Thèse de doctorat
de l'université Pierre et Marie Curie

Spécialité Neurosciences Cognitives

Présentée par

Claire KABDEBON

Sequence encoding in preverbal infants: an electrophysiological perspective

Thèse soutenue le 13 Décembre 2016

Devant le jury composé de :

M.	Luca Bonatti	Rapporteur
Mme	Teodora Gliga	Rapporteure
Mme	Anne Christophe	Examinatrice
M.	Jean-Rémy Hochmann	Examinateur
M.	Lionel Naccache	Examinateur
Mme	Ghislaine Dehaene-Lambertz	Directrice de thèse

REMERCIEMENTS

Pour finir – ou pour commencer – je souhaite profiter de cet espace pour remercier tous ceux qui, de près ou de loin, m'ont accompagnée, soutenue, supportée au cours de ces dernières années de thèse. La palme revient évidemment à Ghislaine, pour m'avoir ouvert les portes de la recherche, pour m'avoir fait confiance alors que je ne savais rien, et pour m'avoir appris à peu près tout ce que je sais du développement et du langage. Elle a su cultiver mon autonomie tout en restant disponible et à l'écoute. Merci surtout pour ta créativité, ta curiosité et ta liberté, j'espère avoir été contaminée ! Je remercie également Marcela Peña pour sa douceur et sa disponibilité, et Marco Buiatti pour m'avoir donné gout au frequency tagging !

Ensuite, il me faut exprimer toute ma reconnaissance à ceux qui constituent le cœur que cette thèse : les quelques 250 bébés que j'ai croisés, physiquement ou virtuellement, au cours de ma thèse, ainsi que leurs parents. Un grand merci à Antoinette pour en avoir vu une partie à Neurospin, et à Anne-Caroline et son équipe de choc du LSCP pour m'avoir permis d'en accueillir une autre partie dans des conditions exceptionnelles à la maternité.

Ce travail n'aurait pas été possible sans le soutien et la bienveillance de toute l'équipe développement de Neurospin ! Merci donc à François qui m'a guidée avec tellement de gentillesse dans mes premiers pas au laboratoire, à Jessica D. pour sa présence toujours rassurante, ses précieux conseils et sa relecture attentive, à Parvaneh bien sûr pour ses encouragements, pour les pauses thé, les discussions, les rires et le soutien psychologique au cours de l'épreuve redoutée de l'écriture, merci aussi à Karima pour sa confiance quand la mienne me faisait défaut, et pour nos échanges toujours tellement intéressants quel qu'en soit le sujet, et merci à Jessica L. pour son soutien et ses réflexions sur le bonheur et l'éducation. J'ai aussi une pensée pour l'ancienne équipe : Alice, Karla, Irène, Cécile, Océane, Hina, et bien sûr pour la relève : Pablo, Leslie, Éric, Marie, Thien-Li... J'ai aussi une pensée pour toi Nelly, ton tour viendra !

Ce travail s'est grandement nourri des interactions que j'ai pu avoir avec Caroline Huron, Virginie van Wassenhove, Sebastien Marti ou Manuela Piazza, ainsi qu'avec tous les brillants esprits du Sequence Meeting : Stanislas Dehaene, Christophe Pallier, Florent Meyniel (et son Gibbs !). J'ai également été heureuse de pouvoir partager des trajets/discussions avec Marie, Baptiste, Laetitia, Murielle, et tous les usagers des lignes 91.10 et 91.06 que j'oublie ! Je remercie enfin celles qui font vivre le laboratoire : Vanna et Laurence.

Je tiens à remercier Luca Bonatti et Teodora Gliga pour avoir accepté la lourde tâche de rapporteur, ainsi que les membres de mon jury de thèse : Anne Christophe, Jean-Rémy Hochmann et Lionel Naccache.

Merci à mes parents, mon frère et ma sœur, qui me donnent la force d'avancer.

Merci, enfin, Ivan. D'être toi, avec moi.

ABSTRACT

To this day, the infant brain is the only known learning system able to apprehend and master the complexity of the human language. Developmental psychologists have dedicated a lot of efforts to break down the mystery of language acquisition, revealing precocious and impressive abilities for processing and encoding speech sequences. The recent emergence of non-invasive neuroimaging techniques provides a new tool to explore language learning mechanisms from a different perspective. In the present thesis, we aimed at investigating the encoding mechanisms of the structural properties of a speech sequence from an electrophysiological perspective. In the first part of this thesis, we provided the developmental neuroimaging community with a methodological contribution. Based on magnetic resonance imaging (MRI) data, we virtually localized the standardized sensor placement system for both electroencephalography (EEG) and near infrared spectroscopy (NIRS) relative to the internal brain structures, and assessed their variability. We additionally provided an infant brain template with an anatomical atlas which will be valuable for studies in which individual anatomical information cannot be obtained. In the second part of this thesis, using high-density EEG, we demonstrated that 8 month-old infants could deploy powerful learning mechanisms for capturing the statistical dependencies between non-adjacent syllable units, in order to chunk a continuous speech stream. Interestingly, a hierarchy of neural processes tracked both the syllables and the chunked constituents of the sequence. Finally, in a third cognitive EEG study, we proposed an experimental design to assess infants' ability to not only extract but also encode the structure of speech sequences into unified mental schemas. The results of this study established that 5 month-old infants could form robust mental representations for repetition-based sequences, allowing them to represent, categorize and operate on multiple structures. Inspection of various neural measurements revealed that several stages of the processing hierarchy were affected by the acquired mental representations. Overall, this thesis complements behavioral research on language acquisition with a window onto the early neural mechanisms allowing sequence encoding, revealing a hierarchy of increasingly complex computations in the encoding of linguistic structures.

RÉSUMÉ

A ce jour, seul le cerveau du nourrisson est capable d'appréhender et de maîtriser la complexité du langage humain. La recherche en psychologie du développement a investi beaucoup d'énergie pour tenter de percer le mystère de l'acquisition du langage, révélant d'impressionnantes capacités précoces permettant le traitement et la représentation de la parole. La récente émergence de techniques de neuro-imagerie non-invasives offre aujourd'hui de nouveaux outils et de nouvelles perspectives pour d'étude des mécanismes d'apprentissage du langage. Cette thèse a pour but d'explorer les mécanismes permettant l'acquisition et la représentation de structures linguistiques, grâce à une approche en électrophysiologie. La première partie de ce manuscrit consiste en une contribution méthodologique à la neuro-imagerie du développement. Sur la base de données acquises en imagerie par résonnance magnétique (IRM), nous avons localisé les positions du système international 10/20 pour le placement d'électrodes – utilisé aussi bien en électroencéphalographie (EEG) qu'en spectroscopie proche infra-rouge (NIRS) – par rapport aux structures cérébrales internes. Cette étude a permis de quantifier la variabilité interindividuelle de ces positions, mais également de construire un modèle de cerveau complété d'un atlas anatomique pour le nourrisson. Dans une seconde partie de cette thèse, grâce à l'EEG haute-densité, nous avons pu démontrer que dès 8 mois, les nourrissons étaient capables de mettre en œuvre de puissance mécanismes d'analyse statistique, permettant d'extraire les dépendances entre syllabes non-adjacentes, pour segmenter un flux continu de parole en unités distinctes. L'analyse des réponses neurales a révélé une hiérarchie de processus cérébraux soutenant le traitement des syllabes mais aussi des unités segmentées. Enfin, dans une dernière partie, nous proposons un paradigme expérimental permettant d'étudier non seulement l'extraction mais aussi la représentation de séquences linguistiques sous la forme d'expressions unifiées. Nous avons pu établir grâce à cette étude que dès 5 mois, les nourrissons étaient capables de former de solides représentations de séquences définies par des répétitions, leur permettant de catégoriser et manipuler plusieurs structures. Pour conclure, cette thèse vient compléter les études comportementales sur l'acquisition du langage, grâce à une approche des processus

cérébraux soutenant l'apprentissage de séquences. La richesse du signal EEG a permis de mettre en évidence une hiérarchie de traitements complexes.

CONTENTS

LIST OF FIGURES

List of Tables

CHAPTER 1

INTRODUCTION

Parents welcome their offspring's first words with special joy and pride because they mark infants' entrance into a truly symbolic system of social interactions. However, the last fifty years of developmental behavioral research, and the recent emergence of non-invasive brain imaging techniques have revealed that these first productions actually issue months of complex linguistic computations, developing from birth on. Before they pronounce their first words, infants have to understand that speech is for sharing information. They have to notice that this special signal is built out of phonemes that combine to form words which refer to objects and concepts. They have to learn the rules that govern the organization of words in order to form sentences. Interestingly, the infant brain, although immature, outperforms in terms of speed or efficiency any other learning device, from the sophisticated adult brain to highly trained animals, or powerful algorithms. By three years of age, regardless of their cultural environment, most infants not only master the underlying structure of their native language, but they also manage to exploit its infinite expressive power to describe new situations, or express new emotions. These amazing learning performances seem to rely on infants' ability to process and encode speech sequences at different representation levels, from their basic acoustic features, to their structural organization, and to readily derive knowledge from each of these description levels. In the present chapter, we explore how this hierarchical encoding of speech sequences develops: we address the learning mechanisms and we investigate the nature of the encoded representations.

1.1 The role of speech

Before capturing its linguistic content, the most global description level for speech is to represent its abstract function for **transferring information from the speaker to the recipient**. Human adults readily understand this communicative function, even when they cannot understand the content of speech. Infants seem to develop this abstract representation of the speech signal before they have learned the structural properties of language and the meanings of words. From birth on, speech is a very special auditory stimulus for which infants show a particular **preference**, compared to other types of auditory stimuli, even when spectral and temporal properties are preserved (Colombo & Bundy, 1981; Spence & DeCasper, 1987; Mehler et al., 1988; Vouloumanos & Werker, 2004, 2007). By 6 months of age, they expect speech to be addressed towards humans rather than objects (Legerstee et al., 2000), and by the end of the first year, they use speech directed at them in ostensive communicative contexts to derive generalizable knowledge (Csibra & Gergely, 2009). Besides this social saliency of language, some recent studies have provided strong evidence that very young infants understand the abstract function of speech. Marno et al. (2015) presented 4 month-old infants with videos of a face articulating speech, while looking at the infant before gazing towards the left or right side of the screen. The face then disappeared, and an object was revealed on the same side of the screen. The authors reported that infants orient faster towards the object when the video was presented with speech compared to backward speech or silence. These results suggested that infant have the intuition that **speech refers** to the objects of the environment.

Martin et al. (2012) further investigated 12 month-olds' understanding that speech not only refers, but also **transfers** information. Infants were presented with a short interaction between two experimenters. During a familiarization phase, the first experimenter chose one object among two, indicating her preference for this target object. In the following scene, the second experimenter was facing the two objects, interacting with both objects, showing no preference. In a final scene, the two experimenters were present in front of the two objects, but the first experimenter could no longer reach the objects. She then turned towards the second experimenter and either pronounced a nonce word "koba" (speech condition) or coughed (non-speech condition). The second experimenter thereafter selected the target object or the non-target object. Authors reported that infants were surprised when the experimenter selected the non-target object, but only in the speech condition. In the non-speech condition, looking-time patterns were similar when the target or non-target objects were selected. Infants appeared to expect that the nonce word, unlike coughing, conveyed information from the speaker to the recipient about the target object. These results were thereafter replicated in 6 month-old infants (Vouloumanos et al., 2014). During the first post-natal year, so far before pronouncing any word, infants therefore understand that speech is more than a very salient stimulus, they also represent its abstract function for communicating **between** individuals, **about** something.

1.2 The sounds of speech

A second representation level for speech is based on its **physical features**. Speech is primarily a continuous acoustic signal characterized by complex variations of energy over time, with quasi-periodic components (vowels), interrupted by spectro-temporal irregularities (consonants). Infants must therefore process these fundamental features which already convey a wealth of information, before decoding its linguistic content. A body of neuroimaging studies have evidenced that from birth on, speech processing relies on a **dedicated cerebral network**, very similar to the one that is described in adults (Dehaene-Lambertz et al., 2002; Peña et al., 2003; Dehaene-Lambertz et al., 2010). Although it is still unclear which acoustic features critically distinguish speech from any other auditory stimuli, these studies indicate that infants benefit from innate constraints allowing them to readily capture these speech-specific features, and channel the signal towards dedicated processing mechanisms.

Infants show remarkable abilities for processing the **temporal dynamics of speech**. From the last trimester of gestation, when audition becomes functional, they can start exploring the external world. The uterine wall filters out a lot of information, but rhythm and melodic contours of sounds are well preserved (Armitage et al., 1980), in particular maternal speech. DeCasper & Spence (1986) evidenced that **fetuses** are able to process this information. In this experiment, pregnant women were asked to read a particular story aloud during the last 6 weeks of pregnancy. The authors reported a listening preference of newborns for this specific story after birth, even when read by another speaker, demonstrating both early rhythmic processing and memory abilities. This early exposure to the dynamics of auditory stimuli might explain neonates' impressive sensitivity for processing prosodic attributes of speech. For example, they can discriminate two languages from two different rhythmic classes, like English and Japanese (Nazzi et al., 2000; Ramus et al., 2000; Mehler et al., 1988). These initial perceptual abilities are thereafter strengthened by powerful learning and maturational mechanisms. After a few months of extra-uterine life, they progressively refine their representation for the sounds of their environment: 5 month-old Spanish infants exposed to both Catalan and Spanish in their everyday environment can discriminate these two roman languages (Bosch & Sebastian-Galles, 2001). Similarly, 9 month-olds prefer to listen to words that exemplify their native language's stress pattern, while 6 month-olds fail to show such a native language stress preference (Jusczyk, Cutler, & Redanz, 1993).

Infants also rely on prosodic processing of human vocalizations to retrieve **par-**

alinguistic information. Based on this strategy, they can discriminate among different speakers. For example they are especially sensitive to their mother's voice (DeCasper & Fifer, 1980), but this preference disappears when the mother reads a page in reverse, from the bottom right to the top left corner, therefore disrupting the prosody of natural language (Mehler et al., 1978). Nevertheless, infants can also represent the formant and timber information of speech, which carry speaker identity. Newborns were shown to discriminate two unknown voices based on isolated disyllabic word utterances (Floccia et al., 2000). Similarly, using Near Infra-Red Spectroscopy (NIRS), Mahmoudzadeh et al. (2013) could record brain responses to a change of voice (male vs female) for minimal stimuli (syllables) in preterm infants, after only 6 months of gestation. The emotional content of speech is also encoded in its prosodic contours. Some behavioral studies have reported different affective reactions (facial expressions, looking time) in response to approving vocalizations compared to disapproving vocalizations, in 4 (Papousek et al., 1990) and 5 month-old infants (Fernald, 1993). A few neuroimaging studies complemented these results, and identified a number of brain regions modulated by the emotional content of speech in 4 to 7 month-old infants (Blasi et al., 2011; Grossmann et al., 2010).

Finally, besides this early ability to process the suprasegmental features of speech, infants show initial sensitivity for discriminating even **subtle acoustic variations**. Neonates and infants not only discriminate the different vowels, they can also perceive the slight voicing difference between /b/ and /p/, or the tenuous change of place of articulation between /b/ and /g/ (Bertoncini et al., 1987; Eimas et al., 1971). Differences in brain responses are also reported following the phonetic change from /b/ to /g/ in preterm infants, after only 6 months of gestation, when neurons are still migrating to form the different cortical layers (Mahmoudzadeh et al., 2013). The neural circuitry of the immature auditory cortex seems nevertheless efficient enough to detect the 20 to 40 ms long variations that distinguish such phonemes. This early phoneme-sensitive cortical network allows infants to initially perceive the acoustic nuances of language, and later derive knowledge from the regularities of these sounds. For example, after a few months of exposure to their native language infants have acquired some expertise about the language-specific constraints on the possible sequencing and positions of phonetic segments allowed in words (Jusczyk, Friederici, et al., 1993; Gonzalez-Gomez & Nazzi, 2012).

Very early on, infants appear to benefit from an efficient neural machinery for representing the various acoustic features of speech. Both **suprasegmental** and **fine-grained** representations of the acoustic environment are processed, and submitted to specific learning mechanisms. Some acoustic specificities of the native language are thereafter progressively acquired through both maturational and statistical learning processes (Pena et al., 2010).

1.3 The units of speech

However, language acquisition is not about mere perceptual abilities. Additional systems are required for storing, comparing, classifying the sensory inputs, in order to **organize** the various percepts into a **finite set of units**.

1.3.1 The phonetic repertoire

Each language is based on a limited number of fundamental sound units that can change a word's meaning: the **phonemes**. The total inventory of phonemes across the world numbers in the thousands, but each language uses only a small subset of these sounds, ranging from 11 (Polynesian) to 141 (Khoisan or "Bushman") (Maddieson & Disner, 1984) phonemes. Adults typically exhibit strong perceptual discontinuities for phonetic variations that induce a change of phoneme in their native language (Liberman et al., 1967). For example, when presented with pairs of stimuli of equal-sized differences, adults reliably discriminate only those differences to which they are able to assign different phonetic category labels in their native language, whereas their perception of differences for stimuli belonging to the same phonetic category remains poor. Interestingly, this **categorical perception for speech** is attuned to the phonetic categories of the native language: while a French listener clearly perceives the /r/ vs /l/ contrast, a Japanese listener cannot disentangle the two percepts (Miyawaki et al., 1975). The first studies of infant speech perception actually explored the ontogeny of categorical phonetic representations, and revealed that, surprisingly, very young infants, just like adults, also discriminate between stimuli that adults label as instances of different categories, holding acoustical differences constant (Bertoncini et al., 1987; Eimas et al., 1971; Morse, 1972; Eimas, 1974; Trehub, 1973; Eimas et al., 1987). More interestingly, young infants discriminate **all phonetics contrasts** whether or not they are used in the language they are learning (Lasky et al., 1975; Streeter, 1976; Trehub, 1976; Aslin et al., 1981), and they progressively loose this ability around the end of the first year (Werker & Tees, 1984).

Several imaging studies have revealed that the phonetic code is actually directly encoded in the **auditory cortices**. In adults, large auditory mismatch responses were recorded using electroencephalography (EEG) in response to native phonetic contrasts, but not to equal-sized non-native phonetic contrasts, or within-category acoustic variations (Dehaene-Lambertz, 1997; Näätäneiv et al., 1997). 3

month-olds exhibit the same response pattern with a strong auditory mismatch for phonetic contrasts compared to within-category acoustic variations (Dehaene-Lambertz & Baillet, 1998), suggesting that the neural architecture of the auditory cortex innately supports categorical phonetic representations. Recent studies in adults further investigated the cerebral organization allowing phonetic processing, and the **posterior superior temporal gyrus** (pSTG), part of classical Wernicke's area appears to play a critical role in the transformation of acoustic information into phonetic representations. Using intracranial recordings in human adults, Chang et al. (2010) evidenced that synthesized speech stimuli varying in small and acoustically equal steps evoked invariant cortical population response patterns in the pSTG, organized along phonetic categories. The authors reported strong concordance between neuro- and psychometric functions, suggesting that this region implements rapid transformation from continuously variable acoustic speech signals to discrete phonetic sound categories. Mesgarani et al. (2014) extended these results using natural continuous speech to reveal the pSTG representation of the entire American English phonetic inventory. These results, together with previous studies showing categorical perception from birth on indicate that the phonetic code is a byproduct of innate neural constraints of the auditory cortex. Besides, these perceptual constraints are shared with other animals which do not benefit from any linguistic experience (Kuhl & Miller, 1978; Morse & Snowdon, 1975; Kluender et al., 1987). Languages simply capitalized on these **natural auditory discontinuities**, to derive rich communicative codes.

During infancy, these innate neural representations for phonemes undergo a **large reorganization** to adapt to the specificities of the native language, and converge towards adult-like representations. In a seminal study, Werker & Tees (1984) evidenced this rapid evolution of perceptual abilities around the end of the first year of post-natal life. The authors reported that 4- to 8-month-old English infants can easily discriminate Hindi and Salish non-native consonant contrasts, but that this discrimination substantially declined by 12 months of age. On the contrary, 12 month-old Hindi and Salish infants can still discriminate these same consonant contrasts. The **linguistic environment** therefore shapes perceptual sensitivity. Infants progressively neglect contrasts that are irrelevant for their native language (Best et al., 1988; Bosch & Sebastian-Galles, 2003), and their ability to discriminate native-language phonetic units improves (Cheour et al., 1998; Kuhl et al., 1992, 2006; Maye et al., 2008). This developmental change is probably due to the interaction of several mechanisms. First, infants appear to analyze the **statistical distribution of sounds** that they hear in ambient language (Kuhl et al., 1997). To demonstrate the power of this learning mechanism, Maye et al. (2002) familiarized 6- to 8-month-old infants with syllables from a phonetic continuum, exhibiting either a bimodal or unimodal frequency distribution along the continuum. After a 2-minute familiarization, discrimination ability for syllables at the endpoints of the continuum was assessed, and only infants exposed to the bimodal distribution were sensitive to the phonetic contrast. Features that occur in a bimodal distribution are treated as phonemic, while features that occur in a unimodal distribution are treated as allophonic variant of a single phoneme. Phonetic representations are therefore shaped by linguistic experience. Second, some **maturational processes** additionally constrain this perceptual reorganization. Yoshida et al. (2010) reported that at 10 months of age, a 2-minute familiarization is no longer sufficient to shape infants' discrimination abilities: the authors had to double the familiarization duration to replicate Maye et al.'s results. This progressive decrease of learning abilities suggests some neuro-biological changes around the end of the first year. Finally, this developmental change appears to be mediated through social interactions (Kuhl et al., 2003). Mere exposition to linguistic input through audio or audio-visual video recordings is not enough to elicit a stable reorganization of the perceptual space; infants have to engage in **communicative interaction** and **joint attention** with other individuals.

