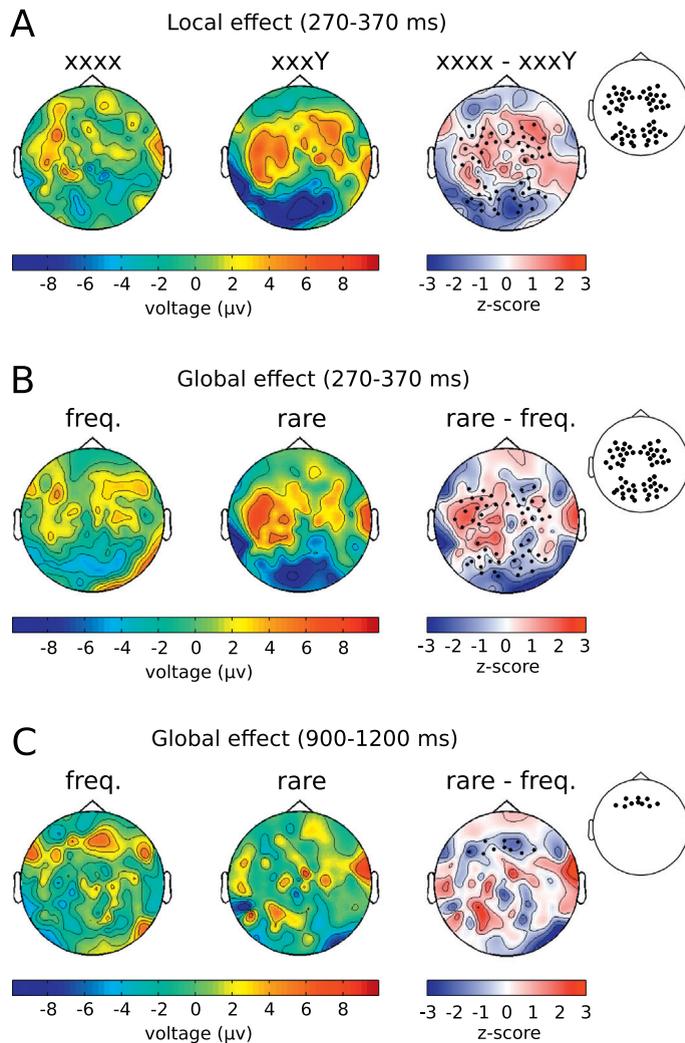


Fig. 5. Topographies of the local and global effects. Each panel shows the projection, on a 2-D representation of the head surface, of the average voltage measured in a time window of interest in the two compared experimental conditions, as well as the corresponding z-scores. Electrodes selected for further statistical analysis are shown at right. (A) Local effect. Averaged ERPs across the time window 270–320 ms after the onset of the fourth sound for local standard (xxxx) and local deviant (xxxY) sequences, and the z-score of the mismatch response (xxxY–xxxx). A bilateral anterior positivity and posterior negativity are evoked by local deviant sounds. (B and C) Global effect. Averaged ERPs across the time window 270–320 ms (in B) and 900–1200 ms (in C) after the onset of the fourth sound in global standard (frequent) and global deviant (rare) sequences, and the z-score of the difference (rare–frequent). A left-lateralized positivity over anterior electrodes, followed by a late negativity over frontal electrodes, are seen for rare violations relative to the rule-governed sequences.

\times stimulus number, $F(1,28) = 6.57$, $p < 0.05$). While the local effect was significant over both left and right anterior electrodes (local deviance \times stimulus number: respectively, $F(1,28) = 6.79$, $p < 0.05$ and $F(1,28) = 8.45$, $p < 0.01$), the global effect resulted in a significant positivity only over the anterior left hemisphere (global deviance \times stimulus number: left $F(1,28) = 18.03$, $p < 0.0001$, right $F(1,28) < 1$).

3.3. Late time-window: frontal negative slow wave

We then examined the presence of a late response to auditory deviance over frontal electrodes and the 900–1200 ms time window (Fig. 5C, Fig. 6B and C). The two-way ANOVA showed a significant global effect

($F(1,28) = 5.75$, $p < 0.05$). A significant interaction of global deviance \times local deviance was observed ($F(1,28) = 4.29$, $p < 0.05$). At this moment and on this cluster, the local effect itself was not significant ($F(1,28) < 1$).

3.4. A global novelty response to xxxx trials

A notable feature of the response to global deviance is that the brain should generate a novelty response to a monotonous xxxx stimulus sequence when this sequence is presented within xxxY blocks and is therefore rare and unexpected. In order to test whether this effect was present in infants, we compared the response to the same xxxx trials in the blocks following the xxxx rule and those following the xxxY rule. A global deviance effect was indeed