1.3.2 The lexicon

While phonetic representations appear to rely on rather basic perceptual and computational abilities shared with other animal species, building a lexicon on the contrary depends upon more sophisticated abilities. Infants have to **segment** the appropriate sound sequences from fluent speech, in order to build stable representations for word units, and **map** them onto objects or events of their environment.

Word segmentation is a daunting task for preverbal infants as continuous speech lacks consistent physical cues for word boundaries. Nevertheless, they manage to deploy a toolkit of strategies for discretizing the speech stream. While there are no robust marker for words' onsets and offsets, the speech stream is naturally organized into phonological phrases thanks to the presence of various non-lexcial segmentation cues such as phonotactics, allphony, coarticulation and stress. A phonological phrase typically consists of one or more lexical items, potentially grouped with some functional elements (e.g. [the student] [is working hard]). It typically spans over a single melodic contour, and it is characterized by pre-boundary lengthening, initial strengthening and reduced coarticulation at edges. The edges of phonological phrases necessarily align with some word boundaries, and these acoustic cues allow a coarse perceptual chunking of spoken language, and thus a first indication for words boundaries. Several studies have evidenced that infants are sensitive to these acoustic features (Hirsh-Pasek et al., 1987; Christophe et al., 1994, 2001; Gerken et al., 1994). Gout et al. (2004) further showed that

10-month-old infants can use this information for on-line word extraction. The authors trained participants to turn their head in response to an isolated word like "paper", and subsequently tested their reactions to sentences containing the word phonemes: "the college with the biggest paper forms is best", and "the butler with the highest pay performs the most". Infants turned significantly more often when the word truly appeared in the sentence. These results demonstrate that infants use their **prosodic processing** skills for discovering word units in fluent speech. Shukla et al. (Shukla et al., 2011) additionally showed that the extracted word-like representations using such mechanisms were not mere sound sequences, but had a linguistic status, and could be used as labels for visual objects.

Infants can also learn some word-like units from **frequent and isolated utterances**. By the age of 5 months, infants prefer listening to their own first name, rather than an unfamiliar one with a similar prosodic contour (Mandel et al., 1995). At 6 months, they recognize frequent words like "mommy" or "daddy" (Tincoff & Jusczyk, 1999), suggesting that they readily capture such recurring sound patterns, often presented in isolation. They can thereafter exploit these highly familiar words to segment and recognize adjoining words. Bortfeld et al. (2005) showed that 6 month-old infants could extract and subsequently recognize a novel word from short speech passages in which the word to be segmented was preceded by their first name. Altogether, these studies indicate that infants represent frequent words presented in isolation as initial anchors in the speech stream, and use them for discovering new units.

However, infant-directed speech does not reliably provide a significant amount of isolated words (Aslin et al., 1996). Infants must therefore use more efficient strategies. Saffran et al. (1996) noted that in natural speech, sounds that often co-occur are most likely word units. They thus proposed that speech could be considered as a predictable sequence of sounds: under this hypothesis, transition probabilities (TPs) from one sound to the next are therefore higher within words than between words. Syllables being privileged computation units very early on (Bertoncini & Mehler, 1981), the authors designed a continuous sequence of syllables in which **transition probabilities** between adjacent syllables were the only cues for discovering the embedded words. More specifically, TPs were maximal (TP=1) within the nonce artificial words, and dropped to 0.3 between syllables that spanned word boundaries. After only 2 minutes of familiarization to this continuous speech stream, 7-month-old infants successfully distinguished the nonce words from part-words (strings consisting of sequences spanning a word boundary). These results were replicated in 5 month-old infants (E. K. Johnson & Tyler, 2010), demonstrating that complex **statistical computations** are functional very early on, and infants can use this strategy to chunk the speech input.

A few studies have further investigated the nature of the output of such statistical learning process. Do infants represent the extracted chunks as relatively coherent sound sequences, with no particular status regarding their native language? Alternatively, do they readily represent these speech chunks as **lexical entries** for their mental lexicon? To address this issue, Estes et al. (2007) familiarized 17 month-old infants with an artificial speech stream in which the only reliable cue to word boundaries were TPs between adjacent syllables. Infants were then taught two novel label-object associations where the labels were words from the artificial language, or part-words. In the test phase, infants were presented with label-object pairings that could be either consistent or inconsistent with what they had just been taught. Their looking patterns indicated that infants only learned the label-object associations when the nonce words were used as labels for objects. Only the high-transitional probability sequences appeared to be represented as word candidates. These results indicate that the learning mechanism used in statistical word segmentation experiments yields representations that have a linguistic status, and can therefore be used as input for future word learning. Hay et al. (2011) extended these results to fluent speech drawn from an unfamiliar natural language. Nevertheless, although there is additional evidence that infants might actively track transitional probabilities in their everyday language exposure (Ngon et al., 2013), the reliability of statistical segmentation remains questioned in more ecological learning conditions (E. K. Johnson & Tyler, 2010).

It seems that no single strategy underlies the discretization of fluent speech into words; infants are rather sensitive to **myriad cues** that help them achieve the word segmentation task. Depending on their expertise in representing the various aspects of the speech input, they appear to attend to different kind of cues. For example, Thiessen et al. (2003) demonstrated that while 6- to 7-month-olds prioritize statistics over rhythmic information, 9-month-olds prefer using their knowledge of the stress pattern of words in their native language to discover new words. Once they have acquired enough linguistic knowledge, they can even rely on grammatical or syntactic cues to discover new words (Bernal et al., 2007). Interestingly, very early on, infants seem endowed with some intuition that the segmented word-sized units of speech are lexical entries related to objects in their environment. This early understanding of the symbolic nature of language might be tightly linked to infants' ability to represent speech as a communicative signal (cf. section 1.1).

1.4 The structure of speech

Perhaps the most intriguing aspect of language acquisition is infants' early ability for grasping and representing the complex **combinatorial principles** that govern the organization of language units into larger constituents, without any explicit instructions.

1.4.1 Long distance dependencies

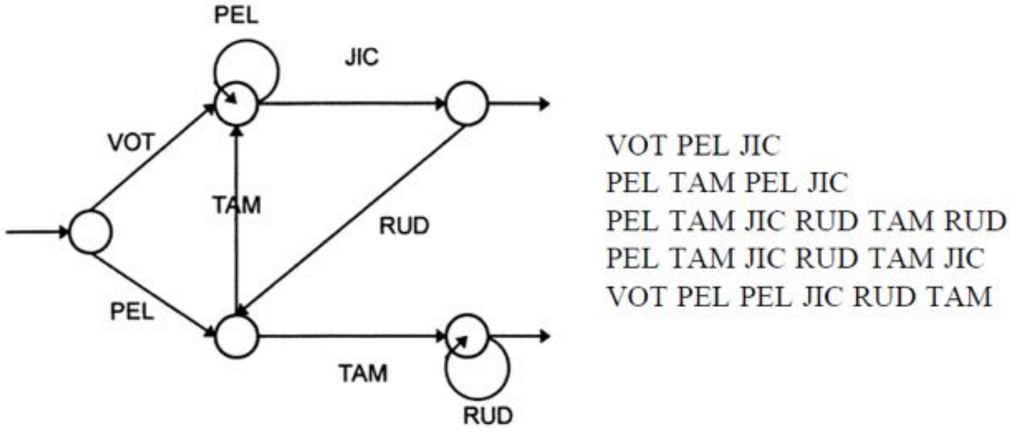
Many languages express some syntactic rules as systematic **dependencies between non-adjacent words or sub-lexical morphemes**. For example, English exhibits several relations between distant morphemes, with the progressive forms of verbs (be + -ing), subject-verb agreement, or number agreement. Similarly, most of the world's languages rely on morphology to mark grammatical distinctions and relations between distant linguistic units. Infants must therefore track dependencies between non-adjacent words or morphemes. Santelmann & Jusczyk (1998) tested English-learning infants with their natural language, and demonstrated that 18 month-olds, but not 15 month-olds, could distinguish between grammatical structures such as "grandma is singing" and ungrammatical structures such as "grandma can singing". 18 month-olds could track the long-distance dependency between the auxiliary and the "–ing" ending as long as the distance between the two units was below 4 syllables. English-learning infants thus appear to consistently acquire such morphosyntactic dependencies during the second year, but the window of analysis remains limited. Gomez (2002) replicated these findings with 18 month-old infants using an artificial language paradigm. Participants were familiarized with strings of three nonce words separated by 250 ms silences, exhibiting systematic dependencies between the first and last units, while the middle element varied (AxC dependencies). At test, listening time was recorded in response to strings that conformed to, or disrupted the AxC dependencies. Crucially the authors showed that an increased variability of the middle element (meaning lower TPs between adjacent units) led to better detection of non-adjacent dependencies. These results indicated that infants turn to exploring non-adjacent dependencies only when adjacent dependencies were less predictable. Infants therefore favor different strategies to make sense of the input depending on its statistical structure. Gomez & Maye (2005) thereafter extended these results showing that by 15 months of age, infants begin to represent long-distance dependencies under a condition of high variability for the middle element. Marchetto et al. (2014) recently showed that 12 month-old infant could also track long-distance dependencies between syllables when regularities were instantiated at the word level.

Using a brain imaging paradigm, Friedericci et al. (2011) evidenced a much earlier sensitivity to long distance dependencies. The authors familiarized German-learning 4-month-old infants with Italian sentences containing two different non-adjacent dependency types between an auxiliary and a verb suffix: namely "sta x-ando" and "puo x-are" (the intervening x element was drawn from a set of 32 verb stems). Some test periods were regularly integrated during exposure, with grammatical and ungrammatical ("sta x-are" or "puo x-ando") sentences. The authors reported a positive electrical component over parietal areas in response to ungrammatical structures, indicating that from 4 months of age, infants could represent long distance grammatical dependencies from natural speech input.

1.4.2 Towards abstraction

Although this early sensitivity for tracking the recurring sounds from the input is particularly impressive, language acquisition entails more than learning surface features' regularities. The different units of an utterance are actually bond together through abstract relations defined by their grammatical categories and the syntax. For instance, a determiner or an adjective relate to a noun, which can be the subject or the language object of a verb, and these different units must occur with a specific ordering depending on the language. Infants therefore have to discover and represent this **hidden web of abstract regularities** from the incoming linguistic input, in order to derive the underlying grammatical structure, and process novel utterances accordingly. In a seminal study, Marcus et al. (1999) investigated the foundations of abstraction abilities, using a minimal syntax. They exposed 7 month-old infants to exemplars of a schema defined by **identity relations** between the constitutive syllables: ABB (e.g. "ga ti ti", "le di di") or ABA (e.g. "ga ti ga", "le di le"). Infants were then tested using novel syllables which were arrayed in either the familiarized or the novel schema (e.g. "wo fe wo" or "wo fe fe"). In this paradigm, infants could not rely on transition probabilities, or any statistical analysis of the surface features to discriminate between the two sequence types; instead, they had to represent identity relations between syllables, and notice the recurring pattern at this abstract representation level. Listening behavior revealed that infants could discriminate between items of the habituation structure and items of the other structure, indicating that they had access to some **abstract features** of speech sequences.

With a similar paradigm, Gomez & Gerken (1999) exposed 12-month-old infants to speech sequences instantiating a simple artificial grammar generated by a finite-state automaton (see Figure 1.1). At test, the authors showed that par-



VOT PEL JIC
PEL TAM PEL JIC
PEL TAM JIC RUD TAM RUD
PEL TAM JIC RUD TAM JIC
VOT PEL PEL JIC RUD TAM

Figure 1.1: **Example of a finite-state automaton used to generate artificial grammatical strings.** A finite-state grammar is characterized by a Markovian process in which a transition from state n to state n+1 produces a word. Strings are generated by starting at the leftmost state and traversing links in the system in the direction of the arrows. Finite-state grammars are limited in terms of their generative capacities, but are nevertheless complex systems in which structure is not limited to first-order dependencies. On the right are several examples of grammatical strings generated by this finite-state grammar.

ticipants could recognize novel sequences governed by the same grammar but instantiated with different units. Further analyses revealed that infants did not actually learn the entire artificial grammar's sequential dependencies, but rather patterns of repetitions (Gomez et al., 2000). Similarly to Marcus et al.'s study, infants could generalize to a novel vocabulary based on identity relations between the successive units. Although these studies did not provide a precise description of the mental representations of such repetition-based sequences, they evidenced **early mechanisms for abstraction**. Besides, these abstraction abilities are not limited to the identity relation: infants readily form abstract representations for numbers (Izard et al., 2008) or ordinal relations (Brannon, 2002), and they can represent and generalize visual sequences of increasing or decreasing magnitude (De Hevia & Spelke, 2010). These initial abilities for representing the abstract features of the speech input lay the foundations for language acquisition.

1.4.3 Climbing the syntactic tree

Abstraction is fundamental for discovering grammatical categories, and representing the syntactic relations between these categories. However, learning the underlying syntactic structure from natural language is much more challenging than grasping a repetition pattern from a highly controlled and minimal artificial stimulation. Infants nonetheless manage to progressively make sense of the input. During the second semester of life, they begin to segregate **function words** from **content words**. Function words typically display specific phonological and acoustic properties: they are unstressed, shorter and simpler compared to content words. Besides, they tend to occur at the edges of prosodic units, and their frequency of occurrence is much higher than that of content words. Thanks to their abilities for representing the acoustic and statistical properties of speech, infants are sensitive to both of these phonological (Shi et al., 1999) and distributional cues(Gervain, Nespor, et al., 2008; Hochmann et al., 2010). By 6 to 10 months of age, infants recognize and store the most frequent function words of their native language (Shi, Marquis, et al., 2006; Shi, Cutler, et al., 2006), and use them as cues to segment adjacent words (Hallé et al., 2008). Young English learners also possess some knowledge of the determiner-noun ordering (Shady et al., 1995). Several studies have additionally provided evidence that, during the second year of life, infants readily abstract the different **categories of function words**, such as determiners (signaling nouns) and pronouns (signaling verbs) (Höhle et al., 2004; Shi & Melançon, 2010). They additionally know that function words, unlike content words, are not associated to specific referents such as objects (Hochmann et al., 2010).

As infants progressively identify the various function words, they can start noticing some abstract regularities between word categories. Using an artificial language learning paradigm, Gomez & Lakusta (2004) evidenced these powerful abstraction abilities in 12 month-old infants. Participants were familiarized to strings comprising two words (e.g. "erd deech"): a pseudo-determiner ("erd") and a pseudo-noun ("deech"). Pseudo-nouns were drawn from two different lists: a list of monosyllabic words, systematically following the pseudo-determiners "erd" or "ong", and a list of disyllabic words, systematically following "alt" or "ush". Infants had to categorize pseudo-determiners from pseudo-nouns, and notice the systematic dependency between the pseudo-determiners and the number of syllables of the pseudo-nouns. At test, infants were presented with strings comprising one of the pseudo-determiners and a novel pseudo-noun, either mono- or disyllabic. The pairing could conform to the familiarization material (e.g. "erd rud" or "ush roosa") or not (e.g. "erd roosa" or "ush rud"). Listening behavior indicated that infants could represent the systematic relationship between the pseudo-determiner and the following pseudo-noun, and form abstract pseudo-noun categories according to syllable number. A following study additionally demonstrated that 22 month-old infants familiarized with this artificial language could readily associate each of these two syntactic categories with semantic categories (e.g. monosyllabic words with pictures of animals, and disyllabic words with pictures of vehicles). Importantly, infants who were not familiarized with the artificial language could not associate the semantic categories to the pseudo-nouns, based on their syllable number (Lany & Saffran, 2010).

In natural language, function words help infants discover the syntactic skeleton of an utterance, and they can progressively notice that, for example, some function words particularly relate to verbs (i.e. actions), and others relate more often to nouns (i.e. objects). They can thereafter use this knowledge to **infer the syntactic category** and the **meaning** of unknown content words. For example, in a behavioral study (Bernal et al., 2007), French 23-month-olds were taught a new word either in a verb context, that was preceded by a pronoun ("Regarde! Elle dase!" / "Look! It's dasing!"), or in a noun context, that was preceded by a determiner ("Regarde! Une dase" / "Look! This is a dase!"), while watching a video of an object performing a simple action. Children were then presented with two videos: the same object performing a novel action and a new object performing the same action, and participants trained with the verb context were asked to point to "celle qui dase" / "the one which is dasing", while participants trained with the noun context were asked to point to "la dase" / "the dase". The results demonstrated that toddlers possess grammatical categories for nouns and verbs, and they can correctly infer the category of the word "dase" based on the syntactic context. Similar results were obtained in Japanese- (Oshima-Takane et al., 2011) and English-learning infants (Waxman et al., 2009). Using the conditioned head-turn procedure, Cauvet et al. (Cauvet et al., 2014) demonstrated that by the age of 18 months, French-learning infants can use function words online as syntactic cues to recognize known nouns and verbs. Additional studies have shown that infants start to notice the abstract relation between determiners and nouns from 14 months of age (Shi & Melançon, 2010; Höhle et al., 2004). At this age, English-learning infants can also discriminate nouns from adjectives (Waxman & Booth, 2001), but they do not map adjectives to object properties before the age of 21 months (Waxman & Markow, 1998).

This body of experimental research suggests that infants show initial sensitivity for function words, providing a basis for syntactic processing of natural speech, and therefore allowing them to abstract some systematic dependencies between grammatical categories. Interestingly, infants readily represent linguistic units as more than mere acoustic or semantic units; they additionally organize them into abstract grammatical categories, a key component for the generativity of natural language. The mechanisms underlying this early intuition for describing words at such an abstract level is poorly understood, and appears to be unique to the human species.

1.5 A symbolic mind?

At this point, it clearly appears that, although infants typically utter their first words around their first birthday, preverbal infants are definitely not prelinguistic. Instead, they are silently building multidimensional representations of the speech input, quietly extracting linguistically relevant regularities from each of these dimensions, and deriving abstract knowledge from these regularities. By 12 months of age, they have already passed many important language-learning milestones, from understanding the abstract role of speech, to grasping the sound structure of their native language, or attaching meaning to a small cohort of frequent words and gathering initial knowledge to understand how words combine to form sentences. This transition from hearing speech as a string of meaningless sounds to perceiving a sequence of recognizable sounds and words seems to rely on the ability to partition the complexity of the input into multiple hierarchical levels of representation that interact and feed or constrain one another. In adults, speech processing is supported by a distributed cerebral network which displays a complex organization with localized brain areas specialized for specific brain processes (Price, 2012). For example, phonological processing seems to be localized in pSTS (Liebenthal et al., 2005; Chang et al., 2010; Mesgarani et al., 2014), while structure-related computations appears to be localized in inferior frontal and posterior temporal regions (Pallier et al., 2011). The different levels of the processing hierarchy therefore seem to rely on different brain modules. Neuroimaging studies have revealed in the past decades that the infant brain initially displays a very similar cerebral architecture with specific areas involved in specific tasks. One can therefore speculate that this modularity together with the wealth of long range connections between the various processing modules and with associative areas might already support a hierarchical processing of speech.

Besides, two mechanisms have classically been proposed to account for this multidimensional knowledge acquisition (Marcus, 2000), and these mechanisms can apply at any level of the hierarchy. First, the human brain appears to benefit from a **data mining device** which aims at gathering a wealth of information, and which can “habituate” to statistically dominant features of the input. This is for example the case when infants process the prosodic features of speech and habituate to the rhythmicity of their native language; they are thereafter surprised when presented with a language belonging to a different rhythmic class (see section 1.2). Similarly, when processing the sound units of speech, the learning system accumulates the frequency of occurrence of various sounds, and transition probabilities from one sound to the next (see section 1.3.2). Some highly frequent sequences emerge, allowing the subsequent discovery of words. This learning mechanism is often referred to as **statistical learning**. However, this system is rather passive and could only account for estimating dominant features, and discriminating elements that deviate from the accumulated regularities. Empirical evidence has demonstrated that infants’ abilities are far more impressive. Besides, as pointed out by Chomsky (Chomsky, 1980) and others, the linguistic input is highly ambiguous, and without representational constraints, infants could not move along the abstraction ladder. An additional system is therefore needed for **deriving actual knowledge** from the input. For example, experimental studies with bilingual infants have indicated that they do not simply accumulate global regularities of the acoustic features from their linguistic environment; they rather construct two separate representational systems for the two languages, and analyze each language separately (Bosch & Sebastian-Galles, 2001; Molnar et al., 2014). Similarly, when extracting highly frequent sound sequences, infants move beyond a simple recognition and readily infer that the segmented string has a referential role. Amazingly, from just a few exposures, the referential value of a word is easily applied to more than the experienced referents, and generalized to the abstract category (Waxman & Markow, 1995; Fulkerson & Waxman, 2007; Ferry et al., 2010; Balaban & Waxman, 1997). Both of these **statistical** and **abstract** learning systems are therefore both necessary and they complement each other: abstract knowledge is often derived from the output of the data mining system, and the output of the abstraction system undergoes statistical analyses. Dawson & Gerken (2012) formalized these two systems by comparing the first to an **associative learner** similar to what is modeled by Hebbian associative networks, and the second to a **rational learner** of the sort that occurs in Bayesian probabilistic inference.

Infants represent a highly challenging population for research, as the experimenters cannot directly ask participants to perform a task, and infants cannot provide verbal reports of what they perceive. This is why infant language researchers have developed ingenious tricks to investigate early linguistic abilities, but they mainly evidence **discrimination** behaviors which could be accounted for by the associative learner. On the other hand, **generalization** paradigms have provided a window to observe the acquisition of abstract representations. Applying this methodology in adults, Peña et al. (2002) could evidence the two mechanisms. The authors first asked whether participants could track non-adjacent dependencies from a monotonous stream of syllables in order to segment the embedded trisyllabic AxC words (the first syllable predicts the third, and the second syllable varies). The authors showed that the statistical learning process could indeed detect the long-distance regularities between A and C syllables, discriminating between words (AxC) and part-words (CA'x). However, participants could not generalize over the x syllable (i.e. to novel instances of the A_C dependency), indicating that they were sensitive to the statistics of the familiarization material, but that they had failed to represent the underlying A_C dependency. The authors thereafter inserted subliminal pauses of 25 ms between words, so that these prosodic-like subtle cues would relieve learners from statistical computations, allowing them to actually encode the long distance dependencies at a more abstract level. After familiarization to this novel sequence, participants indeed generalized the A_C dependency to novel intervening x syllable, preferring rule-words over part-words. These findings illustrate how statistical regularities can be **recoded** into a generic pattern that can be used to interpret new input.

Similarly, Marcus et al.'s study (1999) (see section 1.4.2) illustrated the interplay between the two processing mechanisms. Infants had to encode syllable sequences at an abstract level, i.e. in terms of identity relations between syllables. Then they had to perform statistical analyses of the familiarization material at this abstract level of description, and habituate to the ABB or ABA pattern. At test, the novel structure, processed at the abstract representation level, therefore yielded a surprise response. These findings evidenced that infants could easily access some abstract features of speech. The authors even pushed their results a bit forward and claimed that infants had extracted the "algebraic rule" governing the syllable sequences. Actually, the reported discrimination behavior does not necessarily reflect the explicit recoding of the input into an abstract representation of the governing rule (McClelland & Plaut, 1999); such ABB - ABA discrimination could rather reflect a simple dishabituation from immediate repetition to no repetition in final position. The authors nonetheless raised a particularly interesting hypothesis. While any animal brain appears to benefit from a more or less powerful data mining device – which could indeed be easily implemented through Hebbian associative neural networks – the specificity the human brain might be rooted in the ability to abstract knowledge though an efficient recoding of the input complexity into a **mental symbolic code**. Indeed, infants may have already access to a symbolic system to mentally represent and manipulate the various aspects of the external world, allowing information compression and efficient memory storage. Early word-learning studies have indicated that, very early on, infants can map arbitrary verbal labels onto categories of objects (Waxman & Markow, 1995), suggesting that they already possess this symbolic system. Speech in particular might be a privileged gate towards symbols (Ferry et al., 2010; Fulkerson & Waxman, 2007). The cognitive advantages (Feigenson & Halberda, 2008; Xu, 2002) that stem from labelling objects additionally indicate that this system mediates enhanced processing abilities. Besides, Gerken et al. (2015) recently evidenced structural generalization in 9-month-old infants after hearing a single trisyllabic speech string, without any habituation. These results suggested that infants readily recode the input into an abstract representation of its structure. The exact nature of these representations is still unclear, but we might speculate that it benefits from the attributes of symbolic representation.

1.6 Electrophysiology of language acquisition

The recent emergence of imaging techniques allowed cognitive researchers to inspect the living brain at work, widening not only the range of methods but also the scope of questions that experimental psychologists could address. However, brain imaging techniques remain sensitive to artifacts, and experiments typically need to extend in time in order to obtain decent signal-to-noise ratio and statistical power. Because of these limitations, the use of sophisticated imaging procedures is scarce in developmental research where participants' movements cannot be avoided and their patience is limited. Electroencephalography appears as a reasonable trade-off as it is **non-invasive**, **painless**, and it only requires a **simple** recording apparatus. Besides, this technique is less sensitive to movement artefacts than fMRI, and thus better suited for studying awake infants. The electroencephalogram reflects the summation of synchronous electrical activity of thousands of cortical neurons sharing similar spatial orientation. During post-synaptic depolarization, the negativity at the apical dendrites and the positivity at the cell body create a tiny dipole. When thousands or millions of such dipoles occur at approximately the same time, in close vicinity and sharing the same orientation, the resulting dipole can be measured at the scalp. The amplitude of the measurement at any given point of the scalp surface thereafter depends on the position and the orientation of the underlying generators, as well as on the resistance and shape of the conductive medium. Infants typically display higher skull conductivity compared to adults, and their cranial bones are unfused and separated by fontanels and sutures for at least several months after birth to accommodate birth and the growth of the brain. These differences result in higher voltage amplitudes, and different topographies compared to what is reported in the adult literature. It additionally raises multiple challenges to solve the inverse problem (i.e. localizing the cortical generator sources from the EEG scalp recordings), that are yet to be addressed.

The **placement of electrodes** on the scalp conventionally follows the international 10/20 system (Jasper, 1958), in which electrodes are regularly distributed along the anterior-posterior and lateral axes. The electrodes in this system are labelled according to the cerebral lobes that are thought to lie under the location, according to adult anatomy. A few studies have documented some differences in infants (Blume et al., 1974; Hellstöm et al., 1963), but the pediatric literature lacks precise description of the anatomical relations between the external 10/20 sensors and the underlying brain structures, and their reliability. The 10/20 system has been further extended to high-density electrode nets of 128 electrodes which provide a broad coverage of the scalp surface, and which have become popular in infant research.

The ongoing EEG activity suffers from **poor spatial resolution** compared to hemodynamic measures, but it benefits from a **higher temporal resolution**. EEG recordings indeed capture a wealth of temporal dynamics, and there are limitless possibilities for analyzing electrophysiological data. Neurophysiologists constantly come up with fancy and sophisticated analysis methods and physiological interpretations, adapted to their research questions and exploiting the richness of the recorded data. Due to the difficulties of infants' testing, developmental electrophysiology is still lagging behind what it being developed in adults. **Event Related Potentials** (ERPs) are traditionally used to study cognitive development. Participants are repeatedly presented with a given stimulation, and the ERP is calculated as the average of the EEG responses, locked to the onset of the stimulation, for each electrode site (Luck, 2005). This averaging procedure reduces the ongoing electrophysiological noise in order to recover the neural response to a particular stimulation (Figure 1.2A-B-C). The resulting averaged ERP waveforms consist of a temporal sequence of positive and negative voltage deflections over clusters of electrodes, which are called waves, or components, and which reflect the different processing stages. Cognitive abilities are thereafter questioned by contrasting different experimental conditions and assessing the significance of the observed component differences. If the differences are significant, one can conclude that the different stimulation types are processed differently. While behavioral methods only provide information about the final outcome of cognitive processes, EEG reveals a continuous measure of processing over the scalp, making it possible to determine which spatio-temporal components are affected by a specific experimental manipulation.

Another dimension of the EEG signal which has gained tremendous interest in the past decades is its **oscillatory content**. Indeed, the ongoing cerebral activity contains rich rhythmic fluctuations reflecting either the **spontaneous variations** in the excitability of the underlying neural populations, or **oscillatory bursts** in response to the incoming sensory inputs. The signal can be represented in the frequency domain through signal-processing techniques, and is thereafter described by its amplitude and phase (Figure 3.1D-E-F). This frequency dimension offers another window on cognition, as it can reveal brain processes that classical ERP approaches cannot capture. Besides, it has been convincingly demonstrated that the activity in various frequency bands play a functional role in several cognitive processes, such as attention, time perception, perceptual binding, learning or memory (X. Wang, 2010). It is still unclear how the mechanisms primarily described in adults transfer to infants, but a few studies have evidenced the role of gamma evoked oscillations in visual object processing (Gliga et al., 2010; Southgate et al., 2008; Kaufman et al., 2005). Further studies are needed to question the functional role of other frequency bands in early cognition. Brain oscillations can also be recorded in response to a rhythmic extrinsic driving sensory factor oscillating at a precise frequency (Picton et al., 2003; Norcia et al., 2015). Indeed, cerebral activity is then **entrained** by the stimulation, and it translates into a strong narrow-band increase in EEG spectral power, or phase alignment at the stimulation frequency. This effect is also referred to as **steady-state evoked potential** or **frequency tagging**. This technique allows to tag the early sensory processing of a specific stimulus, with a very high signal-to-noise ratio.

A lot of the existing electrophysiological studies on early linguistic abilities use traditional ERP analyses. Most experiments target auditory processing using classical oddball ERP paradigms, tracking mismatch auditory responses, elicited by a change of phoneme (Dehaene-Lambertz & Pena, 2001; Dehaene-Lambertz & Dehaene, 1994; Dehaene-Lambertz & Baillet, 1998), or by non-native linguistic features (Friederici et al., 2007). A few experiments have also investigated early syntactic processing in toddlers, evidencing a late posterior positivity in response to ungrammatical sentences (Brusini, Dehaene-Lambertz, Dutat, et al., 2016; Brusini,

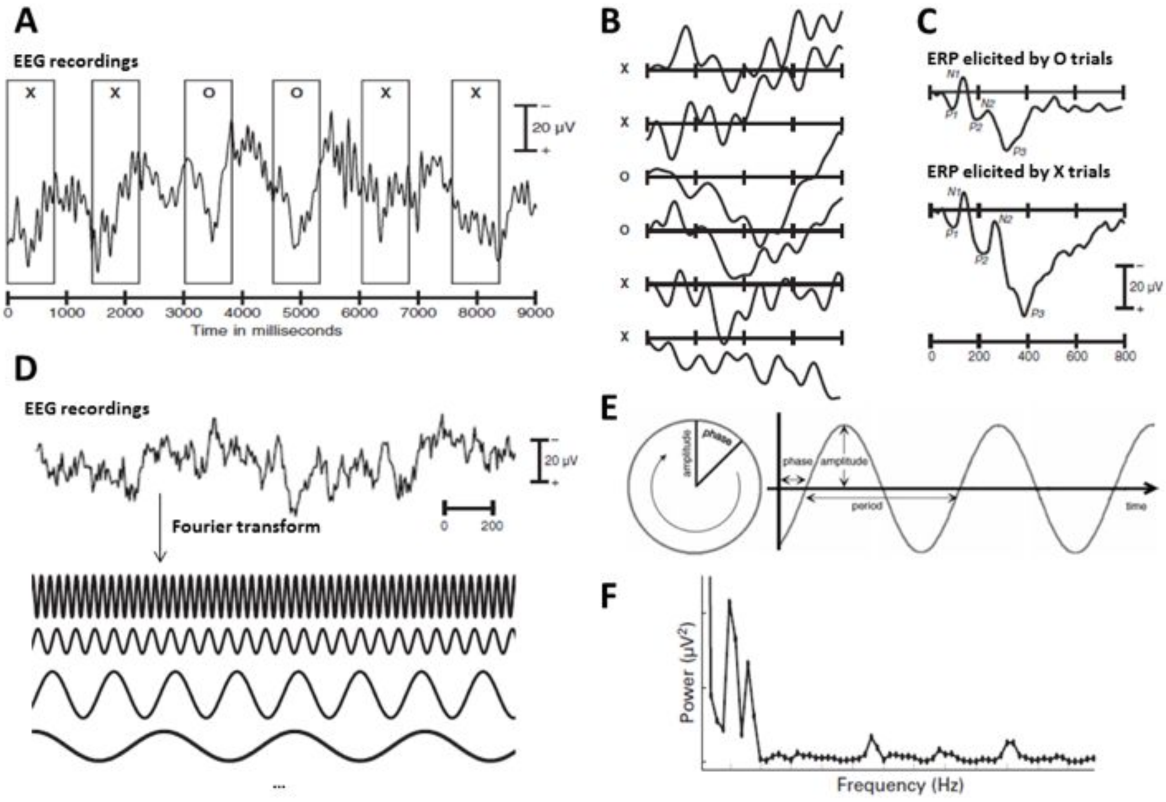


Figure 1.2: **ERP and frequency analyses of the EEG signal A**: The continuous EEG recordings at one electrode site, marked with X and O events, over a 9-seconds period. **B**: The EEG signal is segmented around the events of interest, here X and O events. **C**: The EEG segments are thereafter averaged across similar event types, resulting in ERP responses. **D**: Schematic representation of a frequency analysis: the EEG signal can be decomposed into a set of sinusoids oscillating at different frequencies, with a Fourier transform. **E**: These sinusoids are described by their amplitude and phase. **F**: The power spectrum describes the contribution of each frequency to the original signal. Each point represents the power (i.e. the square of the amplitude) of the sinusoid oscillating at frequency f.

Dehaene-Lambertz, van Heugten, et al., 2016; Bernal et al., 2010; Oberecker et al., 2005; Oberecker & Friederici, 2006). The acquisition of statistical dependencies has additionally been investigated using ERP methodology (Teinonen et al., 2009; Friederici et al., 2011). Such paradigms document infants' early abilities, but they only indirectly address the underlying learning or encoding neural mechanisms. The immature brain is a particularly challenging system to study, as electrical components might be less precisely time-locked to external stimuli depending on the age of the participants, and the EEG signal contains more large amplitude and low frequency endogenous noise (Chu et al., 2014). However, it represents a **promising path** of research for understanding human cognition in general.

1.7 Outline of the thesis

The experimental work presented in this dissertation aimed at exploring the role of both associative and rational learning devices, in the encoding of minimal speech sequences. We additionally questioned the nature of the extracted knowledge, in the framework of language acquisition and from an electrophysiological perspective. We exploited the advantages of EEG paradigms over behavioral ones to complement previous infant studies.

First, this thesis starts with a **methodological contribution** to the developmental imaging community (Chapter 2), in which we assessed the reliability of the 10-20 electrode placement system, initially developed for adult research, in young infants. Based on image processing techniques, this work aimed at quantifying the inter-subject variability of this external sensor placement system, widely used for both EEG and NIRS recordings in developmental research. We additionally provided a brain atlas, reporting anatomical parcellation over an infant reference template brain, which can be useful for studies in which individual anatomic MRIs cannot be obtained.

In a first **cognitive EEG experiment**, we thereafter investigated whether infants could track long-distance dependencies between remote surface features of a continuous speech stream, in order to discover the embedded words (Chapter 3). Transferring some aspects of Peña et al.'s experimental design to infant EEG recordings, we investigated the neural correlates of the two mechanisms previously described. Cerebral activity was recorded during the familiarization period, allowing us to track brain processes at play during **online extraction of the statistical regularities** of the speech stream. Capitalizing on the high temporal precision of EEG recordings, and the rhythmicity of the familiarization stream, we measured neural entrainment as an index of the extraction of the successive linguistic units. In a subsequent test phase, following Peña et al.'s study (2002), we questionned infants' ability to encode the extracted regularities as a unified generic pattern, allowing them to generalize to novel legal sequences. This study was conducted with 8 month-old full-term infants, an age at which they can already track transitional probabilities between adjacent syllables from an artificial speech stream (Saffran et al., 1996). Their performances were compared with 2 groups of pre-term infants, matched for maturation (comparable neural age from conception, but longer ex-utero linguistic experience), or for exposure to a linguistic environment (lower neural age, but comparable ex-utero linguistic experience), in order to investigate the relative influence of neural maturation and linguistic experience on the development of both statistical learning and generalization abilities.

In a final set of **EEG experiments**, we focused more specifically on the **encoding of structural regularities**, and began to question whether the encoded representations possess the required attributes to enter a symbolic system (Chapter 4). Building on Marcus et al.'s study (1999), we investigated whether 5 month-old infants can perform beyond mere discrimination skills between two syllable sequence types, but can rather recode the sequences into generic patterns that are thereafter available for further computations. More specifically, we investigated infants' abilities to track abstract regularities in order to categorize different speech strings based on the accumulated regularities. We assessed whether the extracted linguistic representations could be simultaneously monitored, and paired with an arbitrary visual or auditory item. We measured participants' learning performances directly from cerebral recordings, in order to circumvent any executive control issue. From a methodological perspective, we exploited the richness of EEG data, combining different signal processing approaches to observe how various levels of the processing hierarchy are affected by the acquired abstract representations.

Finally, we closed this thesis with a general discussion of the reported results, and some perspectives to complement these studies (Chapter 5).

CHAPTER 2

ANATOMICAL CORRELATIONS OF THE INTERNATIONAL 10–20 SENSOR PLACEMENT SYSTEM IN INFANTS

Abstract

Developmental research, as well as paediatric clinical activity crucially depend on non-invasive and painless brain recording techniques, such as electroencephalography (EEG), and near infrared spectroscopy (NIRS). However, both of these techniques measure cortical activity from the scalp without precise knowledge of the recorded cerebral structures. An accurate and reliable mapping between external anatomical landmarks and internal cerebral structures is therefore fundamental to localize brain sources in a non-invasive way. Here, using MRI, we examined the relations between the 10-20 sensor placement system and cerebral structures in 16 infants (3-17 weeks post-term). We provided an infant template parcelled in 94 regions on which we reported the variability of sensors locations, concurrently with the anatomical variability of six main cortical sulci (superior and inferior frontal sulcus, central sulcus, sylvian fissure, superior temporal sulcus, and intraparietal sulcus) across these infants. The main difference between infants and adults was observed for the channels O1-O2, T5-T6, which projected over lower structures than in adults. We did not find any asymmetry in the distances between the scalp and the brain envelope. However, because of the yakovlean torque pushing dorsally and frontally the right sylvian fissure, P3-P4 were not at the same distance from the posterior end of this structure. This study should help to refine hypotheses on functional cognitive development by providing an accurate description of the localization of standardized channels relative to infants' brain structures.

2.1 Introduction

The recent development of non-invasive brain imaging techniques has boosted research in cognitive development. Electroencephalography (EEG) and near-infrared spectroscopy (NIRS) are particularly convenient when it comes to neonatal/paediatric brain recordings. Both of these techniques rely on an external placement of the recording sensors; an accurate description of the relations between the external anatomical landmarks and the internal cortical structures is therefore of crucial importance to draw robust interpretations from the recorded activity. It is especially true for NIRS, as it records cortical activity in the crescent of light between a laser emitter and photodiode detectors. Thus, a misplacement of the sensors relative to the cerebral structure of interest can lead to erroneous conclusions.

The international 10-20 system for electrode placement was originally developed to place EEG electrodes on the scalp in a reproducible manner from one recording to the next (Jasper, 1958). This standardized electrode positioning system is based on external landmarks, and a regular spacing between electrodes. It assumes a consistent relationship between scalp locations and underlying cerebral structures. The validity of this assumption has been demonstrated in adults (Homan et al., 1987; Jasper, 1958; Okamoto et al., 2004). However, very few studies have been conducted to tackle this issue during brain development, with only one post-mortem study in 6 infants, younger than 4 months of age (Blume et al., 1974), and a skull X-ray study in 28 infants between one week and thirteen months of age (Hellstöm et al., 1963).

This last study, which demonstrated a fixed location of the 10-20 system relative to fontanella and sutures, relied on the hypothesis that brain structures were also aligned to these skull markers. Yet, the inhomogeneous growth of the different cerebral lobes (Gilmore et al., 2007), the increase of the slope of the sylvian fissure during childhood (Sowell et al., 2002) and the operculation of the inferior frontal region observed during the first post-natal year are some examples of developmental changes that may affect the relations between brain structures and the 10-20 standardized scalp locations. Furthermore, the head shape may vary more during the first months of life than later on due to birth events and sleeping habits that may flatten one side of the head. Two potential sources of inter-subject variability may thus overlap: the external variability of sensor positioning and the internal structural variability.

In this study, we provided a broad description of the cranio-cerebral relationships of the 10-20 standard positions during the first 4 post-natal months, a time of fast developmental changes (i.e. brain volume doubles between birth and 6 months of age), with two distinct approaches to quantify both external and internal variability. We used MRI data, which give access to both external landmarks and cerebral organization, in a cohort of 16 healthy infants and we worked on 3-D reconstruction of the infant's heads and brains thanks to specific algorithms developed in the BrainVisa software (Cointepas et al., 2001) allowing realistic computations and visualization of the relations between external and internal landmarks.