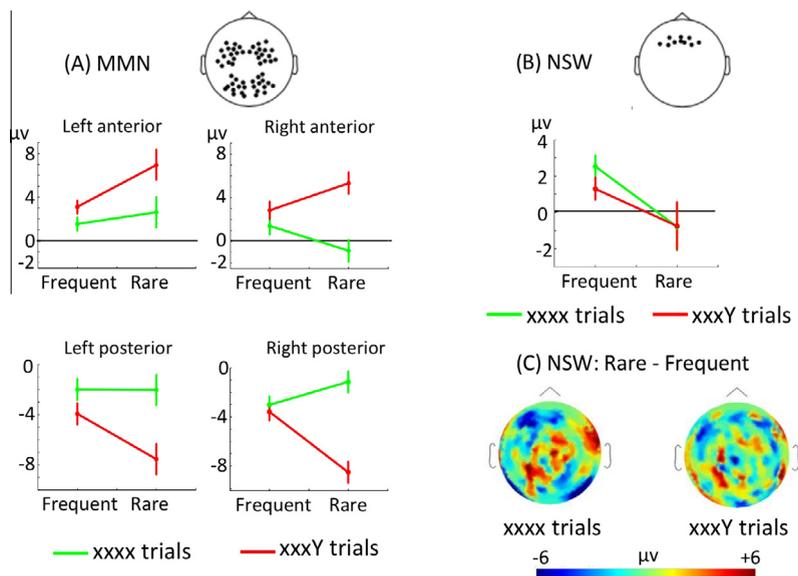


Fig. 6. Grand average of the voltage across the time-windows and clusters of electrodes selected for the MMR (A) and the NSW (B). For the MMR (A), the response was larger for xxxY trials relative to xxxx trials (local effect), both for frequent sequences (global standards) and for rare sequences (global deviants). In addition, over the left anterior cluster, global deviant trials had a higher amplitude than global standard trials ($p < .0001$). This was not the case over the right anterior cluster, leading to a significant interaction of global deviance with hemisphere ($p < .05$). For the NSW (B), a larger negativity was recorded for rare sequences relative to frequent sequences (global effect). The topographies of this global effect (rare–frequent) are presented below, separately for xxxx and xxxY sequences.

observed at both time-windows: at 270–370 ms on the left anterior electrodes (global deviance \times stimulus number: $F(1,28) = 5.58$, $p < 0.05$) and at 900–1200 ms on frontal electrodes (mean of difference = -6.67 μV , $t = -3.08$, $p < 0.01$). This is particularly visible in Fig. 6B for the NSW.

The preceding analysis compared xxxx trials coming from different experimental blocks. As a control, we checked whether these blocks already differed prior to the fourth sound. We compared brain responses to the first three sounds in xxxx and xxxY blocks and in the learning and test phases during the early time-window (i.e. 270–370 ms after the onset of the sounds) on the left anterior electrodes which showed a global effect. We thus performed a repeated-measure ANOVA with stimulus number (first, second or third sound), phase (learning vs. test) and rule (xxxx vs. xxxY) as within-subject factors. The ANOVA showed a significant effect of stimulus number ($F(1,28) = 31.14$, $p < 0.0001$) but no effect of experimental phase ($F(1,28) = 2.23$) or rule type ($F(1,28) = 2.96$) was observed, nor any interaction of these factors. Post hoc analyses restricted to the first and second sounds and to the second and third sounds revealed a significantly larger response to the first than to the second sound ($F(1,28) = 34.34$, $p < 0.0001$) while the response to the second and the third sounds was not significantly different ($F(1,28) < 1$), as already described in Dehaene-Lambertz and Dehaene (1994).

3.5. Source localization

Although our main goal was to determine the temporal dynamics and functional properties of auditory novelty responses in infants, our high-density recordings allow some tentative inferences about the underlying cortical

generators. Fig. 7 displays the difference in source currents associated with the local and global effects in the early and late time-windows studied above. Bilateral temporal, inferior parietal and ventral and dorsal prefrontal cortices are involved in the early local and global effect (i.e. 270–370 ms after the onset of the 4th sound). During the late time window of the global effect (900–1200 ms after the onset of the 4th sound), activity remained in a large network of distributed areas mainly in the left hemisphere involving notably the left inferior frontal region. For illustration, Fig. 7 shows the reconstructed temporal profile of activations in two exemplary regions, the left superior temporal (primarily showing repetition suppression and the local effect) and inferior frontal regions (primarily showing the global effect).

4. Discussion

In agreement with earlier studies using an auditory oddball paradigm (Dehaene-Lambertz & Dehaene, 1994; Friederici et al., 2002), we observed that the detection of a change in a series of repeated sounds induces two successive brain responses in awake three-month-old infants: a mismatch response followed by a late frontal negativity. By using the local–global paradigm (Bekinschtein et al., 2009), we now demonstrate that these responses correspond to partially dissociable processing stages, related respectively to the detection of local probability changes vs. violations of a global sequence.

4.1. An early mismatch response to local violations

We recorded a classical early mismatch response (around 300 ms) when a novel vowel was introduced