We first virtually placed electrodes over infant heads following the standardized 10-20 placement rules. Second, we choose one infant as representative of the group and projected on her the location of the individual electrodes localization after having normalized each infant anatomical image towards this template. Third, we specifically labelled 94 cortical regions (47 on each hemisphere) in our template infant adapting the MNI-space anatomical parcellation proposed for the adult brain by Tzourio-Mazoyer et al. (2002). We were thus able to analyze the electrode placement variability relative to the underlying cortical regions. Fourth, we examined the brain structural variability across our group and computed the main sulcal patterns distribution, to analyse cortical structures variability with respect to the 10-20 system. Finally, we reported electrode-brain distances since NIRS measurements are particularly sensitive to the depth of the cortical surface from the head scalp. Our description should provide an accurate view of the variability of standardized electrode locations over the scalp, and of their relationship with underlying cerebral structures in infants. It also provides the community with an anatomically defined infant atlas in order to study and describe cortical activity.

2.2 Methods

2.2.1 Subjects

Sixteen healthy full-term infants (mean maturational age, that is, chronological age corrected for the gestational age at birth: 9.0 ± 3.6 weeks, range: 3.4-16.3 weeks; 11 boys, 5 girls), were included in this study after their parents gave written informed consent.

2.2.2 Data acquisition

Infants were naturally asleep during MR imaging (no sedation was used). Particular care was taken to minimize noise exposure, by using customized headphones and by covering the magnet bore with special noise protection foam. The study was approved by the regional ethical committee for biomedical research.

T1 and T2 weighted images covering the whole brain were acquired on a 3T MRI system (Tim Trio, Siemens Medical Systems, Erlangen, Germany) using a 32-channels head coil. To minimize specific absorption rate (SAR) and noise exposure, we used radio-frequency (RF) impulsions with "no SAR", and the "whisper" gradients mode. The total acquisition time was 5min32s (T1w=2min48s; T2w=2min44s). T1w images were obtained with a 3D fast gradient recovery sequence (MPRage, TE/TR/TI=4.25/1100/2000ms, parallel imaging GRAPPA reduction factor 2, partial Fourier sampling factor 6/8). Sagittal slices were acquired with a spatial resolution of 1mm isotropic (field of view=192mm; acquisition matrix=192×192, no interpolation at reconstruction; slice thickness=1.1mm; 176 slices). T2w images were obtained with a 2D turbo spin echo sequence (TSE, TE/TR=149/4500ms, 4 concatenations, parallel imaging GRAPPA reduction factor 2). Axial slices were acquired with a spatial resolution of 1mm isotropic (field of view=192mm; acquisition matrix=192×192, no interpolation at reconstruction; slice thickness=1.1mm; 92 slices).

2.2.3 T1w versus T2w images

In this study, we distinguished the variability for sensor placement over the skull from the variability of inner cerebral structures. At this age, the contrast between white and gray matter is weak in T1w images, and T2w images are preferred to analyse brain structures (Barkovich, 2000). By contrast, T1w images provide better information about the head shape, with a good contrast for the fat of the skin. We thus used T1w images to extract head shape and study sensor placement over the head, and T2w images to examine cortical organisation. T1w images and T2w images were registered to each other in each infant and depending on the analysis, we used one or the other sequence.

2.2.4 Template and atlas definition

Among our infants, we chose a 7.1 week-old girl as a template, because her age was close to the population's mean age and her head was regular and symmetrical. We checked that the head was symmetrically positioned in the head coil to avoid that cerebral spinal fluid (CSF) settles on one side. We choose to adapt the Tzourio-Mazoyer et al.'s atlas (2002) to the infant's brain. We adapted this particular MNI-space anatomical parcellation of the adult brain because, on the one hand, it provides a standardized anatomical labelling that is widely used notably in different softwares (e.g. SPM and Brainstorm), and on the other hand, it relies on identification of primary and secondary sulci, which are already clearly visible in the newborn's brain. Instead of manually drawing sulci landmarks on axial slices as in it was done in Tzourio-Mazoyer et al. (2002), we benefited of the 3-D reconstruction of the grey-white matter interface and of the automatic recognition of the sulci through the BrainVisa pipeline (Cointepas et al., 2001) to semi-automatically draw the ROIs on the brain surface.

More specifically, we performed the following steps. The first and most difficult step in infant is to obtain a correct reconstruction of the grey-white matter interface. The inner cortical surface was segmented using a semi-automatic segmentation pipeline dedicated to T2w MRI images of the infant brain (validation of this pipeline can be checked in (Leroy, Mangin, et al., 2011)), followed by manual correction, when local misplacements occurred. As maturation progresses, the T2w signal becomes darker and the contrast between the white and the grey matters decreases. Inaccuracies are thus likely to occur in the fast maturing regions (e.g. calcarine and central sulci). As the human eye is better to follow the cortical ribbon, these regions were carefully checked to correct misplacements. We thus obtained a 3D reconstruction of the inner cortical surface, over which we performed an anatomical parcellation (Figure 2.4). Through a BrainVisa pipeline (Cointepas et al., 2001), primary and secondary sulci were automatically extracted and labelled (Figure 2.1). Once a correct grey-white matter interface is obtained, this step does not raise specific difficulties in infants as compared to adults. Infants' sulci are generally simpler than adults due to the fact that the tertiary gyration is just starting at this age. The sulci were visually checked and relabelled when necessary, directly on the 3D ribbon-like representations of the sulcal patterns – using Anatomist (Riviere et al., 2000).

On the basis of the segmented sulci, we delineated 47 anatomical regions of interest (ROIs), in each hemisphere. As explained in Tzourio-Mazoyer et al. (2002), almost all ROIs were delineated by sulcus fundi. Thus CK delineated the parcels on the 3-D reconstruction of the grey-white matter interface using Anatomist drawing tool dedicated to brain folding patterns (Le Troter et al., 2012). This tool automatically extracts the line of the sulcal fundi on the mesh of any cortical surface by following the shortest path between any two points with a negative mean curvature. When necessary, T2 slices aligned on the 3D mesh using the Anatomist software (Riviere et al., 2000) were also considered to resolve ambiguous situations. When specific distances to a given structure were used to define a brain region in the adult's atlas (e.g. supplementary motor area), a 1.4 ratio was used to decrease this distance according to the size ratio between the adult and infant brains (Dehaene-Lambertz et al., 2002). Each region was reviewed (FL, CK and

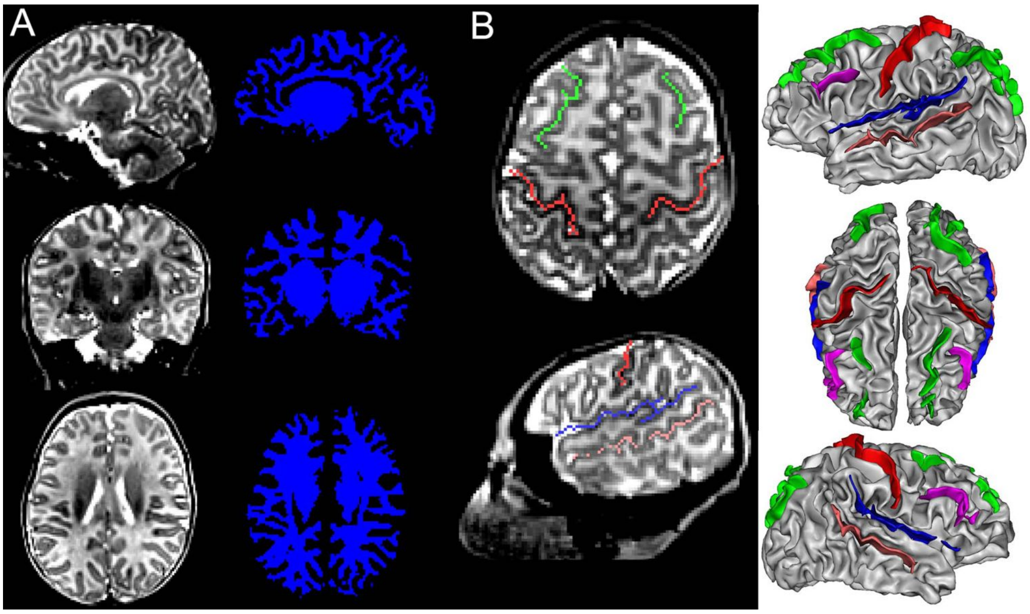


Figure 2.1: **Cortical segmentation and sulci extraction A**: Template infant's T2w images in sagittal, coronal and axial sections with the corresponding white matter mask, segmented using the semi-automatic segmentation pipeline (Leroy, Mangin, et al., 2011). **B**: Template infant's T2w images in axial section with the projection of the left and right central sulci and in sagittal section with the projection of the left superior temporal sulcus. Sulci were computed from a dedicated BrainVisa algorithm (Cointepas et al., 2001), and manually checked and relabeled when necessary.

GDL), and in difficult cases, a consensus was obtained.

For practical reasons, two anatomical regions (pallidum and hippocampus,) among the 45 defined in Tzourio-Mazoyer et al. (2002) were not included in our parcellation, because T2w images contrast and resolution were too poor at these locations. We could also not delineate a region labelled "inferior parietal cortex" in the right hemisphere of our template infant. It was defined by Tzourio-Mazoyer et al. (2002) as the remaining part of the inferior parietal lobule that was not belonging to the supramarginal or to the angular gyrus. This observation was template specific, and should not be considered as a general fact about infant neuroanatomy. We brought some additional refinements to the atlas defined in (Tzourio-Mazoyer et al., 2002), indeed, the temporal regions lining the inferior plane of the sylvian fissure cannot be considered as homogeneous. First, we labelled Heschl's gyrus, being the most anterior gyrus on the upper surface of the temporal lobe. Thus gyrus is bordered posteriorly by Heschl's sulcus. The planum temporale was then defined as the triangular structure lying on the upper surface of the temporal lobe, inside the sylvian fissure, its anterior limit being Heschl's sulcus. For the definition of the posterior border of the planum temporale, a variety of strategies have been developed in the past. Consistently, we defined the posterior limit of the planum temporale as either the last coronal section before the temporal and parietal cortex fuse, either medially or laterally, or the last coronal section before a change in the slope of the continuous plane characterising the planum temporale. Any further territory on the floor of the sylvian fissure posterior to that limit was labelled as planum parietale, while the region on the floor of the sylvian fissure anterior to the Heschl's gyrus, and posterior to the superior temporal pole (defined in (Tzourio-Mazoyer et al., 2002)). The proposed parcellation (Figure 2.4) is a single-subject atlas as several atlases proposed in adults (Talairach & Tournoux, 1988; Tzourio-Mazoyer et al., 2002), and is therefore specific to our template (i.e. a 7.1 week-old infant) but as it kept the adequate resolution of single subject, it can be used to reconstruct sources in a template brain, as proposed in Brainstorm (Tadel et al., 2011) for example.

2.2.5 External 10-20 sensors placement

The international 10-20 standard positions were virtually set on each individual infant head reconstruction using a semi-automated algorithm (Figure 2.2-A). The infants' head volumes were first extracted from the T1w MRI images, and reconstructed in 3D. Then, the four standard fiducial positions (nasion, inion, left and right tragi) were located by CK over the head 3D reconstruction, with the help of T1w MRI volumes, based upon external anatomical considerations. First, the nasion was defined as the distinctly depressed area directly between the eyes, just

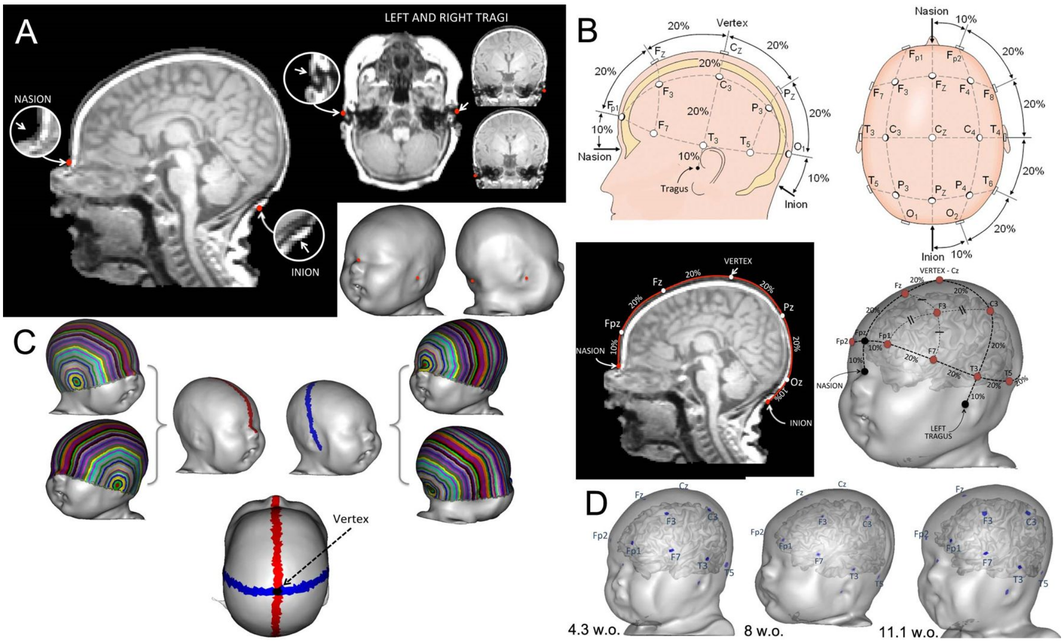


Figure 2.2: **The international 10-20 system for electrode placement. A**: The four anatomical reference points presented on one individual infant, selected to be the template. **B**: The international 10-20 standard cranial positions as defined in (Jasper, 1958) and reported on the template infant. **C**: An algorithm based on geodesic distance maps was used to automatically define electrode positions. Here is shown how the vertex was defined on the template brain. **D**: Example of electrode automated localization on three different subjects of different ages (4.3-week-old on the left, 8-week-old on the center, 11.1-week-old on the right).

superior to the bridge of the nose. The inion was the lowest point of the skull from the back of the head, and we systematically located this point on the inferior side of the external occipital protuberance. Regarding the left and right pre-auricular points, originally used for the lateral measurement in the 10-20 system, we choose to replace them with the upper limits of the left and right tragi, as it is actually often the case in clinical situations, due to difficulty in their detection. It has been shown that this subtle difference in pre-auricular point definition has a very limited impact on the 10-20 standard positions (Jurcak et al., 2007). As a result, for each ear, the tragus was defined as the upper point of the small pointed eminence of the external ear.

The four fiducial positions were then used to generate a system of geodesic lines that run across the head and intersect at intervals of 10% and 20% of their total length following the standardized 10-20 description (Figure 2.2-B). An iterative algorithm following these definition rules was designed to automatically locate the 19 sensors in each infant. The geodesic distance maps were computed from the landmarks over the skull, using BrainVisa image processing tools (Cointepas et al., 2001). We defined two sets of points with these distance maps: those equidistant from the left and right tragi, and those equidistant from the nasion and the inion. The vertex was then computed as the intersection of these two sets of points (Figure 2.2-C).

The line joining the nasion and the inion through the vertex was computed as the intersection of the plane defined by these 3 points and the baby's skull. A geodesic distance map was computed along this line, and we set the frontal sensor Fz and the parietal sensor Pz along the geodesic line at 30% and 70% of the total distance, from the nasion, respectively; additionally the midline points Fpz and Oz were located at 10% and 90% of the total distance. Similarly, the line joining the two tragi through the vertex was computed as the intersection of the plane defined by these 3 points and the baby's skull, and the four points T3, T4, C3, C4 were set along the left-right tragi geodesic line (at respectively 10%, 90%, 30% and 70% of the total length). Again, the line joining the midline points Fpz and Oz through T3 was computed as the intersection of the plane defined by these 3 points and the baby's skull, and the four transversal points Fp1, F7, T5 and O1 were located along this line on the left side (at 10%, 30%, 70% and 90% of the entire distance away from Fpz); and the same procedure was conducted on the right side to mark Fp2, F8, T6 and O2. Eventually, we defined the four points F3, F4, P3 and P4 along geodesic lines at the equidistant locations from the neighbouring sensors. The sensor locations were then stored as a binary mask for each infant, that is a volume with one voxel per sensor location set to 1, and the other voxels set to 0.

2.2.6 Brain structures segmentation

In order to investigate the variability of the inner brain structures at the population level, we identified some relevant cortical landmarks over the infant group, and modelled the sulcal presence probability and inter-subject variability. Because of both sulcal variability and potential local inaccuracies in cortical segmentation of the most mature and twisted parts of the brain (e.g. the calcarine sulci), we restricted our set of sulci to the most reliable ones, namely, the lateral sylvian fissure (SF), the central sulcus (CS), the superior temporal sulcus (STS) and its ascending branch, the superior frontal sulcus (SFS), the inferior frontal sulcus (IFS) and its anterior part, and the intraparietal sulcus (IPS). The sulcal patterns extraction algorithm crucially depends on the quality of the MRI images, and it failed for one infant, who was thus removed from this analysis. For each infant, the extraction of the inner cortical surface was performed using an automatic segmentation pipeline, dedicated to T2w MRI images of the infant brain (see (Leroy, Mangin, et al., 2011) for a validation of this method in infants of the same age than here). For all infants, a neuropaediatrician (H.S.) visually checked and manually corrected when local misplacements occurred, and each segmented brain was thereafter reviewed (H.S. and F.L.). We obtained the interface between cortex and white matter for each hemisphere, saved as binary images, as well as 3D reconstructions. We identified sulci through a BrainVisa pipeline (Cointepas et al., 2001), providing a first automatic extraction and labelling, from the grey-white interface reconstruction. A manual checking and relabelling was performed when necessary by two of the authors (CK and FL). For difficult cases, a consensus was obtained. The manual labelling was performed directly on the 3D meshes – 3D ribbon-like representations of the sulcal patterns – using Anatomist.

2.2.7 Normalization

As mentioned before, we distinguished two approaches for non-linear normalisation in the two following sections: based on either the T1w images, when we intended to align preferentially the skull shapes for the measures of inter-subject sensors variability, or the T2w images, when the alignment of the cortical structures was particularly at stake for inter-subject sulcal mapping. For each subject, we non-linearly normalized the individual T1w (or T2w) image onto the single-subject template T1w (or T2w) using the Statistical Parametric Mapping toolbox in matlab (SPM8), and visually checked the match of each individual image to the template.

2.2.8 Inter-subject sensors variability

We applied the T1w normalization matrix to the sensor binary images, using the nearest neighbour interpolation in each infant. Then the registered binary masks were averaged over the template head and smoothed using a small Gaussian filter (filter size: 1.1mm, isotropic). It produced an “average” sensor net made of 19 regular regions – one for each sensor.

For a more accurate description of the sensors’ locations with respect to the underlying brain structures, we projected the set of the registered sensors onto the template brain envelope. For the projection, we defined sensors as spheres centred on the standard locations, with a 5mm diameter. The brain envelope is a smooth surface overlapping the crown of gyri, where folded patterns are removed. It was computed from the binary images of the grey-white matter interface: applying a morphological closing (structuring element size: 15 mm), followed by a morphological dilation (structuring element size: 2 mm). These envelope images were also reconstructed in 3D. The projection of a given sensor was defined as the closest point of the brain envelope from the average sensor. A Gaussian smoothing (filter size: 2.2mm, isotropic) was then applied to the projected set of average sensors to regularize the distribution estimate. Over the template brain envelope, we delineated two iso-surfaces for each of the 19 sensors, one containing 90% of the population and the other 50% (Figure 2.5-B). These isosurfaces were computed, for each electrode, by gradually integrating the distribution estimate from the highest likelihood to the lowest. We additionally computed the standard deviation of the spatial distribution for each sensor, and along the three directions.

2.2.9 Inter-subject sulcal mapping

To examine the localization and variability of the inner cerebral structures, we built an atlas of sulcal patterns at the population level, modelling sulcal presence probability across infants. For each sulcus label, we built frequency maps, in which each voxel stores the number of subjects whose sulcus was identified at this location. We applied the T2w normalization matrix to the sulcal binary images, using the nearest neighbour interpolation in each infant to project each map in our reference template. To reduce the dispersion of individual sulcus location due to our small population size (15 infants), we performed an isotropic Gaussian smoothing (filter size: 1.7mm, isotropic) on each sulcus. We then estimated the spatial probability of finding a sulcus in a given voxel by dividing each voxel by a normalization factor: the sum of all the voxels of the 3D space. This normalization ensures that the probability integrates to 1 over the spatial 3D space. We obtained sulcal SPAM (Statistical Probabilistic Anatomy Map) (Evans et al., 1994; Perrot

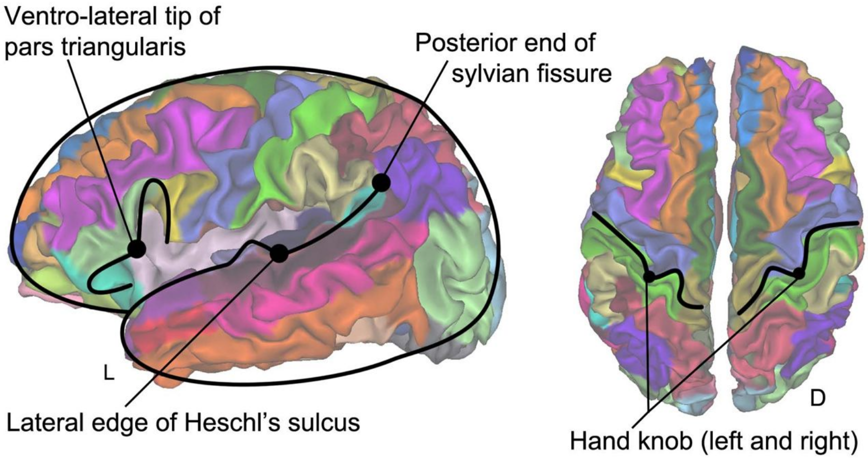


Figure 2.3: **Cortical landmarks definitions for depth measurements**, with the posterior end of the lateral fissure, the ventro-lateral tip of the pars triangularis, the lateral edge of the Heschl's sulcus, and the most posterior point of the hand knob, in the depth of the central sulcus (Left – Dorsal).

et al., 2011). We thereafter represented the sulcal models as volumes, using an isosurface corresponding to 60% of the whole probability mass, computed by integrating the probability from the highest likelihoods to the lowest (Figure 2.7-B).

The isosurface provides localisation information about the structures. However, this 3D representation comes with a main drawback: it mixes information about the size of the sulcus and the inter-subject variability. In other words, a large 3D sulcal model might either mean a large sulcus or a variable structure. To overcome this ambiguity, we defined dispersion as a “density” measurement along each sulcus, defined from the smoothed sulcal frequency maps. For each section along the considered sulcus, we defined dispersion as the ratio between the sum of the voxels in the section, and the number of non-null voxels in the section.

2.2.10 Depth measurements

We first generated a mean depth map over the skull of the template subject (Beauchamp et al., 2011; Okamoto et al., 2004), showing the distance distribution from the brain envelope surface to the skull, averaged over the 16 infants. For each infant, we computed, on non-normalized brains, the brain-scalp distance at each point of the head, using distance maps, and we saved this depth information as a volume. We thereafter registered the computed depth volumes to the template space using the non-linear transformations computed on the T1w images.

We then investigated sensor-brain distances. To address this question, we considered two different approaches to extract distances on the original images (i.e. non-normalized). First, we performed, for each infant, a basic computation of the distance between each sensor on the scalp and the closest point of the brain tissue, defined on the brain envelope from the sensor. However, often experimenter might be more interested by the distance between the sensors and some remarkable cerebral structures. Thus, in a second approach, we manually defined four cortical landmarks in each of the 16 infants (Figure 2.3) with the help of the T2w images and the grey-white interface reconstructions. We identified the most posterior point of the hand knob, in the depth of the central sulcus, the ventro-lateral tip of the pars triangularis, below the triangular sulcus, the lateral edge of Heschl's sulcus, and the posterior end of the sylvian fissure, each landmark on the left and right hemispheres. The closest sensors, and their contro-lateral equivalents were then identified for each cortical landmark, and the landmark-sensor distances were computed. Note here that the cortical landmarks were manually selected on the grey-white interface, not the envelope. Since structural asymmetries are particularly at stake in developmental studies, we thereafter examined whether landmark-sensor distances were similar on the left and right hemispheres. For each sensor left-right pair, we thus computed an asymmetry coefficient of the distance to ipsi-lateral landmark (D_{left} and D_{left}) defined as $2 \times \frac{D_{left}-D_{right}}{D_{left}+D_{right}}$, and we tested whether the left (or right) distance was significantly larger using a two-tailed paired t-test between left and right distances for each landmark. We additionally computed the distance between the tragus and the posterior end of sylvian fissure, and performed the same tests for asymmetry. These regions support asymmetric functions and it is important in brain imaging studies investigating these functions to ensure that the external sensors, in particular NIRS sensors, are not themselves located asymmetrically relative to these landmarks.

2.3 Results

2.3.1 Template atlas

The anatomical parcellation of the template brain is depicted in Figure 2.4, with the corresponding labels detailed in Table 2.1. We additionally computed the surface areas of the 47 regions of interest in each hemisphere.

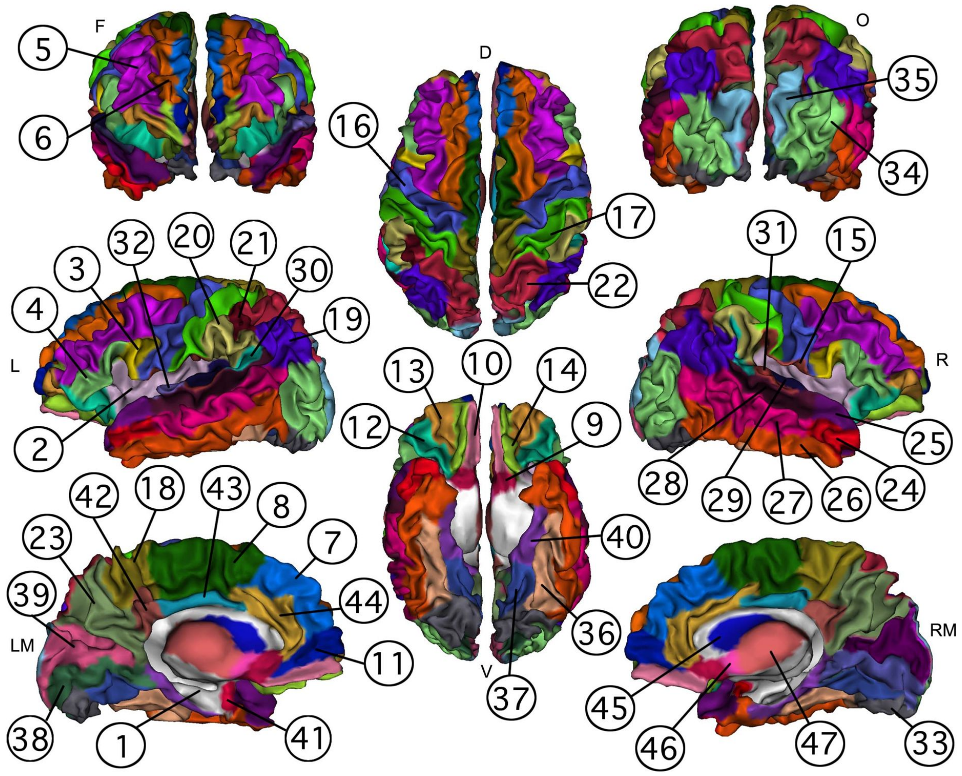


Figure 2.4: **Brain atlas.** Different views of the anatomical parcellation of the template infant's brain (from top to down and left to right: Frontal – Left lateral – Left Medial ; Dorsal – Ventral ; Occipital – Right Lateral – Right Medial).

			Left hemisphere		Right hemisphere	
	Region	Label	Surface (mm^2)	Percentage of hemispheric surface	Surface (mm^2)	Percentage of hemispheric surface
	Lateral ventricule and remaining tissues	1	3904	9.72%	3402	8.38%
	Insula	2	1037	2.58%	1044	2.57%
Frontal	**Lateral surface**					
	Inferior frontal gyrus, opercular part	3	611	1.52%	499	1.23%
	Inferior frontal gyrus, triangular part	4	862	2.15%	812	2.00%
	Middle frontal gyrus	5	2137	5.32%	2631	6.48%
	Superior frontal gyrus	6	1468	3.66%	1469	3.62%
	Medial surface					
	Medial superior frontal gyrus	7	741	1.84%	918	2.26%
	Supplementary motor area	8	1407	3.50%	1265	3.11%
	Orbital surface					
	Olfactory bulb	9	217	0.54%	193	0.48%
	Gyrus rectus	10	321	0.80%	431	1.06%
	Orbitary frontal medial	11	333	0.83%	304	0.75%
	Orbitary inferior frontal gyrus	12	673	1.68%	691	1.70%
	Orbitary middle frontal gyrus	13	502	1.25%	508	1.25%
	Orbitary superior frontal gyrus	14	519	1.29%	520	1.28%
Central	**Lateral surface**					
	Rolandic operculum	15	617	1.54%	660	1.63%
	Precentral gyrus	16	1662	4.14%	1504	3.70%
	Postcentral gyrus	17	1383	3.45%	1450	3.57%
	Medial surface					
	Paracentral lobule	18	847	2.11%	1056	2.60%
Parietal	**Lateral surface**					
	Angular gyrus	19	1316	3.28%	1721	4.24%
	Supramarginal gyrus	20	1192	2.97%	969	2.39%
	Inferior parietal gyrus	21	449	1.12%		
	Superior parietal gyrus	22	1238	3.08%	1317	3.24%
	Medial surface					
	Precuneus	23	1190	2.96%	1629	4.01%
Temporal	Middle temporal pole	24	229	0.57%	454	1.12%
	Superior temporal pole	25	641	1.60%	717	1.77%
	Inferior temporal gyrus	26	2164	5.39%	1789	4.40%
	Middle temporal gyrus	27	1707	4.25%	1873	4.61%
	Superior temporal gyrus	28	833	2.08%	641	1.58%
	Heschl gyrus	29	453	1.13%	292	0.72%
	Planum parietale	30	288	0.72%	278	0.68%
	Planum temporale	31	223	0.56%	189	0.46%
	Planum polare	32	316	0.79%	270	0.66%
Occipital	**Lateral surface**					
	Inferior occipital gyrus	33	723	1.80%	1081	2.66%
	Middle occipital gyrus	34	1581	3.94%	2101	5.17%
	Superior occipital gyrus	35	471	1.17%	510	1.26%
	Medial surface					
	Fusiform gyrus	36	1143	2.85%	782	1.93%
	Lingual gyrus	37	494	1.23%	499	1.23%
	Calcarine sulcus	38	703	1.75%	788	1.94%
	Cuneus	39	863	2.15%	794	1.96%
Limbic	Parahippocampal gyrus	40	385	0.96%	339	0.84%
	Amygdala	41	84	0.21%	105	0.26%
	Posterior cingulate gyrus	42	353	0.88%	331	0.81%
	Middle cingulate gyrus	43	370	0.92%	338	0.83%
	Anterior cingulate gyrus	44	704	1.75%	700	1.72%
Nuclei	Putamen	45	60	0.15%	55	0.14%
	Caudate	46	179	0.45%	214	0.53%

Table 2.1: Labels and corresponding surfaces of the 47 anatomical regions of interest defined in each hemisphere

2.3.2 10-20 standard positions localisation and variability

The standard deviations of the spatial distributions of the 10-20 sensors over the scalp are presented in Table 2.2, based on the 16 infants normalized to the template, and expressed in the template subject coordinate system. Figure 2.5-A and 2.5-B depict probability maps for sensor locations over the template head surface and the template brain envelope surface. Fp1 and Fp2 were the most stable sensors, with the smallest standard deviation. Temporal and parietal sensors showed a large scattering. Paired t-test on the variability in each direction on the left versus right channel locations showed a larger dorso-ventral variation on the left relative to the right hemisphere (p=.027; proportion of left hemisphere sensors distributions larger than their right counterparts = 75 %) . Thus, the scattering of channel locations tended to be larger on the left than on the right hemisphere.

2.3.3 Brain structures localization and variability

We also examined the cerebral structure variability: Figure 2.7-B displays the sulcal models for six main cortical sulci over the infant population, providing estimates of the presence probability for each sulcus. Note that in line with the literature (Glasel et al., 2011), the sylvian fissure and the superior temporal sulcus showed a posterior ascending branch on the right side, while they remained rather horizontal on the left side. Figure 2.6 displays a measurement for dispersion over SPAM models: it indicates the degree of overlap between subjects' sulci. The intraparietal sulcus and frontal sulci exhibit strong variability, while the sylvian fissure and the superior temporal sulcus are rather stable sulci. Inter-subject variability is especially visible at the ends of the sulci, due to the differences in size.

2.3.4 Relations between scalp sensors and cerebral structures

The projection of the electrodes over cerebral structures is convergent across both approaches (electrode distribution projected on the template and template electrodes projected on sulcal probability maps) suggesting a stable relation between external landmarks and cerebral structures (Table 2.3). Figure 2.7-A and Table 2.3 summarize the anatomical relationships between the international 10-20 standard positions and the underlying cortical structures. Sensors Fp1 and Fp2 projected over the orbitary part of the frontal lobe, F3 and F4 over the middle frontal gyrus, C3 and C4 over the postcentral gyrus, and T5 and T6 projected on the posterior part of the inferior temporal gyrus. F7 and F8 overlaid the inferior frontal gyrus, pars triangularis. Interestingly, anterior temporal sensors did not

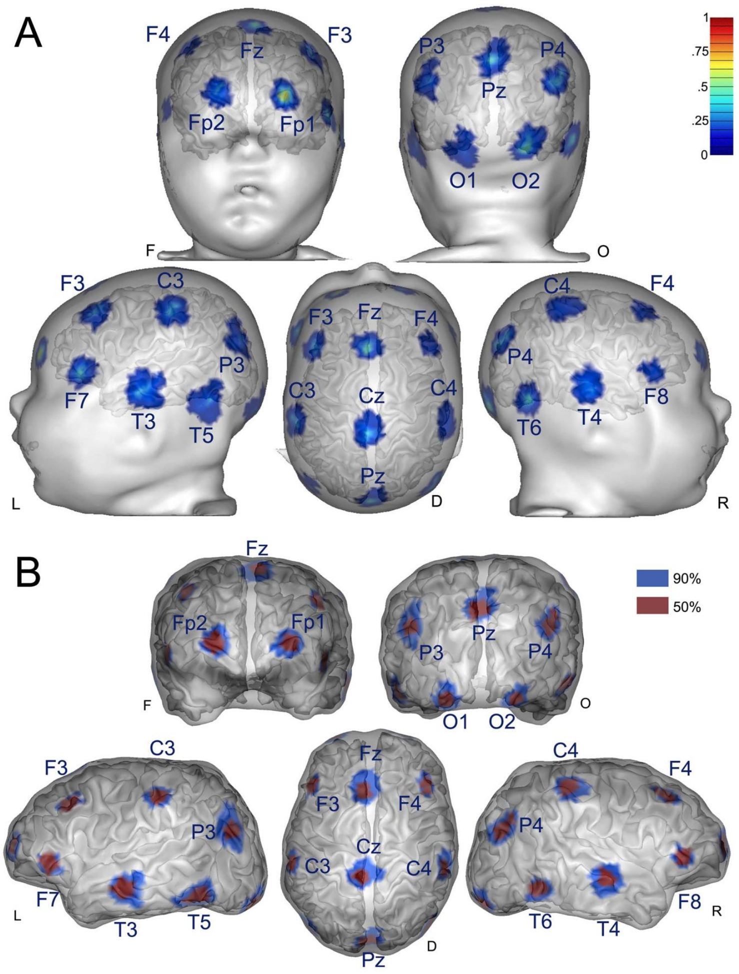


Figure 2.5: **10-20 standard positions and their cortical projections. A**: International 10-20 sensors' location distributions computed across the 16 infants, and showed over the skull of the template infant, with the underlying grey-white matter interface (from left to right and top to down: Frontal – Occipital; Left – Dorsal – Right). **B**: International 10-20 standard positions distributions, projected on the brain envelope, with the underlying grey-white matter interface. Red surfaces encompass 50% of the subjects and blue surfaces 90% (from left to right and top to down: Frontal – Occipital ; Left – Dorsal – Right).

Electrode	Left-Right standard-deviation (mm)	Antero-Posterior standard-deviation (mm)	Dorso-Ventral standard-deviation (mm)	Standard-deviation (mm)
Frontal				**4.02**
Fp1	2.09	1.36	2.45	3.49
Fp2	2.17	1.70	2.55	3.76
Fz	2.47	2.25	1.90	3.84
F3	2.28	2.74	2.65	4.45
F4	2.25	2.77	2.22	4.20
F7	1.98	2.57	2.79	4.28
F8	1.81	2.32	2.91	4.14
Central				**4.92**
Cz	2.45	3.36	1.71	4.49
C3	2.28	3.21	3.23	5.09
C4	2.11	4.13	2.29	5.17
Temporal				**5.26**
T3	1.51	4.03	3.94	5.83
T4	1.32	2.83	3.78	4.90
T5	2.12	3.53	4.25	5.92
T6	1.54	2.64	3.11	4.37
Parietal				**5.31**
Pz	2.74	2.31	3.64	5.10
P3	2.44	2.47	4.40	5.60
P4	2.68	3.36	2.96	5.21
Occipital				**5.13**
O1	2.87	2.60	3.76	5.40
O2	3.16	2.03	3.08	4.86
L vs R pvalue t-test(7)	0.52	0.75	0.03*	0.08

Table 2.2: Variability of the 10-20 international standard positions

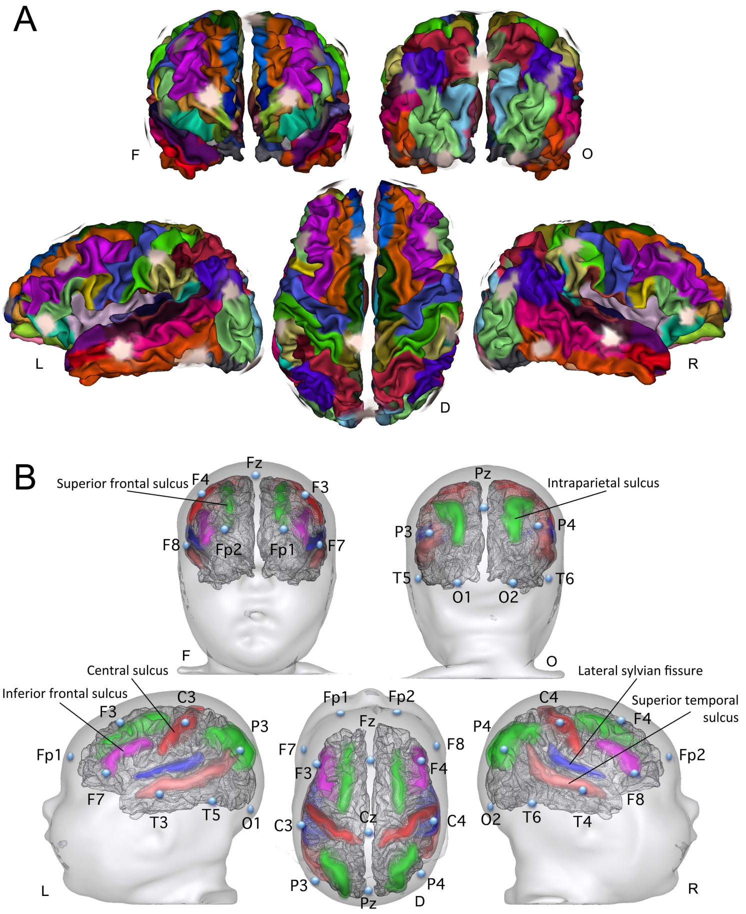


Figure 2.6: **Cranio-cerebral correlations.** **A**: Projected distributions of the 10-20 standard positions over the segmented anatomical brain regions (from left to right and top to down: Frontal – Occipital ; Left – Dorsal – Right). **B**: Internal structures distributions: sulcal presence probability modeled by SPAM, based on 15 infants, and the overlying international 10-20 electrode positions (from left to right and top to down: Frontal – Occipital ; Left – Dorsal – Right). 3D convenient representation is modeled by an isosurface corresponding to 60% of the whole probability mass, for each sulcus.

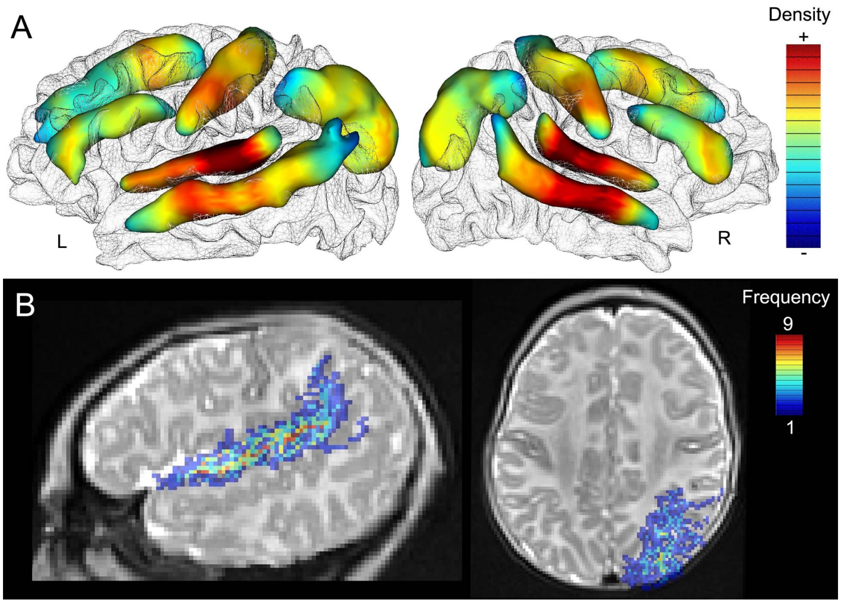


Figure 2.7: **Density values along sulci.** **A**: Inter-subject structural variability modeled by a density measurement along the 60% isosurfaces of SPAM models (Left – Right), arbitrary units. **B**: Two examples of inter-subject variability over the template MRI: the frequency maps of the 15 normalized right sylvian fissures and intraparietal sulci. For the sylvian fissure, on the left (sagittal view), the 15 normalized sulci show a higher degree of overlap as compared to the intraparietal sulci, on the right (axial view).