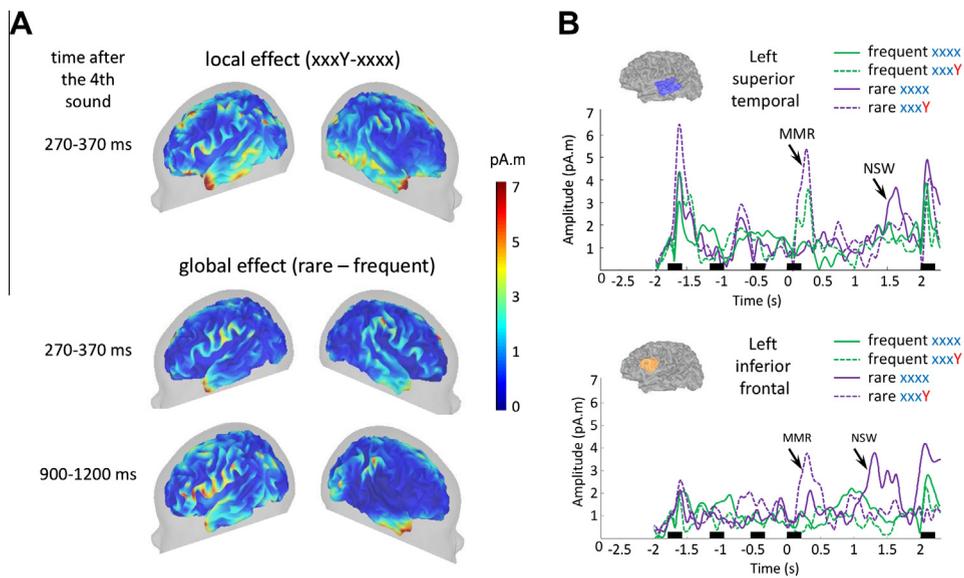


Fig. 7. Approximate reconstruction of cortical sources for the local and global effects, as computed from the grand average. (A) In each panel, the colors, projected on an infant's left- and right-hemispheric cortical surfaces, code for the difference in source currents associated with the experimental comparisons depicted in Fig. 5. Bilateral temporal, inferior parietal and ventral and dorsal prefrontal cortices are involved in the early local and global effect (i.e. 270–370 ms after the onset of the 4th sound). Left inferior frontal areas are involved in the late global effect (900–1200 ms after the onset of the 4th sound). (B) Time courses represent the mean activity in two representative regions of interest: left inferior frontal cortex and left superior temporal cortex. Two temporally distinct responses to violations are clearly visible in the time windows where a mismatch response (MMR) and a Negative Slow Wave (NSW) are seen on the scalp (see Fig. 5). The left superior temporal cortex shows primarily an early MMR to local deviance (xxxY relative to xxxx trials). The left frontal region is involved in the detection of global deviants, with an earlier response when the rare sequence is xxxY than when it is xxxx.

following a series of repeated vowels (xxxY trials). As is well known from earlier research (e.g. Dehaene-Lambertz & Dehaene, 1994; Friederici et al., 2002; Stefanics et al., 2007), this effect, analogous to the adult MMN, indicates that the infant brain is sensitive to elementary statistical regularities, such as the fact that one sound is presented more frequently than the other.

There is a debate as to whether the adult MMN can be explained solely by a passive process of adaptation to repetition (May & Tiitinen, 2010) or implies an active mechanism of prediction and/or comparison of the past with the present (Näätänen et al., 2007; Wacongne et al., 2012; Winkler, 2007). According to predictive coding models, the amplitude of the auditory response is determined by the difference between the current incoming auditory input and its prediction based on the temporal contingencies experienced in the recent past (Bendixen, SanMiguel, & Schröger, 2012; Bendixen, Schröger, & Winkler, 2009; Friston, 2005; Wacongne et al., 2012; Winkler & Czigler, 2012). Wacongne et al. (2012) proposed a neuronal implementation of such a predictive model where the mismatch response is generated by “prediction errors” neurons which sum excitatory input from the thalamus and inhibitory input from a population of predictive neurons in supra-granular layers.

In adults, the predictive coding hypothesis is supported by the observation of an MMN to omitted sounds in a patterned sequence (Bendixen et al., 2009; Wacongne et al., 2012), or to rare repeated sounds (AA) amidst frequent different ones (AB) (Horváth & Winkler, 2004; Wacongne et al., 2011). Both of these findings are difficult to explain

by a purely passive adaptation process. In infants, although fewer data is available, He, Hotson, and Trainor (2009) also showed that 4-month-old infants display a mismatch response to rare repeated pairs (AA) amidst frequent different ones (AB), suggesting that a similar activation predictive mechanism may already be present at four month of age (evidence was ambiguous in younger infants).

4.2. The early response is modulated by global context

Our three-month-olds' local mismatch response was larger on xxxx blocks, where xxxY deviants were rare, than on xxxY blocks, when the deviants were frequent and largely predictable. Similarly, the response to xxxx trials was larger when these trials were rare (in xxxY blocks), than when they were frequent (in xxxx blocks). As a result, the local effect (difference between xxxY and xxxx trials) was larger when these sequences were rare than when they were frequent (Fig. 6). These effects also induced a significant difference between rare and frequent segments (early global effect), mostly visible on the left anterior channels.

Although these findings adds complexity to the simple two-stage picture of the local–global design, with an early MMR to local violations and a late SNW to global violations (Bekinschtein et al., 2009), a modulation of the early MMR according to context is not surprising given that the structure of the blocks affects sound transition probabilities. In adults too, although the first publication of the local–global design did not detect any interaction of the two effects (Bekinschtein et al., 2009), a more recent multivariate

pattern analysis of adult data demonstrated such an effect as early as 150 ms in attentive subjects (King et al., 2013).

There are several possible interpretations of this early global difference. One interpretation that we can exclude is a stronger attention to one type of block than to the other. All audio–visual blocks comprised face movements and speech stimuli attractive for infants of this age and an analysis performed on the first three sounds of trials showed no difference between blocks. Furthermore, because our rare vs. frequent comparison is orthogonal to the block factor, a difference in attention between blocks would not be sufficient to explain this early global difference.

A more likely interpretation is that the amplitude of the auditory response is modulated by the frequency of occurrence of sounds and their transitions, not only on a given trial, but accumulated from one trial to the next as implemented in a recent model of the adult MMN (Wacongne et al., 2012). It has already been shown that the mismatch response is modulated by deviant frequency in adults: its amplitude is increasingly larger as the event frequency is low (Matuoka et al., 2006; Sato et al., 2000). It is also elicited after a smaller number of repetitions when the repeated sound is kept constant across trials than when it changes from one trial to the next (Winkler et al., 1996). Here by necessity, the proportion of Y vowels was higher in xxxY blocks than in xxxx blocks (18.75% vs. 6.25% during the test phase). Thus, as in adults, the infant mismatch response to a local change appears to be modulated by the block structure, being larger on xxxx blocks where the Y sound is less frequent.