Electrode	Anatomical region	Electrode	Anatomical region
Fp1	Orbitary superior frontal gyrus	Fp2	Orbitary superior frontal gyrus
	Orbitary inferior frontal gyrus		Orbitary inferior frontal gyrus
	Middle frontal gyrus		Middle frontal gyrus
F3	Middle frontal gyrus	F4	Middle frontal gyrus
F7	Inferior frontal gyrus, triangular part	F8	Inferior frontal gyrus, triangular part
	Orbitary middle frontal gyrus		Orbitary middle frontal gyrus
	Anterior to IFS		*Anterior to IFS*
C3	Postcentral gyrus	C4	Postcentral gyrus
	Supramarginal gyrus		Supramarginal gyrus
	Posterior to CS		*Posterior to CS*
T3	Middle temporal gyrus	T4	uperior temporal sulcus
	Inferior temporal gyrus		Middle temporal gyrus
	Anterior portion of STS		*Anterior portion of STS*
T5	Inferior temporal gyrus, posterior part	T6	Inferior temporal gyrus, posterior part
P3	Angular gyrus	P4	Angular gyrus
	Middle occipital gyrus		Middle occipital gyrus
O1	Middle occipital gyrus	O2	Middle occipital gyrus
	Inferior occipital gyrus		Inferior occipital gyrus
Fz	Supplementary motor area	Cz	Paracentral lobule
	Medial superior frontal gyrus		Supplementary motor area
Pz	Superior parietal gyrus		
	Superior occipital gyrus		

Table 2.3: Projections of the 10-20 standard positions onto the template's cerebral structures (listed in a rank order), and in italic, onto the probabilistic sulci maps (SPAM).

project exactly over the same structures on the left and right hemispheres: T3 being slightly more ventral (left middle/inferior temporal gyri) than T4 (right superior/middle temporal gyri). This asymmetry is less obvious when SPAMS are considered, both electrodes projecting on the ventral part of the sulcus. Occipital sensors O1 and O2 were located above the inferior and middle occipital gyri, while parietal sensors P3 and P4 overlapped the angular gyrus and the middle occipital gyrus. The midline standard positions Fz, Cz, and Pz, were located on, or near, the inter-hemispheric fissure.

2.3.5 Depth measurements

Considering NIRS recordings, the distance between the captors and the brain is of crucial importance. For each infant, we thus measured the distance between the skull and the brain envelope from all points on the head surface and created

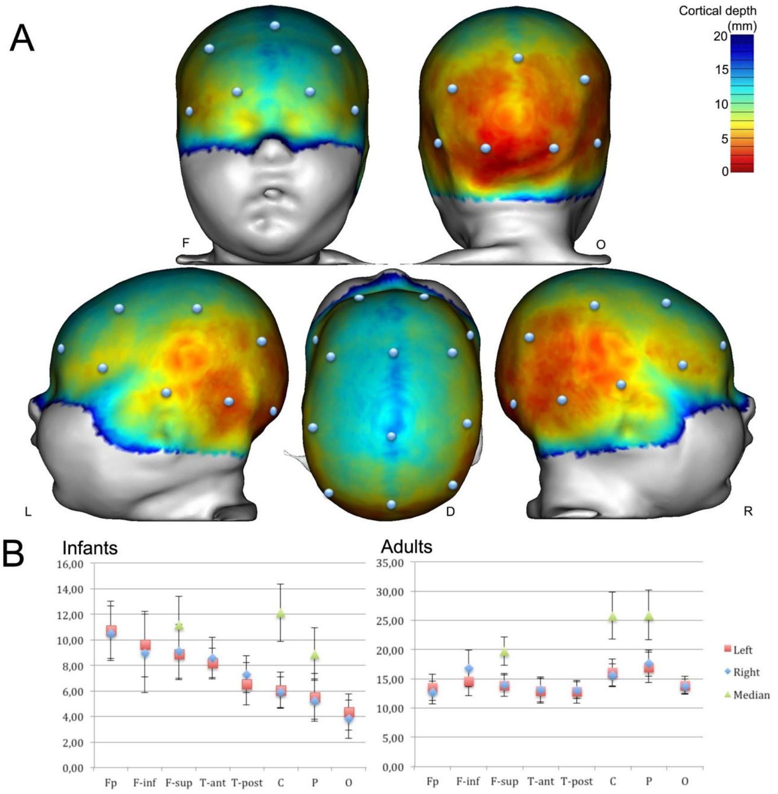


Figure 2.8: **Cortical depth measurements.** **A**: Cortical surface depth distribution after inter-subject integration, exhibited over the head surface (in millimeters). **B**: Cortical depth (in millimeters) as a function of electrode location on the skull, for the 16 infant subjects on the left, and as reported by Okamoto et al. (2004) for 17 adult subjects on the right.

	Mean	Standard Deviation
Infants population		
Left to Right tragi (mm)	102.90	5.56
Nasion to Inion (mm)	122.60	5.49
Tragi line to Vertex (mm)	97.97	7.02
Mean Cerebral Envelope (cm^3)	481.61	68.46
Radius of the corresponding sphere (mm)	48.54	0.22
Template		
Left to Right tragi (mm)	93.52	
Nasion to Inion (mm)	117.77	
Tragi line to Vertex (mm)	92.89	
Mean Cerebral Envelope (cm^3)	412.79	
Radius of the corresponding sphere (mm)	46.20	

Table 2.4: Head measures

an averaged cortical surface depth map over the infant group. We observed an anterior-posterior gradient with occipital cortex being closer to the surface than frontal cortex (Figure 2.8-A). The skull was also more distant from the cortical envelope along the midline due to the superior sagittal sinus running in the longitudinal fissure.

To quantify this description, we first measured the distance between the 10-20 standard positions on the head surface and their cortical projection over the brain envelope for each infant. We confirmed the anterior-posterior gradient for brain depth with smaller electrode-envelope distances for posterior electrodes, and larger distances for the most anterior positions. This gradient is not present in adults (Figure 2.8-B), data extracted from Okamoto et al. (2004). However, just as adults, the most distant electrodes from the brain are the midline positions (Cz, Fz, Pz). Both in the infant and adult groups, there are no difference between the left and right sides. Depth and asymmetry measurements are summarized in Table 2.4.

We additionally measured the distances between four precise cortical landmarks and the overlying 10-20 positions. For the hand knob, the ventro-lateral end of the pars triangularis, and the lateral edge of Heschl's sulcus, distances were similar in the right and left hemispheres, but the posterior end of the sylvian fissure was significantly closer to P3 and more distant from C3 on the left and thus tended to be closer to C4 and more distant from P4 on the right because of its different shape and slope in the left and right hemisphere,. However, we found no asymmetry in distances between tragus and the posterior end of sylvian fissure, a distance reported in preterms by Mahmoudzadeh et al. (2013).

2.4 Discussion

Our goal was to provide the community of developmental cognitive researchers with an accurate description of the international 10-20 standard electrodes locations relative to the underlying cerebral structures in infants, and an anatomical parcellation of a 2-month-old infant brain to localize the brain functional activity once data obtained by EEG or NIRS are registered to this template. As the brain is growing very fast after birth, we limited our age-range to the first four postnatal months (3 to 16 weeks). We investigated whether the relationships between external landmarks and brain structures were reliable during this developmental period when the head and the brain undergo quick growth (0.5 cm per week). We provided two complementary measures for cranio-cerebral relationships. With the first measure, we quantified the variability of the standard positions between infants, with respect to the underlying brain structures of the template infant. To complement this first approach, the second measure focused on the localisation and variability of the internal cerebral structures over the group, with respect to the 10-20 theoretical standard positions of the template infant. We also provided additional information about brain-skull distances, and studied potential asymmetries of the distance between external sensors and cerebral landmarks.

2.4.1 The infant template

As template, we chose an infant with a symmetrical head (i.e. not distorted by birth or sleeping habit) and high quality T1w and T2w images, in the median of our age range. For the anatomical parcellation, we had no particular difficulties to follow the principles used by Tzourio-Mazoyer et al. (2002) in adults, based on sulcal delineation confirming that the main macroscopical landmarks are well developed from birth on. We added the planum temporale, planum parietale and planum polare as they are important structures that were not reported in the Tzourio-Mazoyer et al.'s atlas (2002), because of a specificity of their template subject who had not the classical larger left planum. We did not parcel the pallidum and hippocampus, which are small and internal structures, difficult to segment in the infant brain.

As it is the case in the Tzourio-Mazoyer et al's (2002) and Talairach and Tournoux's (1988) atlases, our labelling relies on a single subject. As opposed to multi-subject-based probabilistic atlases (Altaye et al., 2008; Kuklisova-Murgasova et al., 2011; Fonov et al., 2011; Akiyama et al., 2013), single-subject atlases have a main drawback: the brain shape and anatomical structures might be biased by the subject's particularities. However, it has recently been reported (Tsuzuki & Dan, 2013) that, when it comes to functional group analyses, anatomic structural differences between single-subject and multi-subject based atlases are not substantial. Besides, this approach allows obtaining a precise description of the cerebral structures, which is necessary to report data, and also to compute sources of EEG/MEG recordings, as it is often not possible to obtain the MRI of each subject in an infant population. This atlas should increase the precision of the observed effects for researchers as far as they respect the 10-20 system to locate their recording channels (or know the location of their channels relative to this system).

2.4.2 Cranio-cerebral correlations

In our group of infants, the channels locations of the 10-20 system were rather stable, and consistent with adult studies. Among the standard positions, frontal Fp1 and Fp2 were the most stable ones, with the weakest variability, whereas more posterior positions, in particular parietal and occipital points, exhibited larger variability. This may be explained by infant head shape variability due to skull remodeling by sleeping habits or following vaginal birth (e.g. the inion can be particularly difficult to identify on MRI in infants with a flat occiput). This feature is most likely related to the posterior variability in the human head shape, as Okamoto et al. (2004) and Jurcak et al. (2007) reported the same observation in adults. The location of the channels relative to the template's brain structures can be examined in Figure 2.7-A, which reports the channels distribution and variability across infants relative to the template brain. Since, the standard locations are examined with respect to the template brain, the description provided here is not purely probabilistic. Still, it provides a first description of the cerebral structures underlying the 10-20 standard locations in a typically developing population of infants with variable head shapes. Figure 2.7-B, on the other hand, reports the sulcal distribution and variability across infants relative to the template's 10-20 standard positions. It provides a probabilistic model of the inner cerebral structures in an infant population (9.0 ± 3.6 week-old infants). Both of these approaches are complementary.

Our results were roughly similar to the locations described by (Blume et al., 1974) in post-mortem brains (6 infants younger than 4 months of age and a two-year-old). The differences are likely due to differences in methodology. Infants are lying down in an MRI scanner and were sat to mark the channels location in (Blume et al., 1974). Due to its weight, the brain might move backward relative to the skull in the scanner but the tight connections at its basis (white matter tracts, sylvian arteries, cerebral falx) might limit its backward movement. By contrast, post-mortem brains are no more inflated by the blood pump and might be packed down by gravity. They also might suffer from oedema due to death-anoxia creating inaccuracies. In any case, the standard locations along the anterior-posterior axis were similar in both studies whereas the supra-sylvian positions projected somewhat higher along the vertical axis in our study. For instance, we observed that F7-F8 laid at the anterior tip of the inferior frontal sulcus, while this pair of locations projected below the frontal lobe in the post-mortem brains. This difference might be due to the sitting position in (Blume et al., 1974) or to the inclusion of preterm infants in their dataset (several infants of this study are described as not having clear frontal sulci which appear at the end of gestation). Indeed, the operculation of the inferior frontal region, which pushes down the inferior frontal region toward the temporal lobe, only occurs during the last weeks of gestation and first months of post-term life. But the infants' ages are underspecified in this study, so we cannot assess the validity of this hypothesis.

The main difference between our report in infants and Okamoto et al. (2004) in 17 adults (22-51 years) using MRI, is the projection of the lower electrodes (01-02, T5-T6). These pairs of electrodes laid above the inferior limit of the brain (consistent with Blume et al.'s study in infants), whereas Okamoto et al. reported projections over the middle occipital and middle temporal gyri in the adult population. This difference might be related to the globularization of the modern human brain driven by parietal cortex expansion during development (Gunz et al., 2012) or/and by the delayed development of the face (in particularly its lower part), relative to the brain during childhood. By contrast, the anterior temporal location (T3-T4) and the frontal locations projected on similar structures at both ages. For example, C3-C4 were reported over the post-central region in most of the adults, just as in our infant study. This was also the case in (Steinmetz et al., 1989; Towle et al., 1993), but note that two other studies (Homan et al., 1987; Jasper, 1958) located these electrodes over the pre-central region in adults.

Beyond the variability of electrode placement, the location, shape, size of the cortical structures might themselves differ from one subject to the other. Using SPAMs (Perrot, et al., 2011), we provided a probabilistic model for the localisation of 6 main sulci (Figure 2.7-B). We complemented this localisation information with a measure of dispersion along the sulci (Figure 2.6). The intraparietal sulcus was particularly variable. Indeed, it is often segmented into different subunits by several crossing-folds. The inferior frontal sulcus also exhibited a high dispersion, segmented into subunits by crossing-folds as well. The sylvian fissure and the superior temporal sulcus, on the contrary, appeared to be the most stable sulci, notably the central region of the sylvian fissure, at the level of Heschl's gyrus, which exhib-

Sensor	Mean Distance (mm)	Standard Deviation (mm)	Asymmetry (R vs L)	
			Coefficient	p-value t-test(15)
Fp1	10.75	2.23	-2%	0.63
Fp2	10.48	2.12		
Fz	11.15	2.23		
F3	8.88	2.03	2%	0.50
F4	9.07	2.10		
F7	9.63	2.56	-10%	0.21
F8	8.92	3.05		
Cz	12.12	2.23		
C3	6.03	1.41	-2%	0.72
C4	5.87	1.20		
T3	8.19	1.14	3%	0.43
T4	8.54	1.60		
T5	6.53	1.65	12%	0.09
T6	7.28	1.44		
Pz	8.89	1.99		
P3	5.53	1.77	-6%	0.39
P4	5.17	1.60		
O1	4.35	1.39	-16%	0.16
O2	3.77	1.47		

Table 2.5: Sensors-brain envelope distances and left-right asymmetry measures

ited a strong consistency between subjects. The enhanced inter-subject variability, at the ends of sulci, is mostly due to differences in the sulci lengths.

2.4.3 Electrode-brain distance information

The distance from the scalp to the cortical surface is of crucial importance in NIRS and EEG measurements. We therefore examined inter-subject and inter-channel variations of electrode-brain distances. There was an increase of scalp-brain distances from the occipital to the frontal locations and from the inferior to the superior locations (Figure 2.8). Because infants are lying down in the MRI bore, we cannot rule out that the brain weight pushed the CSF forward in lying position but the position of the electrodes along the anterior-posterior axis was similar to Blume et al.'s (1974) description obtained in sitting infants, and the brain is tightly connected, limiting its movement within the skull. This gradient was not observed in reclining adults (Okamoto et al., 2004) and the scalp-brain distance was rather stable for the different channels at this age (Figure 2.8-B). At both ages, the most distant channels were the median channels, because of the superior sagittal sinus running in the inter-hemispheric region.

Researchers must thus be aware that their sensors might be more or less distant from the brain. Sensor-brain distances, which affect signal amplitude, can vary by a factor of 3 between locations (4 to 12 mm). Furthermore, distances between important landmark and their closer scalp electrodes were often superior to 20 mm, and for some landmarks close to 30 mm (e.g. Cz to the hand knob). Odabaee et al. (2013) showed that the EEG amplitude in neonates decays rapidly and reached the noise floor at around 5cm from the amplitude peak for the 12-20Hz frequency band. Although for slower latencies, the decline was less steep, this study confirms models demonstrating that due to the medium higher conductivity in infants (especially of the skull) scalp EEG is more focal than in adults (Grieve et al., 2003, 2004). It is noteworthy to signal that Odabaee et al's (2013) observation of a steeper slope over parietal areas than over inter-hemispheric fissure is probably due to the increased electrode-brain distance over the midline. The inter-subject variability may seem small (3 to 8 mm, 5 mm in average, for the distances reported in table 2.5) but given the amplitude decreasing slope described in (Odabaee et al., 2013), it represents a 10% variability in the amplitude between subjects. A high-density coverage of the head is thus crucial for a correct spatial sampling of the EEG activity in infants as already advocated in (Grieve et al., 2003, 2004; Odabaee et al., 2013).

Distance to the internal structures is also at stake for NIRS measurements as the measure is done in the crescent of light between the light emitter and the detector. The diodes spacing commonly used in infants, is sufficient to reach the brain. For example, a spacing of 20 mm between source and detector allowed reaching a depth of 10 mm with the same iso-sensitivity (Patil et al., 2011). However, NIRS experimenters should keep their sensors on the skull convexity and move them above the region of interest. It might be useful to increase the source-detector distances above 20 mm if the goal is to record from frontal areas and anterior temporal areas (Figure 2.8) at the expense of a decrease in signal intensity.

2.4.4 Left-right asymmetries

Anatomical asymmetries are observed from the preterm period on (Chi et al., 1977; Dubois et al., 2008; Glasel et al., 2011; Habas et al., 2012; Sowell et al., 2002; Toga & Thompson, 2003; Dubois et al., 2010) and several maturational markers indicate a different developmental time-course of the left and right human hemispheres (Chiron et al., 1997; Dubois et al., 2009; Leroy, Glasel, et al., 2011; Lin et al., 2012). To study functional asymmetries, it is thus necessary to control for the known structural asymmetries which can affect channel/brain structures relations.

Beauchamp et al. (2011) recently reported significant left-right asymmetries in brain-scalp distances measured from T1w images in 0-18 month-old infants, at different cortical landmarks. In line with this idea, we performed brain-scalp measurements for each of the 10-20 standard positions, and found no asymmetry at these points. Note however, that we actually computed distances between the brain envelope and the scalp; we might therefore have occulted some anatomical asymmetries in the depth of the sulci but this is also the case in (Beauchamp et al., 2011) where the authors chose external brain landmarks (i.e. not in the depth of sulci). We also computed distances between some remarkable cortical structures and the closest 10-20 standard positions. In particular, for the lateral edge of the Heschl's sulcus, we found neither asymmetry in distances to standard positions T3-T4 nor T5-T6, while Beauchamp et al. (2011) reported a strong asymmetry for distances to the skull. Similarly, we did not find asymmetry when measuring the distance from the tragus to the posterior end of sylvian fissure, replicating Mahmoudzadeh et al.'s reports (2013) in preterm infants. J.D. and G.D.L. were present during data acquisition and were attentive to a symmetrical placement of the infant head in the coil to reduce CSF accumulation on one side, as subsequent left-right differences in measures approximations which inevitably accumulate at each reconstruction step along the MRI data processing line. A correct placement in the coil is more difficult to ensure when data are coming from data-base and may partially explain the difference between these two sets of data (see for example Figure 2.2 in (Beauchamp et al., 2011), in which the presented axial image is tilted to the left of the image).

We observed only one significant distance asymmetry for the posterior end of the sylvian fissure ($p=.016$, table 2.5), due to the known difference in the slopes of the left and right sylvian fissure (the right fissure is shorter, steeper and more dorsal than the left from the preterm period on). P3 and P4 (and also C3-C4) were not at the same distance from the posterior end of the sylvian fissure. Thus, sensors were at the same distance from the brain envelope on the left and right sides, but the best location to record from homologous posterior peri-sylvian structures might be slightly different on the right and left sides. This should be taken into account when functional asymmetries are discussed.

2.5 Conclusion

In this study, we presented an infant brain atlas based on a single subject (7.1 week-old girl), complemented with a probabilistic description of the 10-20 standard positions with respect to the underlying brain structures, and a prob-

Structure	Electrodes	Mean distance (mm)	Standard deviation (mm)	Asymmetry (R vs L)	
				Coefficient	p-value t-test(15)
Hand Knob	Left-Cz	34.77	2.97	1%	0.76
	Right-Cz	35.36	4.54		
	Left-C3	23.80	3.87	-6%	0.31
	Right-C4	22.60	4.23		
Ventro-lateral end of the pars triangularis	Left-F3	30.10	6.87	4%	0.49
	Right-F4	30.86	3.59		
	Left-F7	20.71	4.53	-3%	0.67
	Right-F8	19.90	3.39		
Lateral edge of Heschl's sulcus	Left-T3	20.13	4.62	-8%	0.34
	Right-T4	18.60	4.59		
	Left-T5	35.87	5.25	0%	0.99
	Right-T6	35.79	4.85		
Posterior end of the sylvian fissure	Left-C3	22.32	6.20	-27%	0.07
	Right-C4	17.01	5.73		
	Left-P3	25.95	5.19	21%	0.016*
	Right-P4	31.98	5.94		
	Left-Tragus	72.01	6.40	-3%	0.34
	Right-Tragus	71.89	8.30		

Table 2.6: Distance from four cortical landmarks to closest international 10-20 standard positions (and tragus), and left-right asymmetry measures

abilistic model of sulcal patterns. The template and all maps provided in this article can be publicly downloaded (http://www.unicog.org/pm/pmwiki.php/Site/InfantTemplate). We hope that this work helps the community converge to a unified analysis framework for infant brain recordings. Still, further studies are needed in order to cover the various stages of brain development, and to grow the body of knowledge in the field of neonatal/paediatric neuroimaging.

Acknowledgments

This research was supported by La Fondation Motrice, la Fondation de France, la Fondation Bettancourt, la Fondation Fyssen, l'Ecole des Neurosciences de Paris and the McDonnell foundation. We are particularly grateful to Neurospin staff for their daily help and at last but not least we thank infants and parents for their participation.

CHAPTER 3

ELECTROPHYSIOLOGICAL EVIDENCE OF STATISTICAL LEARNING OF LONG-DISTANCE DEPENDENCIES IN 8-MONTH-OLD PRETERM AND FULL-TERM INFANTS

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.

3.1 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 et al., 2008, 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 et al., 2010; Shi, Werker, & Cutler, 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 et al., 1993; E. K. Johnson & Tyler, 2010; Saffran et al., 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 et al., 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 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 15month-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, Gomez 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 et al., 2010; Brusini, Dehaene-Lambertz, Dutat, et al., 2016). At a much younger age, 4-months, Friederici et al. (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 et al., 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 et al., 2010, 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 et al., 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 3.1). 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 et al., 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 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 et al., 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:

- 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 et al., 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 4Hz) and word (around 1Hz) frequencies being rather low, the normalization term captures the spontaneous background 1/f noise (steeper in infants than in adults).

- A growing body of research shows that phase-locking to speech rhythm (rather than power spectrum variations) reliably tracks speech comprehension (Ahissar et al., 2001b; Kerlin et al., 2010; Luo & Poeppel, 2007; Peelle et al., 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 et al., 2010) and attentional processes (Ward, 2003) or to more specific speech processes, such as top-down processes in the beta-band (Fontolan et al., 2014) and long-distance relation computation when adults listened to sentences, in the alpha-band (Meyer et al., 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.

3.2 Methods

3.2.1 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. From 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 section 3.2.4). 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 ± 1.8 wGA) and full-terms between 38 and 41 wGA (mean = 39.5 ± 0.8 wGA). At birth, all infants had:

- APGAR scores higher than 6 and 8 respectively at 1 and 5 min
- normal weight, size, and cranial perimeter for their gestational age
- normal otoacoustic emissions
- 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 Sótero del Río 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 et al., 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 ± 0.7, 34.9 ± 0.6 and 46.5 ± 0.7 weeks, respectively.

3.2.2 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 3.1). 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 3.1). 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 to 0.42) within words and 0.5 (0.43 to 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 (Pena et al., 2002; Buiatti et al., 2009) and infants (Marchetto & Bonatti, 2014). Contrary to the 200 ms pause used by Marchetto and Bonatti, this pause was subtle and not consciously perceived by adults (Pena et al., 2002).

If like the adults in Pena et al. (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 3.1), 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 AxC 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),

Learning Speech stream	Test 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

Table 3.1: Material used to generate the artificial speech stream and the test items

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.

3.2.3 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 kΩ. No behavioral measurement was carried out.

3.2.4 Data processing and statistical analysis

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 µV, or when deviation between fast and slow running averages exceeded 150 µV. 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.

3.2.4.1 Learning part: Frequency tagging analysis

Signal processing: 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:

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

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 et al., 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.

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

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

Statistical analysis: 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.

3.2.4.2 Learning part: 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 3.2.4.1). 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).

3.2.4.3 Test part: ERP analysis

Signal processing: 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: 696ms). 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 3.2.1). 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.

Statistical analyses: To identify the latencies and topographies of the auditory ERPs in our subjects for this type of stimuli, we first averaged 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 (Figure 3.2) 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 (t1) encompassed the responses to the last two syllables of the word and a second (t2) encompassed the late slow wave (Figures 3.2 and 3.3). 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.

3.2.4.4 Test part: 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 3605ms-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 1605ms. 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 (pvalue<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.

3.2.4.5 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 ($\frac{RW-PW}{RW+PW}$) 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.

3.3 Results

3.3.1 Learning part: Neural correlates of online statistical computations

A large cluster of electrodes showed significant entrainment at the syllable frequency (f_3 = 4.17Hz; $p_{corr} < 0.001$). Crucially, entrainment was also significant at the word frequency (f_1 = 1.39Hz) in two clusters over temporal regions: a left cluster ($p_{corr} = 0.001$) spreading over a temporo-frontal region, in between C3, F7 and T7 and a right cluster ($p_{corr} = 0.005$) located on T8 and posterior, towards P8. No significant entrainment was observed at a two-syllable frequency (no significant cluster was found) (Figure 3.1).

The phase-locking values were averaged over these significant clusters separately at

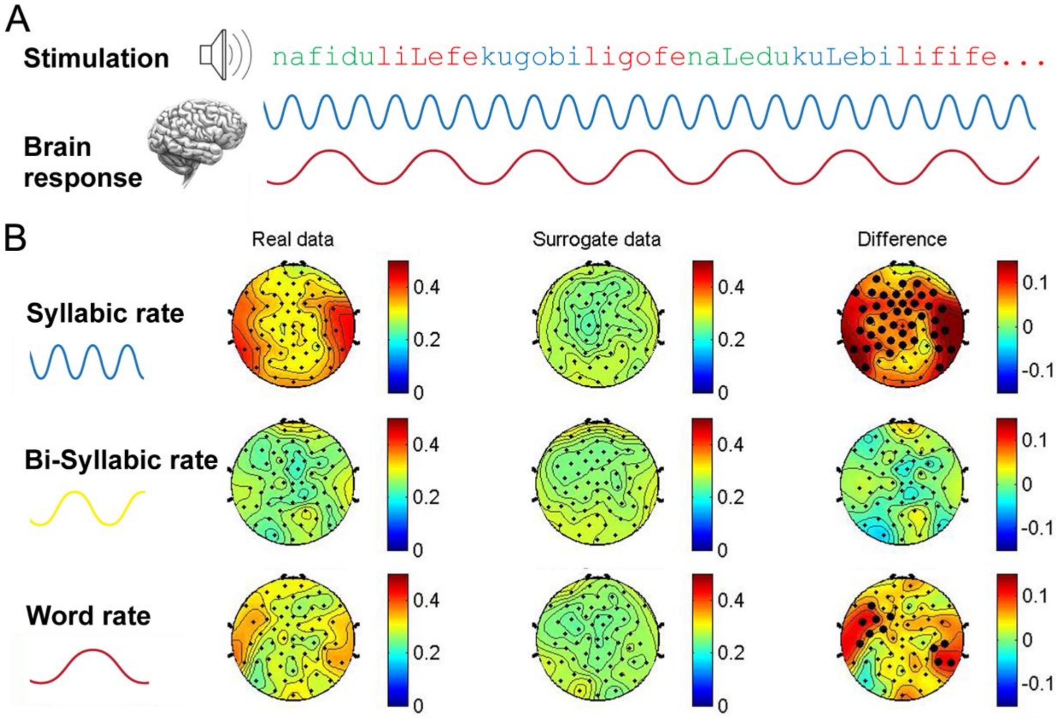


Figure 3.1: **Frequency tagging during the learning stream.** **A**: Schematic representation of the expected brain activity in response to the stimulation, at syllabic and word frequencies. The oscillatory activity is phase locked to the onset of syllables (blue) and words (red). **B**: Phase-locking 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 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.

3.3.2 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 to 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 syllables 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 × Polarity × Hemisphere × Group). *Rule-words* evoked a larger auditory response, notably over the left hemisphere, yielding a marginal interaction Condition × Polarity ($F(1, 42) = 3.03$, $p = 0.08$, Figure 3.2). Because of the significant interaction Hemisphere × Polarity ($F(1, 42) = 15.98$, $p < .001$), we computed post-hoc analyses on each hemisphere. A significant Condition × Polarity interaction was ob-

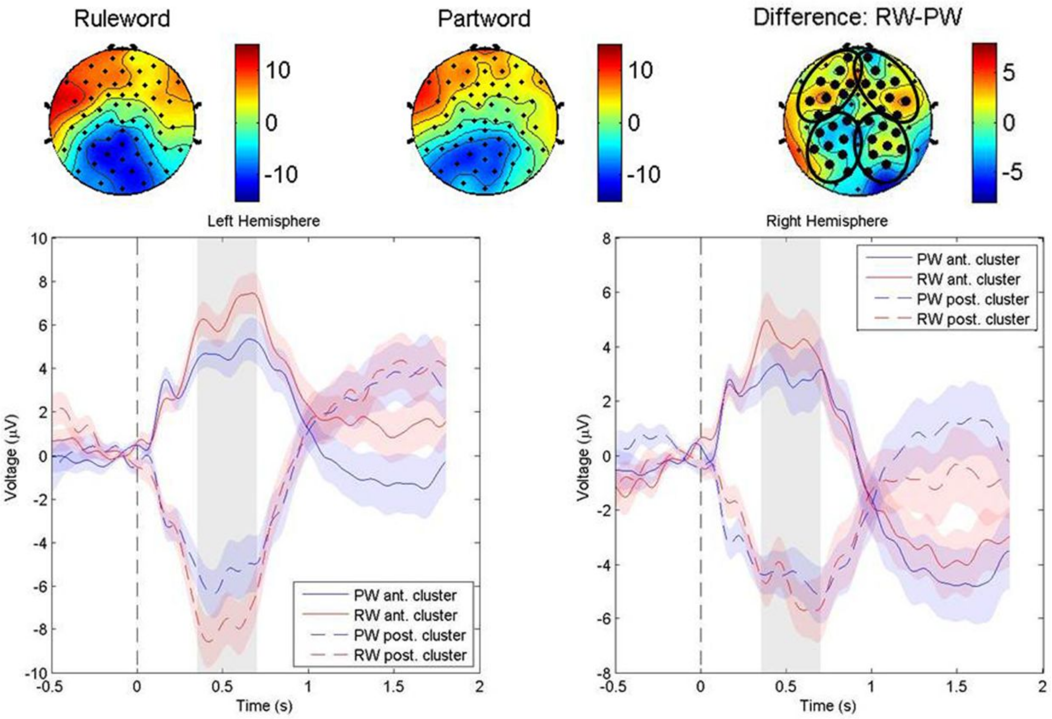


Figure 3.2: **Event-related potentials in response to *rule-words* (red) and *part-words* (blue). Early response:** Top panel shows the topography of the response to *rule-words* and *part-words*, and the difference (RW-PW). Lower graphs show the time course of the voltage averaged over the circled groups of electrodes, for the left and right hemispheres (dotted lines: posterior clusters, plain lines: anterior clusters).

served 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 t2 ([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 × Polarity × Group). The Condition × Polarity interaction was significant ($F(1,42) = 7.41$, $p = 0.009$,

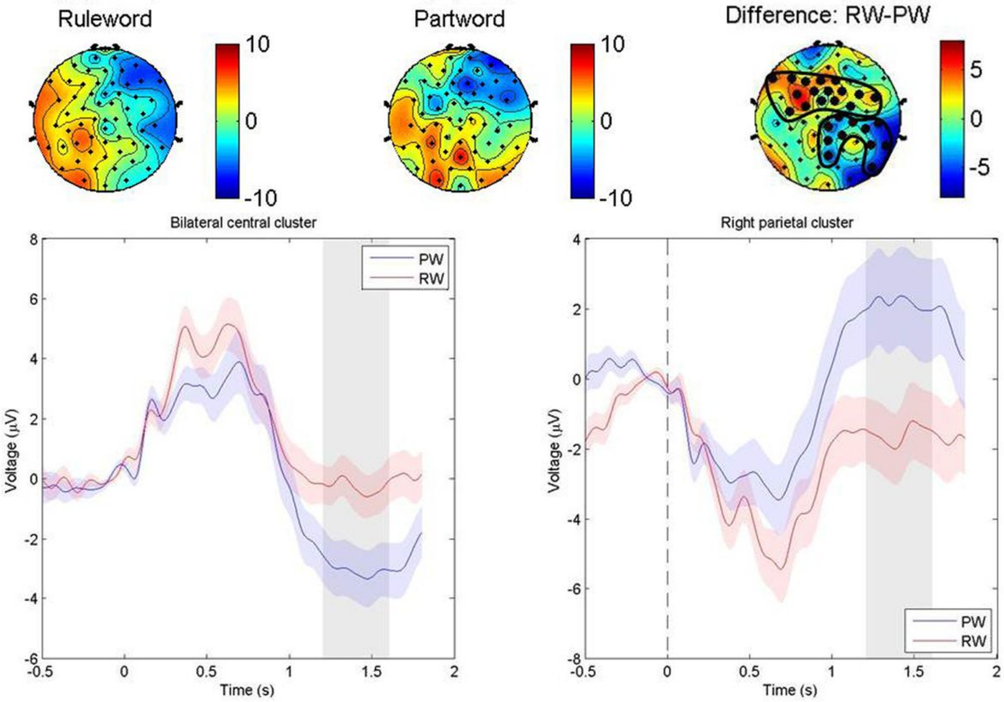


Figure 3.3: **Event-related potentials in response to *rule-words* (red) and *part-words* (blue). Late response:** Top panel shows the topography of the response to *rule-words* and *part-words*, and the difference (RW-PW). Lower graphs show the time course of the voltage averaged over the circled groups of electrodes from the bilateral central cluster, and the negative right parietal cluster.

Figure 3.3). The Condition effect was significant on each cluster (F(1,42)=8.34, p=0.006 for the positive cluster, and $((1, 42) = 4.87$, $p = 0.033$ for the negative cluster).

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

3.3.3 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 ($p_{corr} = 0.017$) at the syllable frequency ($f = 4.17$Hz, Figure 3.4A) 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 (Figure 3.4B). 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, $p_{corr} = 0.018$) 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, $p_{corr} = 0.004$) 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 towards electrode F4.

3.3.4 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*, Figure 3.4C).

3.4 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

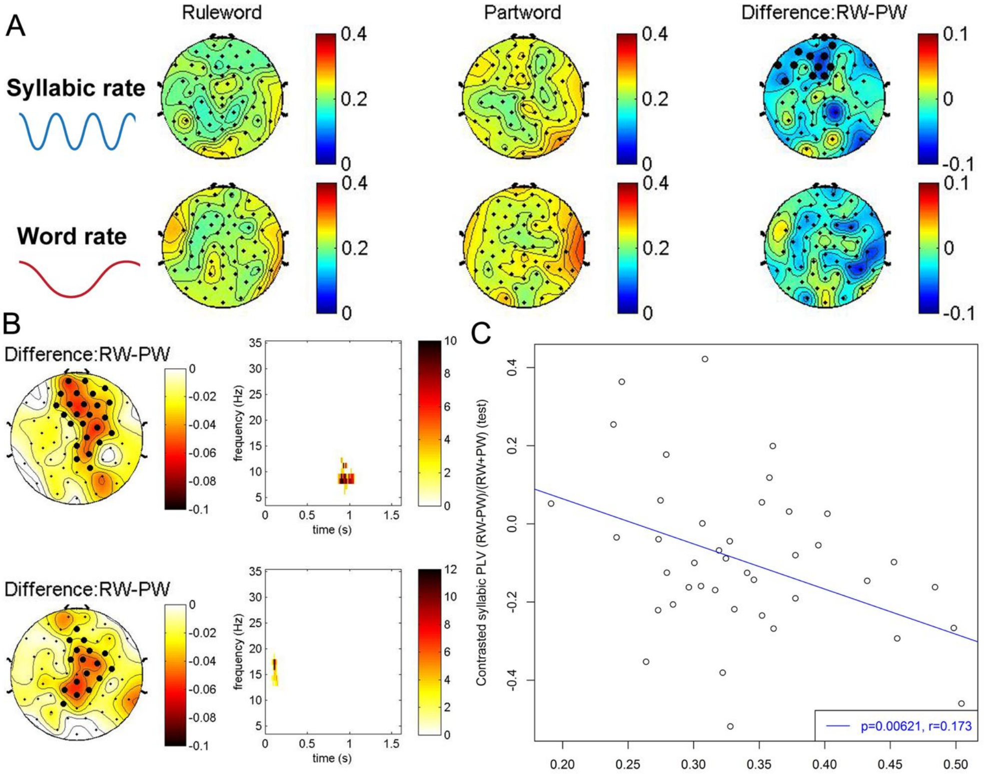


Figure 3.4: **Phase locking value (PLV) analysis in response to *rule-words* and *part-words*.** **A**: PLV at syllabic and word rates in response to *rule-words* (first column) and *part-words* (middle column), and the difference (RW-PW, third column). Electrodes showing a significant difference between 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.

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.

3.4.1 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., 2001a; Luo & Poeppel, 2007; Shannon et al., 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 et al., 2007) and infants (E. K. Johnson & Jusczyk, 2001; Shukla et al., 2011) have shown that prosodic cues reduce the influence of TP computations in speech segmentation. Morgan et al. (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 et al., 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

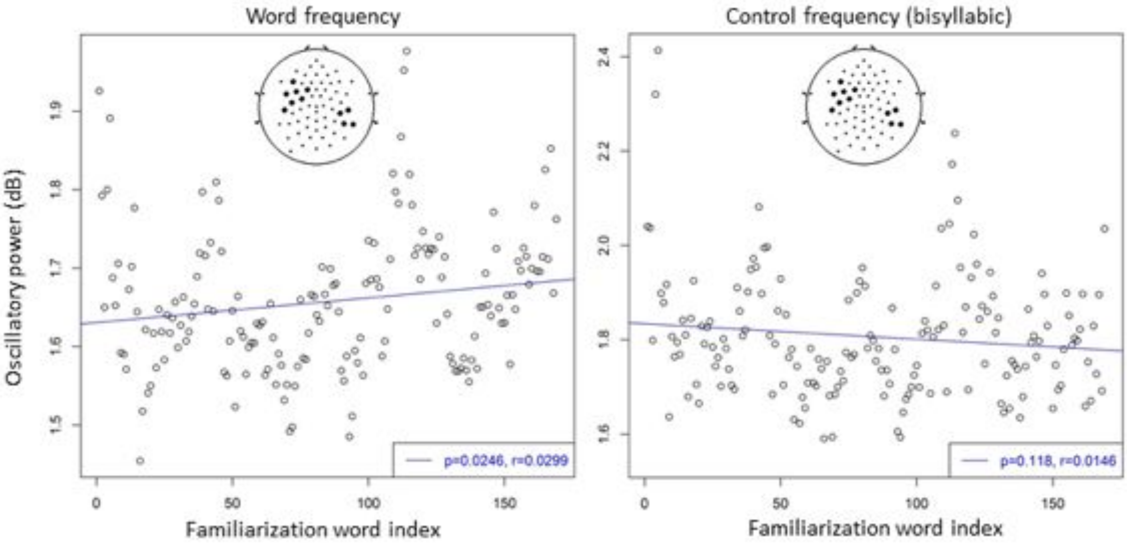


Figure 3.5: **Evolution of the oscillatory power** throughout the familiarization period, at word frequency and bisyllabic frequency. The power is averaged over the significant bilateral cluster for word entrainment, and over participants.

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. More interestingly, the progressive increase of oscillatory power at this precise frequency throughout the 180 words of familiarization (Figure 3.5) suggests that neural entrainment gets gradually stronger as infants accumulate more and more evidence for segmentation.

Further control experiments should be held in infants in order to compare their performances with the adult results reported in Buiatti et al. (2009), and test their segmentation abilities in the absence of subliminal pauses on the one hand and in the absence of long-distance dependencies on the other hand. One could also imagine an intermediate control experiment during which subliminal pauses would progressively be removed. The inspection of the oscillatory power throughout the familiarization period would therefore inform us whether entrainment can be maintained once the word-internal statistical dependencies have been detected.

3.4.2 Infants differently respond to *rule-words* and *part-words* during 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 100ms. 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 (Figures 3.3 and 3.4A). 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* (Figure 3.3). Late responses are often recorded in attentive infants when unexpected events occurred (Csibra et al., 2008), for example after a rare event in oddball designs (Basirat et al., 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.

3.4.3 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 et al., 2004; Mintz et al., 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 (Marchetto & Bonatti, 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.

3.4.4 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. (2001b), 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.

3.4.5 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, Gomez 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).