What about xxxx trials? Even there, a similar explanation may hold. During the test phase, the occurrence of x stimuli was 93.75% in the blocks following the xxxx rule, vs. only 81.25% in the xxxY blocks. Furthermore, in terms of transition probabilities, the x–x transition occurred on 91.7% of test trials in xxxx blocks, vs. 75% in xxxY blocks. Here again, a modulation of early auditory responses by the frequency of occurrence of sounds and/or their transitions, and thus by local sound predictability, can explain the difference between frequent and rare xxxx segments. This interpretation suggests that a record of event frequencies, or conditional transition probabilities, is accumulated from one trial to the next, being only limited by the decay of sound representations in auditory memory.

We cannot exclude, however, another interpretation, according to which even global rules impact on early responses. The two rules that we tested, a strict repetition in xxxx blocks and a final deviant sound in xxxY blocks, are simple enough that they might be partially learned at an early auditory level (Bendixen et al., 2012), attenuating the auditory response by a higher-order “global” expectation of the overall sequence pattern. Two experiments in infants suggest that some simple regular sequences may indeed modulate responses in the auditory cortex. Stefanics et al. (2007) recorded a weaker negativity in neonates between 150 and 250 ms in a regular sequence (xxxxY) relative to a random sequence of the same x and Y tones, suggesting an amplitude modulation of the early response by sound predictability (see also He et al., 2009). Carral et al. (2005) observed a mismatch response

in neonates to changes in the direction of pitch change between two tones in a pair (e.g. a deviant descending-pitch pair within a series of ascending-pitch pairs) independently of their absolute frequency. These experiments emphasize that even in infants, prediction of the following sound is an early-functioning learning mechanism.

In adults, the early global effect was only observed in attentive adults (King et al., 2013). Here, our infants were awake and attentive, as they were strongly attracted to the speaking faces, as in He et al.’s study (2009), but they were asleep in Stefanics et al.’s (2007), and in Carral et al.’s (2005) studies. Further studies are needed to specify what type of auditory patterns may or may not be learned in inattentive or even sleeping infants. Ascending and descending pairs (Carral et al., 2005) and regular vs. random presentation (Stefanics et al., 2007) might be considered as *gestalts* immediately identified at the auditory level (Endress, Dehaene-Lambertz, & Mehler, 2007), whereas more complex patterns, such as the repetition of xxxY segments, might need an attentive participant before being internalized at the auditory level.

4.3. Left–right hemispheric differences

Spatial differences in electrophysiological recordings must be treated cautiously, as event-related potentials are limited in their spatial resolution and different cortical sources may mix on the scalp, making it difficult to infer cortical sources from surface recordings. Reconstruction of cortical sources should also be interpreted with caution, given the anatomical asymmetries present in the human brain from the fetal age on (Glaser et al., 2011), the lack of individual MRI in our data-set, and the inevitable reduction in accuracy incurred when modeling the group average rather than individual subjects. However, the distinct difference in amplitude modulation observed over the left and right anterior channels for the early local and global effect might suggest that the two hemispheres do not react to exactly the same information. Numerous studies have shown that the left and right hemisphere follow a different developmental time-course from the preterm period on, with numerous regions of the left hemisphere slightly delayed relative to the right. Sulci appear generally earlier in the right than in the left hemisphere (Dubois et al., 2008), cerebral blood flow is larger in the right than in the left hemisphere during infancy (Chiron et al., 1997; Lin et al., 2012) and the superior temporal region show indices of faster maturation during the first months of life (Leroy et al., 2011). Furthermore, processing differences are also observed early on: processing of fast temporal transitions is preferentially performed in the left hemisphere whereas spectral information is preferentially channeled to the right hemisphere already during infancy (Bristow et al., 2009; Dehaene-Lambertz et al., 2010; Mahmoudzadeh et al., 2013; Perani et al., 2010; Telkemeyer et al., 2011). The present experiment provides yet another observation of the different functional properties displayed very early on by the left and right hemispheres in human infants. It might suggest that left and right auditory buffers might be of different size, allowing the left hemisphere to keep track of a sound during a

longer time-window, or that top-down modulation from higher-levels regions are stronger on the left than on the right hemisphere: Whereas maturation indices are delayed in the left temporal region relative to the right, the reverse pattern is observed in Broca's area (Leroy et al., 2011) and in the arcuate fasciculus (Dubois et al., 2009), suggesting a stronger short-term auditory loop in the left hemisphere than in the right (Dehaene-Lambertz et al., 2006; Leroy et al., 2011) from the first months of life on.

4.4. A Negative Slow Wave (NSW) response to global violations

During the late time-window (900–1200 ms), we observed a specific response to violations of the global sequence pattern. A bilateral frontal negativity was observed only for the global, but not local, deviance. A similar late response has been previously reported by Dehaene-Lambertz and Dehaene (1994) and Friederici et al. (2002) in awake infants after a deviant sound. This response was similar in topography and latency to the negative slow wave evoked by deviant visual events in visual oddball experiments, and which is interpreted either as an orientation response to novelty, or as a generalized arousal effect (see Nelson, 1994, and for review, Csibra et al., 2008). Taken together, these results therefore suggest that, similarly to the adult P300 response, the infant's negative slow wave indexes a multimodal inferential system capable of detecting cross-modal regularities (Ritter, Sussman, Deacon, Cowan, & Vaughan, 1999) and that infants not only encode the present incoming stimulus, but are also able to place it in a more global and abstract context.

Our auditory–visual paradigm was designed to maximally engage the infant. At this age when infants learn the vowel repertoire of their native language and to associate mouth movement and speech sounds (Kuhl & Meltzoff, 1982; Patterson & Werker, 2003), talking faces are strong attractors for infants. Thus even in the case of repeated xxxY, the change of vowel might still have grabbed infants' attention, but rare xxxY among frequent xxxx were obviously noticeable given the large auditory and visual difference between /a/ and /i/. This huge effect explains that our first analysis also uncovered a late local effect during the time window of the NSW. Most important, however, is the observation that the mere repetition of the same vowel, when it was unexpected (rare xxxx among frequent xxxY) elicited a long-lasting NSW (around a second). That NSW can only be explained by the novelty of the overall sequence, signaling an orientation of infants' attention due to the violation of their expectation.

Detecting that a monotonic sequence xxxx is deviant suggests a hierarchical novelty-detection process: as shown by the early mismatch response, the infant brain reacts to deviant final sounds (xxxY), but it is also capable of detecting the absence of such a local deviant in the xxxx sequence, which, in turn, signals that this sequence is in fact novel. The existence of a brain response to rare xxxx trials has therefore been interpreted as fitting with the hierarchical Bayesian framework (Wacongne et al., 2011), according to which the brain generates predictive signals

at multiple hierarchical levels, and event-related response reflect a sequence of error responses that index the difference between the predicted and actual signals (Friston, 2005; Rao & Ballard, 1999). Like adults', infants' brain responses to auditory violations would be separated into two stages: the first one mainly driven by the local probability of occurrence of a sound, and the second sensitive to the overall regularity of the global sequence of sounds relative to past sequences. Hierarchical Bayesian inference would thus already be available to infants, as also suggested by other predictive and surprise paradigms (e.g. Gopnik & Schulz, 2004; Gweon, Tenenbaum, & Schulz, 2010; Tenenbaum et al., 2011; Téglás et al., 2011).

Although source localization is only indicative in infants, it ascribed the late global effect to a prolonged activity in a left inferior frontal source over Broca's area whereas the sources of the local and global mismatch response were predominantly assigned to an early phasic activation in the superior temporal cortex (Fig. 7). Although speculative at this time, it is interesting to underline potential similarities between our infant sources and the adult sources reported using MEG by several teams using similar auditory paradigms (Halgren et al., 2011; Wacongne et al., 2011). These authors reported a sequence of activations, first confined to temporal areas, then involving a broader network expanding notably into frontal areas around 250 ms. Once again, infants appear to show a similar sequence, only at a much slower pace, compatible with their reduced synaptic connectivity and weak myelination (see also Kouider et al., 2013 for another delayed response signaling higher-level processing). The involvement of a similar cortical network in infants and adults should obviously be addressed in future studies using more precise localization methods such as fMRI.

5. Limits, implications and perspectives

The present experimental design leaves open the degree of abstractness of the global sequence pattern that infants detected, a point that will require further experimentation. Infants may be memorizing the recurrence of a particular sequence of vowels in a given 5-min block (e.g. a-a-a-i). Indeed, a memory for specific words (Benavides-Varela et al., 2011; Shi, Werker, & Morgan, 1999) and sentences (Dehaene-Lambertz et al., 2006) is known to be available in the first few months of life. Alternatively, they may be abstracting away from the individual phonetic items and storing a more abstract rule such as xxxY (“3 identical items are followed by a different one”) or yet more abstract rules (e.g. “the last item is different”).

The ability of older infants to quickly infer abstract regularities after a brief exposure to sequences of auditory or visual stimuli has been vastly studied (review in Aslin & Newport, 2012; Johnson et al., 2009; Kirkham, Slemmer, Richardson, & Johnson, 2007; Marcus, Fernandes, & Johnson, 2007; Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Saffran, Pollak, Seibel, & Shkolnik, 2007; Saffran & Thiessen, 2003). In a classical study, Marcus et al. (1999) showed that 7-month-old infants which were familiarized to rules like ABA (e.g. ga ti ga, po tu po) or ABB (e.g. ga ti ti,

po tu tu) were able to extract the abstract rule and generalize it to new syllables, thus detecting sequences that violated the learned rule. Using the same type of paradigm, Frank, Slemmer, Marcus, and Johnson (2009) observed that 5-month-old infants could learn ABA or AAB rules when they were presented in an audio–visual modality: a syllable was accompanied by a colored looming shape (e.g. “ba-octagon de-square ba-octagon” for ABA rule). However, no significant effect of abstract rule extraction was observed when the rule was presented in the auditory or the visual modalities alone. Similarly, Gervain, Macagno, Cogoï, Peña, and Mehler (2008), using near-infrared spectroscopy, observed that neonates were sensitive to the repetitive structure ABB of auditory pseudowords relative to a random structure ABC, but did not differentiate the ABA structure from the random ABC, suggesting sensitivity to adjacent repetitions only. To our knowledge, there is no other report in the literature on the rule-learning abilities of younger infants. In the future, the present paradigm could be used to determine whether three-month-old infants can already detect an abstract rule, such as “the last item is different”, for instance by varying the items during the learning phase (e.g. mixing a-a-a-i and i-i-i-a) and testing for brain responses to rule-respecting and rule-violating sequences with a variable degree of generalization to novel items (see Fitch & Friederici, 2012).

Besides its relevance for temporal-sequence learning, the local–global test was first introduced as a probe of conscious-level processing in adult brain-lesioned patients with vegetative state and related disorders (Bekinschtein et al., 2009). The rationale was that (1) the identification, over multiple trials, of a recurring global sequence such as xxxY requires a temporal scope which is beyond the typical exponential decay of unconscious processes, and requires an explicit maintenance in conscious working memory (Dehaene & Naccache, 2001); and (2) global deviants elicit a late P3b electrophysiological response which, in many paradigms contrasting conscious vs. non-conscious stimuli, has been repeatedly demonstrated to correlate with conscious-level processing in healthy subjects (Donchin & Coles, 1988; Sergent et al., 2005; review in Dehaene & Changeux, 2011). Subsequently, empirical studies using the local–global paradigm in adult brain-lesioned patients with coma, vegetative state or minimal consciousness indeed demonstrated that the presence of the global P3b response almost invariably indexes a conscious subject (Bekinschtein et al., 2009; Faugeras et al., 2011), while the earlier local response (mismatch negativity) remains present in many comatose subjects (Fischer, Luauté, Adeleine, & Morlet, 2004; Fischer, Luauté, & Morlet, 2010).

Following this logic, the unambiguous finding that the brain of three-month-old infants responds to global auditory violations suggests the tentative conclusion that an elementary form of conscious-level processing may already be operative in three-month-old infants. This conclusion fits with our recent observation of a tight parallel between infants’ and adults’ event-related responses to visible and invisible masked faces, also suggesting that a late frontal negative response characterizes conscious access to visual stimuli in infants (Kouider et al., 2013). Lagercrantz and Changeux (2009, 2010) review the

considerable, though still highly fragmentary, pediatric and neuroimaging evidence suggesting that even newborns may already present integrative responses to pain, olfactory, visual and auditory stimulation. Long-distance bi-hemispheric cortical networks thought to support a “conscious global workspace” (Dehaene, Kerszberg, & Changeux, 1998; Dehaene & Naccache, 2001) are already anatomically present (Dubois et al., 2008; Takahashi, Folkner, Galaburda, & Grant, 2012) and may already be functionally operative in the first few months of life (Dehaene-Lambertz et al., 2002; Fransson et al., 2007; Kouider et al., 2013).

At present, any conclusion on this topic must remain speculative, as we cannot exclude that specific mechanisms of rule extraction may be operative non-consciously in three-month-old infants, but may require consciousness at a later age. Extensions of the present paradigm, comparing responses arising from various levels of cortical integration in infants and adults with variable degrees of residual consciousness, could be useful in order to establish more unambiguous behavioral and electrophysiological signatures of the level of consciousness during infancy.

6. Conclusion

The present study provides evidence that hierarchical brain mechanisms underlying temporal sequence learning are already present at three months of life. At least two distinct cortical systems, one sensitive to local probabilities and the other to global regularities, are present in young infants. As in adults, violations of local and global regularities affect distinct time windows: only global, but not local violations, elicit a late anterior negative slow wave over bilateral frontal areas. In that respect, the present results fit with previous brain imaging studies demonstrating that frontal regions are already active in three-month-old infants (Bristow et al., 2009; Dehaene-Lambertz et al., 2002). To our knowledge, the present study provides the earliest evidence so far that young infants can process multiple levels of statistical violations, and supports the emerging Bayesian framework (Tenenbaum et al., 2011) according to which hierarchical learning mechanisms may be operative early on during development.

Acknowledgments

This research was supported by European Research Council (NeuroConsc Program), CEA (Commissariat à l’Energie Atomique et aux Energies Alternatives) and INSERM (Institut National de la Santé et de la Recherche Médicale). We are particularly grateful to Giovanna Santoro and Gaëlle Mediouni for their help in baby recruitment and testing.

References

- Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, 21(3), 170–176.
- Atienna, M., Cantero, J. L., & Gomez, C. M. (1997). The mismatch negativity component reveals the sensory memory during REM sleep in humans. *Neuroscience Letters*, 237(1), 21–24.

- Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences of the United States of America*, 106(5), 1672–1677.
- Benavides-Varela, S., Gómez, D., Macagno, F., Bion, R., Peretz, I., & Mehler, J. (2011). Memory in the neonate brain. *PLoS One*, 6(11), e27497.
- Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: A review. *International Journal of Psychophysiology*, 83(2), 120–131.
- Bendixen, A., Schröger, E., & Winkler, I. (2009). I heard that coming: Event-related potential evidence for stimulus-driven prediction in the auditory system. *The Journal of Neuroscience*, 29(26), 8447–8451.
- Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., et al. (2009). Hearing faces: How the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience*, 21(5), 905–921.
- Carral, V., Huottilainen, M., Ruusuvirta, T., Fellman, V., Näätänen, R., & Escera, C. (2005). A kind of auditory “primitive intelligence” already present at birth. *European Journal of Neuroscience*, 21(11), 3201–3204.
- Celsis, P., Boulanouar, K., Doyon, B., Ranjeva, J. P., Berry, I., & Chollet, F. (1999). Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *NeuroImage*, 9, 135–144.
- Cheour, M., Čeponienė, R., Leppänen, P., Alho, K., Kujala, T., Renlund, M., et al. (2002). The auditory sensory memory trace decays rapidly in newborns. *Scandinavian Journal of Psychology*, 43(1), 33–39.
- Chiron, C., Jambaque, I., Nabbout, R., Lounes, R., Syrota, A., & Dulac, O. (1997). The right brain hemisphere is dominant in human infants. *Brain*, 120, 1057–1065.
- 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, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200–227.
- Dehaene, S., Kerszberg, M., & Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences of the United States of America*, 95(24), 14529–14534.
- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79(1–2), 1–37.
- Dehaene-Lambertz, G., & Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature*, 370(6487), 292–295.
- Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, 298(5600), 2013–2015.
- Dehaene-Lambertz, G., & Gliga, T. (2004). Common neural basis for phoneme processing in infants and adults. *Journal of Cognitive Neuroscience*, 16(8), 1375–1387.
- Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Mériaux, S., Roche, A., Sigman, M., et al. (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.
- Dehaene-Lambertz, G., Montavont, A., Jobert, A., Alliol, 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.
- Dehaene-Lambertz, G., & Peña, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport*, 12(14), 3155–3158.
- Donchin, E., & Coles, M. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, 11(3), 357–374.
- Dubois, J., Benders, M., Cachia, A., Lazeyras, F., Ha-Vinh Leuchter, R., Sizonenko, S. V., et al. (2008). Mapping the early cortical folding process in the preterm newborn brain. *Cerebral Cortex*, 18(6), 1444–1454.
- Dubois, J., Hertz-Pannier, L., Cachia, A., Mangin, J.-F., Le Bihan, D., & Dehaene-Lambertz, G. (2009). Structural asymmetries in the infant language and sensori-motor networks. *Cerebral Cortex*, 19(2), 414–423.
- Endress, A. D., Dehaene-Lambertz, G., & Mehler, J. (2007). Perceptual constraints and the learnability of simple grammars. *Cognition*, 105(3), 577–614.
- Faugeras, F., Rohaut, B., Weiss, N., Bekinschtein, T. A., Galanaud, D., Puybasset, L., et al. (2011). Probing consciousness with event-related potentials in the vegetative state. *Neurology*, 77(3), 264–268.
- Fischer, C., Luauté, J., Adeleine, P., & Morlet, D. (2004). Predictive value of sensory and cognitive evoked potentials for awakening from coma. *Neurology*, 63(4), 669–673.
- Fischer, C., Luauté, J., & Morlet, D. (2010). Event-related potentials (MMN and novelty P3) in permanent vegetative or minimally conscious states. *Clinical Neurophysiology*, 121(7), 1032–1042.
- Fischer, C., Morlet, D., Bouchet, P., Luauté, J., Jourdan, C., & Salord, F. (1999). Mismatch negativity and late auditory evoked potentials in comatose patients. *Clinical Neurophysiology*, 110(9), 1601–1610.
- Fitch, W., & Friederici, A. D. (2012). Artificial grammar learning meets formal language theory: An overview. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1598), 1933–1955.
- Frank, M. C., Slemmer, J. A., Marcus, G. F., & Johnson, S. P. (2009). Information from multiple modalities helps 5-month-olds learn abstract rules. *Developmental Science*, 12(4), 504–509.
- Fransson, P., Skiöld, B., Horsch, S., Nordell, A., Blennow, M., Lagercrantz, H., et al. (2007). Resting-state networks in the infant brain. *Proceedings of the National Academy of Sciences of the United States of America*, 104(39), 15531–15536.
- Friederici, A. D., Friedrich, M., & Weber, C. (2002). Neural manifestation of cognitive and precognitive mismatch detection in early infancy. *Neuroreport*, 13(10), 1251–1254.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836.
- Garrido, M. I., Friston, K. J., Kiebel, S. J., Stephan, K. E., Baldeweg, T., & Kilner, J. M. (2008). The functional anatomy of the MMN: A DCM study of the roving paradigm. *NeuroImage*, 42(2), 936–944.
- Garrido, M. I., Kilner, J. M., Kiebel, S. J., & Friston, K. J. (2007). Evoked brain responses are generated by feedback loops. *Proceedings of the National Academy of Sciences of the United States of America*, 104(52), 20961–20966.
- Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences of the United States of America*, 105(37), 14222–14227.
- Glaser, H., Leroy, F., Dubois, J., Hertz-Pannier, L., Mangin, J.-F., & Dehaene-Lambertz, G. (2011). A robust cerebral asymmetry in the infant brain: The rightward superior temporal sulcus. *NeuroImage*, 58(3), 716–723.
- Gopnik, A., & Schulz, L. (2004). Mechanisms of theory formation in young children. *Trends in Cognitive Sciences*, 8(8), 371–377.
- Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences of the United States of America*, 107(20), 9066–9071.
- Halgren, E., Sherfey, J., Irimia, A., Dale, A., & Marinkovic, K. (2011). Sequential temporo-fronto-temporal activation during monitoring of the auditory environment for temporal patterns. *Human Brain Mapping*, 32(8), 1260–1276.
- He, C., Hotson, L., & Trainor, L. (2009). Development of infant mismatch responses to auditory pattern changes between 2 and 4 months old. *European Journal of Neuroscience*, 29(4), 861–867.
- Horváth, J., & Winkler, I. (2004). How the human auditory system treats repetition amongst change. *Neuroscience Letters*, 368(2), 157–161.
- Johnson, S. P., Fernandes, K. J., Frank, M. C., Kirkham, N., Marcus, G., Rabagliati, H., et al. (2009). Abstract rule learning for visual sequences in 8- and 11-month-olds. *Infancy*, 14(1), 2–18.
- Kabdebon, C., Leroy, F., Simonnet, H., Perrot, M., Dubois, J., & Dehaene-Lambertz, G. (submitted for publication). Anatomical correlations of the international 10–20 sensor placement system in infants.
- King, J. R., Faugeras, F., Gramfort, A., Schurger, A., El Karoui, I., Sitt, J., et al. (2013). Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. *NeuroImage*, 83, 726–738.
- Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., & Johnson, S. P. (2007). Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development*, 78(5), 1559–1571.
- Kouider, S., Stahlhut, C., Gelskov, S. V., Barbosa, L. S., Dutat, M., de Gardelle, V., et al. (2013). A neural marker of perceptual consciousness in infants. *Science*, 340(6130), 376–380.
- Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, 218(4577), 1138–1141.
- Lagercrantz, H., & Changeux, J.-P. (2009). The emergence of human consciousness: From fetal to neonatal life. *Pediatric Research*, 65(3), 255–260.
- Lagercrantz, H., & Changeux, J.-P. (2010). Basic consciousness of the newborn. *Seminars in Perinatology*, 34(3), 201–206.
- Leroy, F., Glaser, H., Dubois, J., Hertz-Pannier, L., Thirion, B., Mangin, J.-F., et al. (2011). Early maturation of the linguistic dorsal pathway in human infants. *The Journal of Neuroscience*, 31(4), 1500–1506.

- Lin, P.-Y., Roche-Labarbe, N., Dehaes, M., Fenoglio, A., Grant, P. E., & Franceschini, M. A. (2012). Regional and hemispheric asymmetries of cerebral hemodynamic and oxygen metabolism in newborns. *Cerebral Cortex*. <http://dx.doi.org/10.1093/cercor/bhs023>.
- Mahmoudzadeh, M., Dehaene-Lambertz, G., Fournier, M., Kongolo, G., Goudjil, S., Dubois, J., et al. (2013). Syllabic discrimination in premature human infants prior to complete formation of cortical layers. *Proceedings of the National Academy of Sciences of the United States of America*, 110(12), 4846–4851.
- Mäntysalo, S., & Näätänen, R. (1987). The duration of a neuronal trace of an auditory stimulus as indicated by event-related potentials. *Biological Psychology*, 24(3), 183–195.
- Marcus, G., Fernandes, K., & Johnson, S. (2007). Infant rule learning facilitated by speech. *Psychological Science*, 18(5), 387–391.
- Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283(5398), 77–80.
- Matuoka, T., Yabe, H., Shinozaki, N., Sato, Y., Hiruma, T., Ren, A., et al. (2006). The development of memory trace depending on the number of the standard stimuli. *Clinical EEG and Neuroscience*, 37(3), 223–229.
- May, P., & Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, 47(1), 66–122.
- Näätänen, R., Gaillard, A., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, 42(4), 313–329.
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, 118(12), 2544–2590.
- Nelson, C. A. (1994). Neural correlates of recognition memory in the first postnatal year of life. In G. Dawson & K. Fischer (Eds.), *Human behavior and the developing brain* (pp. 269–313). New York: Guilford Press.
- Patterson, M. L., & Werker, J. F. (2003). Two-month old infants match phonetic information in lips and voice. *Developmental Science*, 6(2), 193–198.
- Pegado, F., Bekinschtein, T. A., Chausson, N., Dehaene, S., Cohen, L., & Naccache, L. (2010). Probing the lifetimes of auditory novelty detection processes. *Neuropsychologia*, 48(10), 3145–3154.
- Perani, D., Saccuman, M., Scifo, P., Spada, D., Andreolli, G., Rovelli, R., et al. (2010). Functional specializations for music processing in the human newborn brain. *Proceedings of the National Academy of Sciences of the United States of America*, 107(10), 4758–4763.
- Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.
- Quartz, S. R., & Sejnowski, T. J. (1997). The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences*, 20(4), 537–556. discussion 556–596.
- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87.
- Richards, J. E. (2005). Localizing cortical sources of event-related potentials in infants' covert orienting. *Developmental Science*, 8(3), 255–278.
- Ritter, W., Sussman, E., Deacon, D., Cowan, N., & Vaughan, H. G. (1999). Two cognitive systems simultaneously prepared for opposite events. *Psychophysiology*, 36(6), 835–838.
- Saffran, J. R., Pollak, S. D., Seibel, R. L., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition*, 105(3), 669–680.
- Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, 39(3), 484–494.
- Sato, Y., Yabe, H., Hiruma, T., Sutoh, T., Shinozaki, N., Nashida, T., et al. (2000). The effect of deviant stimulus probability on the human mismatch process. *NeuroReport*, 11(17), 3703–3708.
- Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, 8(10), 1391–1400.
- Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, 72(2), B11–B21.
- Squires, N., Squires, K., & Hillyard, S. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, 38(4), 387–401.
- Stefanics, G., Háden, G., Huotilainen, M., Balázs, L., Sziller, I., Beke, A., et al. (2007). Auditory temporal grouping in newborn infants. *Psychophysiology*, 44(5), 697–702.
- Tadel, F., Baillet, S., Mosher, J., Pantazis, D., & Leahy, R. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis [electronic version]. *Computational Intelligence and Neuroscience* (13p.).
- Takahashi, E., Folkerth, R., Galaburda, A., & Grant, P. (2012). Emerging cerebral connectivity in the human fetal brain: An MR tractography study. *Cerebral Cortex*, 22(2), 455–464.
- Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J., & Bonatti, L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science*, 332(6033), 1054–1059.
- Telkemeyer, S., Rossi, S., Nierhaus, T., Steinbrink, J., Obrig, H., & Wartenburger, I. (2011). Acoustic processing of temporally modulated sounds in infants: Evidence from a combined near-infrared spectroscopy and EEG study. *Frontiers in Psychology*, 1, 62.
- Tenenbaum, J., Kemp, C., Griffiths, T., & Goodman, N. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, 331(6022), 1279–1285.
- Vul, E., & Pashler, H. (2012). Voodoo and circularity errors. *NeuroImage*, 62(2), 945–948.
- Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *The Journal of Neuroscience*, 32(11), 3665–3678.
- Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T. A., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 108(51), 20754–20759.
- Wetter, S., Polich, J., & Murphy, C. (2004). Olfactory, auditory, and visual ERPs from single trials: No evidence for habituation. *International Journal of Psychophysiology*, 54(3), 263–272.
- Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of Psychophysiology*, 21(3), 147–163.
- Winkler, I., Cowan, N., Csépe, V., Czigler, I., & Näätänen, R. (1996). Interactions between transient and long-term auditory memory as reflected by the mismatch negativity. *Journal of Cognitive Neuroscience*, 8(5), 403–415.
- Winkler, I., & Czigler, I. (2012). Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. *International Journal of Psychophysiology*, 83(2), 132–143.