3.5 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 (E. K. 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 discrimination responses during test. Temporal imprecision has been proposed as one of the mechanism at the origin language impairment (Abrams et al., 2006).

Acknowledgments

CK, MB and GDL were supported by the Bettancourt Fundation, Robert Debré Foundation and MP by the grant Fondecyt 1141040.

CHAPTER 4

ELECTROPHYSIOLOGICAL CORRELATES OF CHAINED MENTAL OPERATIONS IN 5 MONTH-OLD INFANTS

4.1 Introduction

The specificity of human language unquestionably lies in its complex structure, unrivaled in any other animal communication system. Yet, human infants rapidly apprehend its underlying organization and master its limitless expressive power within a few months. Neither intensively trained animals, nor the mature adult brain, nor sophisticated computational models reach these performances and speed for learning a new language. The remarkable learning skills of human infants primarily rely on their ability to rapidly represent the input at various levels of description, ranging from the simple physical features, and the statistical distribution of its constitutive units, up to more abstract relations between the different units. This abstract representation level is a powerful tool, supporting the discovery of grammatical categories, or the hierarchical relations that govern their usage, as well as the acquisition of other complex cognitive skills like mathematics or music.

In order to progress through the hierarchy until the more abstract representations postulated in language, one needs to integrate various features – abstract or not – of the input, and recode it into a single mental variable that is then available for further computations. We can then chain multiple complex operations. This mental ability allowed the development of the rich symbolic systems we constantly use to represent the various aspects of the external world, but little is known about the ontogeny of this mechanism. It typically operates on physical objects when several elements are identified as belonging to the same semantic group: this abstract category is then recoded into a single mental variable which can be associated to a verbal label during word learning. Recent studies evidenced that infants can associate some familiar auditory word forms to the object categories they refer to, from 6 months of age (Bergelson & Swingley, 2012, 2013; Shukla et al., 2011). Verbal labels can even trigger the recoding process (Waxman & Markow, 1995; Ferry et al., 2010). It also operates over more abstract entities, supporting for example linguistic computations, in the case of language. We can recode the various lexical units into subject, object, verb or determiner, based on their morphological or phonological attributes and manipulate them according to the syntactic rules that apply to each of these categories. This recoding process applies at any level of the hierarchy, allowing the manipulation of increasingly complex objects, from lexical units to entire clauses. This mechanism also underlies mathematical as well as musical formalism, some typical human-specific competences. In the present study, we investigate whether preverbal infants can represent abstract linguistic sequences as a unified mental representation.

In a seminal study, Marcus et al. (1999) substantiated abstract computations in preverbal infants. They habituated 7 month-old infants to triplet sequences conforming to one of two abstract patterns based on adjacent (ABB, e.g. ba-po-po) or non-adjacent repetitions (ABA, e.g. ba-po-ba). Crucially, triplets are built out of distinct syllables, and can only be discriminated on the basis of their pattern of repetitions. After experiencing only 16 familiarization items, 7 month-old infants generalized their habituation to novel instantiations of this pattern, and showed enhanced interest when presented with a different pattern. Based on these generalization performances, the authors argued that preverbal infants had encoded the abstract "algebra-like" rule representing the structure. Their claim triggered an intense debate (Seidenberg & Elman, 1999; McClelland & Plaut, 1999; Marcus, 1999b; Christiansen & Curtin, 1999; Altmann & Dienes, 1999; Marcus, 1999a), questioning the need for pattern encoding in the generalization process they had evidenced.

Crucially, in this experiment, as well as in most of the following ones on abstract rule learning, the authors use a habituation/dishabituation paradigm, implementing only a single consistent pattern. Therefore, the exact nature of the discrimination mechanism is not entirely addressed. Do infants' discrimination abilities only reflect basic sequence processing mechanisms such as repetition detection, ordinal information, or edge biases? Alternatively, do these performances involve the encoding of the sequence information into a single unified representation that can guide actions, or play a role in inference? Can infants manipulate several abstract representations simultaneously? Most of the literature on abstract rule learning has questioned the role of speech, demonstrating that although not specific to speech-like stimuli (C. Dawson & Gerken, 2009; Saffran et al., 2007; S. P. Johnson et al., 2009; Ferguson & Waxman, 2015), these abstract representation abilities are tightly linked to language (Frank et al., 2009; Thiessen, 2012; Ferguson & Lew-Williams, 2016). Interestingly, Marcus et al. (2007) evidenced that 7-month-olds could generalize abstract rules acquired with speech stimuli to new types of stimuli such as tones, timbres or animal sounds. These findings demonstrate that the acquired abstract representations are not tied to the initial domain, but can rather be transferred to new stimuli types.

A few studies asked whether abstract representations could be used for further computations in infants. Kovacs & Endress (2014) investigated 7 month-olds' ability to use ABB/ABA representations in order to discover higher-level patterns. The authors report mixed results: infants succeed in the task only for some specific repetition-based patterns, and it is not clear whether they can actually manipulate multiple abstract representations. Kovacs & Mehler (2009) report that 12 months-old bilingual infants are able to simultaneously represent two abstract speech structures and pair each structure to a specific spatial location on a screen, as measured with anticipatory eye movement. Surprisingly, infants raised in monolingual environment failed to learn both pairings, but rather exclusively encoded the spatial location associated to the structure based on adjacent repetitions (see also (Kovacs, 2014)). In the visual modality, Hochmann et al. (2016) recently demonstrated that 14 month-olds could use the abstract representation of same in match to sample and non-match to sample experiments. The present study builds on these results, and aims at a better understanding of these abstract representations, and their neural underpinnings.

In two experiments, we asked whether 5 month-old infants could recode an abstract structure into a mental variable that could be used as input for a second computation. More specifically, we designed a paradigm in which infants had to extract and categorize trisyllabic nonce words according to their structure in order to associate each abstract pattern to an arbitrary label. After a brief exposure to the abstract patterns consistently followed by their associated arbitrary labels, we test infants' learning performances by introducing some incongruent associations. Should infants have extracted the abstract patterns, and learned the associations, we expect a surprise response to this incongruency. Using electroencephalography (EEG), combining event-related potentials (ERP) and a frequency tagging approach, we were able to ask which neural mechanisms of the processing hierarchy were affected by infants' expectations. In Experiment 1, infants were exposed to AAB and ABA patterns, and each pattern was associated to a specific image.

In Experiment 2, we confirm and extend the results of Experiment 1. Infants were exposed to ABB and ABA patterns, associated either with an image for one pattern and to a nonce word for the other pattern.

4.2 Experiment 1

4.2.1 Methods

4.2.1.1 Participants

A total of 47 healthy born at term infants were recruited to participate in this study. Among these 47 infants, 2 infants were excluded from analysis due to a technical problem during recording (the synchronization between the stimulation and recording systems was lost). 13 infants were excluded from the analysis of the visual part of the experiment, because of a too small number of artifact-free trials (less than 10 trials in one of the two conditions of interest), leading to a final sample of 32 infants for visual responses analyses (14/18 girls/boys, 21 weeks ± 2 weeks, range 17-26 weeks). Similarly, 6 infants were rejected from the analyses of the auditory part, leading to a final sample of 39 infants for auditory responses analyses (16/23 girls/boys, 21 weeks ± 2 weeks, range 17-26 weeks). The study was approved by the regional committee for biomedical research and parents provided their written informed consent before the experiment.

4.2.1.2 Stimuli

Auditory stimuli consisted of trisyllabic nonce words conforming to either an AAB structure (i.e. the third syllable is different from the two repeated first syllables), an ABA structure (i.e. the second syllable is different from the first and third repeated syllables) and an ABB structure (i.e. the first syllable is different from the two last repeated syllables). To construct the triplets, we used two different vocabularies, one for each experimental part. One vocabulary consisted of 15 CV syllables made of the consonants { 'b', 't', 'k' } and the vowels { 'a', 'u', 'ou', 'in', 'e' } and the second of 15 CV syllables made of consonants { 'p', 'd', 'g' } and vowels { 'an', 'eu', 'i', 'o', 'on' }. The syllables were generated with a fixed duration of 240 ms and the same flat intonation using the MBROLA text-to-speech software (Dutoit, 1997), with French diphones, and digitized at 22050 Hz. The syllables were concatenated to form trisyllabic words (duration: 720 ms), ensuring that the consonant and vowel from syllable A were systematically different from consonant and vowel from syllable B. We constructed 6 distinct sets of 120 words each, corresponding to the three structures (AAB, ABA, and ABB), built out of the two vocabularies.

Visual stimuli consisted of two colored cartoon pictures: a yellow lion and a red fish (Figure 4.1-A). Each visual stimulus was systematically presented over a flickering background oscillating between gray and black at a specific rate of 10Hz or 15Hz depending on the identity of the image (e.g. 10Hz-lion, 15Hz -fish). The flickering frequency associated to each image was counterbalanced across participants. These stimulation frequencies were chosen as a tradeoff between different criteria. First, the stimulation frequencies should not be too high so that the immature visual cortex of 5 month-old infants can follow the stimulation. Second, the stimulation frequencies should not be too slow so that the entrained steady-state does not interfere with the slow dynamics of visual evoked responses. Finally, the two flickering frequencies should be sufficiently distant from one another in order to be separately analyzed.

4.2.1.3 Experimental procedure

The infant was seated in front of a computer screen on the caregiver's lap in a soundproof Faraday booth, after a high-density EEG net (128 channel, EGI; Eugene, OR, USA) was placed on her head relative to anatomical references. Caregivers were instructed not to interfere and to only focus on pacifying the infant. The researcher watched the caregiver and infant from a video camera located above the screen.

A trial consisted in the presentation of an auditory word followed by one of the two visual stimuli (Figure 4.1-A). We used a pair of cartoon eyes located in the middle of the screen as a fixation point. The beginning of a trial was signaled by the eyes blinking for a total duration of 733 ms. Then, a trisyllabic word (AAB, ABA or ABB word, depending on the trial type), was presented, followed by a 980 ms silent period before the occurrence of the carton picture presented on a flickering background during 1000 ms. Then the cartoon eyes reappeared and started blinking again indicating the beginning of the next trial.

Infants were first engaged in a training phase of 36 trials, subdivided into three blocks of 12 trials. This phase introduced them to the consistent dependency between each triplet structure and its associated cartoon picture. During the first 12 trials, one triplet structure (e.g AAB) was followed by one of the pictures (e.g. 10Hz-fish), then the second triplet structure (e.g. ABA) was presented during the following 12 trials associated with the other picture (e.g. 15Hz-lion). Finally during the last 12 trials of the training phase, both structures were randomly presented, followed by the paired visual stimulus. All triplets were randomly drawn from one vocabulary, the other one being kept for the test phase. The order of presentation of the two structures (AAB and ABA), the training vocabulary and the association between a given rule and a given cartoon picture were counterbalanced across subjects.

In the test phase, the vocabulary was changed, and triplets conforming to one of the three structures were randomly drawn in order to have: 1/3 AAB trials, 1/3 ABA trials and 1/3 novel ABB trials. In 75% of AAB and ABA trials, the visual label was congruent with the previously learned association (e.g. AAB words followed by a 10Hz-fish and ABA words by a 15Hz-lion), but in 25% of these trials, the visual label was incongruent with the learned association (e.g. AAB - 15Hz-lion, ABA - 10Hz-fish). Triplets following the novel ABB structure were randomly associated with both images (50%-50%) to prevent any specific learning. Overall, the two visual labels and the three structures were presented with equal frequency. The experiment was stopped when infants became fussy or inattentive.

4.2.1.4 EEG recordings and data pre-processing

EEG was continuously recorded at 250Hz, using a 128-electrode HydroCel net (EGI; Eugene, OR, USA), referenced to the vertex. The onset of the trisyllabic word was sent to the EEG amplifiers thanks to a click signal placed on one channel of the audio file. Similarly, the onset of the image was recorded thanks to a photo diode placed on the screen to detect the display of a white rectangle located on the corner of the image. Infants did not hear the click sent to the EEG amplifiers nor see the white rectangle hidden behind a black mask with the diode.

Pre-processing comprised a high-pass filter above 0.2Hz. Data were further low-pass filtered under 15Hz for the ERP analyses, 40Hz for the frequency tagging analyses and 80Hz for the time-frequency analyses due to the signal to noise constraints required in each analysis. Filtered data were segmented into epochs from -500ms to +1800ms relative to the onset of the stimulus of interest (either words or pictures). Artifact rejection was performed using custom-made MATLAB scripts based on the EEGLAB toolbox (Delorme & Makeig, 2004). For each channel, epochs were rejected either when the fast average amplitude exceeded 250 µV, or when deviation between fast and slow running averages exceeded 150 µV. Electrodes were rejected if more than 60% of the epochs were marked bad, and trials were rejected when more than 70% of the electrodes in that trial were marked bad. For each trial with less than 30% of bad channels, bad electrodes were interpolated using the EEGLAB spherical spline interpolation function. Non-rejected trials were mathematically referenced to the average of the 128 channels, and subsequently baseline-corrected on the 200 ms preceding the onset of the stimulus of interest. The averaged activity was then computed in each condition. The effects

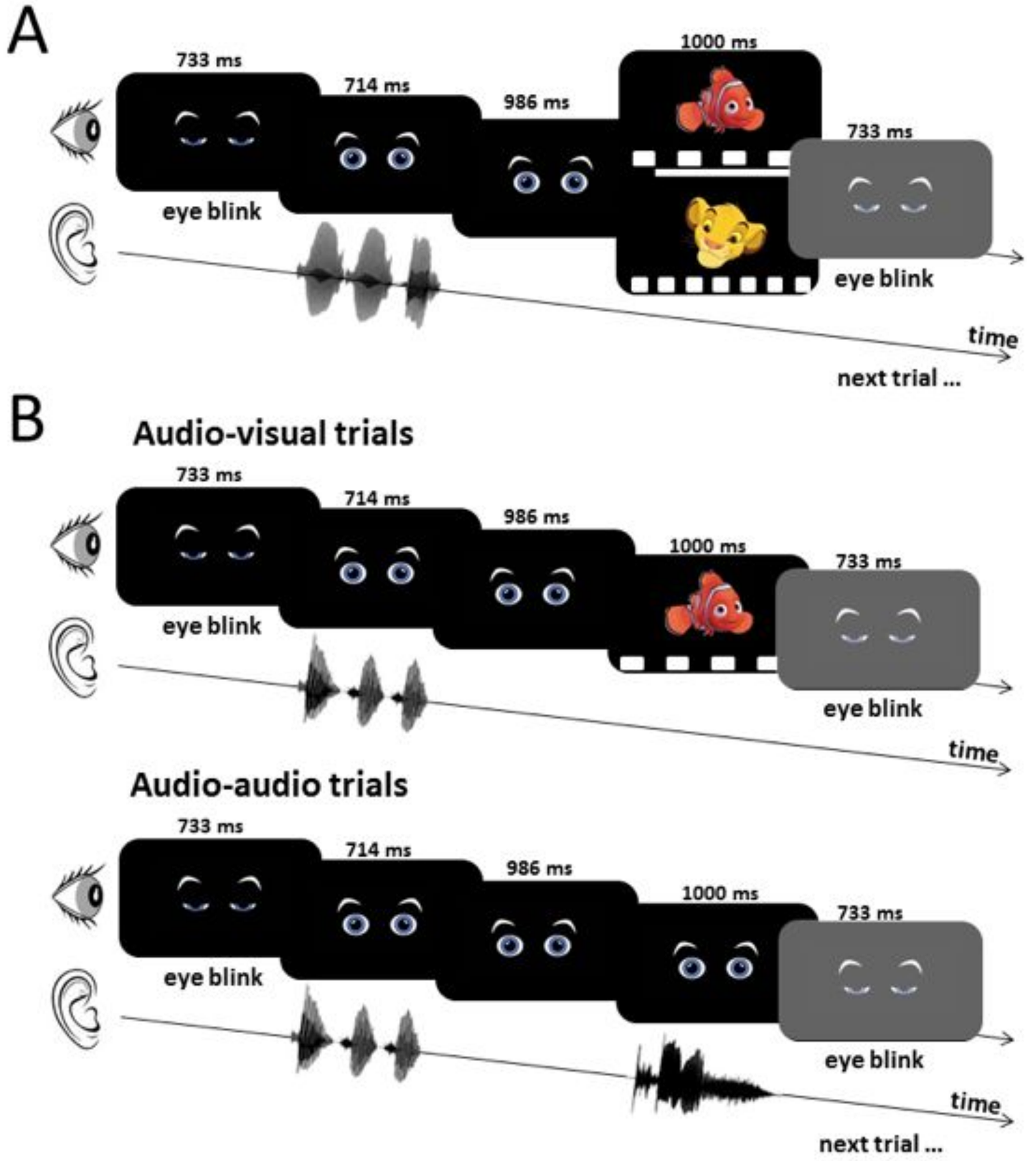


Figure 4.1: **Schematic description of the experimental procedure.** **A**: Trial structure for Experiment 1. The trial starts with the eyes blinking in the middle of the screen. Then infants are presented with a triplet, followed by one second of silence, before the appearance of the visual item. Each visual item is presented on a background flickering at a specific frequency. **B**: Trial structure for Experiment 2. The procedure is very similar to Experiment 1, except that one of the two visual items is replaced by an auditory item (*"Schtroumpf"*). Infants are therefore equiprobably presented with audio-visual trials for which the triplet is followed by a visual item, and audio-audio trials for which the triplet is followed by an auditory item.

we report here are those observed during the test phase.

4.2.1.5 Statistical Analyses

In order to assess whether infants had performed the task, we inspected congruency effects using different approaches. We first computed the typical event related potentials, and investigated whether the different components were affected by experimental conditions. We combined a literature driven analysis for the early and well-described N290 and P400 components, with a less supervised approach using non-parametric statistics for the late and diffuse component. The frequency tagging approach thereafter allowed us to inspect whether early visual processes were modulated by experimental conditions. The obtained steady-state visual evoked potentials (SSVEPs (Regan, 1989)) benefit from two main advantages. First, these responses are relatively immune to artifacts, and thus show a high signal-to-noise ratio. Second his methodology allows a straightforward analysis procedure, as it is objective and predictive: the response appears exactly at the periodic stimulation frequency over occipito-temporal electrodes. Finally, we inspected the time-frequency decomposition of the visual responses, as cortical oscillations are known to be involved in many perceptual and cognitive operations (X. Wang, 2010), especially for cross-modal integration (Senkowski et al., 2008). We used an unsupervised statistical analysis approach, performing non-parametric statistics.

4.2.1.5.1 Visual responses

Visual trials (i.e. locked to the onset of the image) were sorted by condition (Congruent or Incongruent) in each infant. Infants with less than 10 trials in one of the conditions were rejected from further analyses. On average, we obtained 71 artifact-free trials per subject in the Congruent condition, and 22 artifact-free trials per subject in the Incongruent condition. Given that the proportion of Congruent and Incongruent trials was strongly imbalanced (75% vs 25% respectively), we selected a subset of the Congruent trials in order to match the number of Incongruent trials. For each incongruent trial, we selected the closest congruent trial, controlling for the identity of the image (e.g. for each fish-Incongruent trial we selected the closest fish-Congruent trial).

ERP analyses We first collapsed all conditions together in order to identify the main ERP components evoked by the visual stimulation. We observed the classical components expected at this age: an early N290 over occipito-medial electrodes (110-290 ms), a P400 over occipito-temporal electrodes (300-550 ms), and a Late Slow Wave with a more distributed topography over the scalp (1000-1800 ms). We analyzed the N290 and P400 using a cluster-based approach. For both components, voltages were averaged across the clusters and time-windows of interest and thereafter submitted to an ANOVA, with condition (Congruent vs Incongruent) as a within-subject factor. Because late components are less characterized in the infant literature, we could not run our analyses on an obvious set of electrodes. We thus used non-parametric statistics: a clustering analysis coupled with a randomization procedure (Maris & Oostenveld, 2007). After computing a two-tailed paired t-test between the incongruent and congruent conditions for each sample, spatio-temporal cluster statistics were defined as the sum of the t-values above a threshold ($\alpha = 0.1$) on contiguous points both in terms of space and time, separately for positive and negative t-values. Because our data were transformed to an average reference, ERP components can present two near-simultaneous activations of opposite signs over two distinct electrode sets. We thus implemented an additional step allowing for the detection of such dipolar components (Izard et al., 2008). We combined clusters of opposite signs over two distinct electrode sets, overlapping in time, into one dipolar cluster. The dipolar cluster statistic was defined as the sum of the absolute value of the statistics of its constituting positive and negative clusters. To evaluate the significance of the effects, we then recomputed the same analysis on 5000 sets of randomly permuted data. A permutation was defined by randomly attributing the label “Congruent” or “Incongruent” to the two averaged responses to congruent and incongruent trials for each infant. For each permutation, we extracted the value of the maximum effect strength for the total time interval, thus building a distribution of maximal effect strength. The corrected p-values corresponded to the rank of the experimental data within this distribution of maximal effect strength, divided by the total number of permutations. This method allowed revealing the clusters of electrodes and the time-windows that were significantly affected by our experimental conditions.

Frequency tagging analyses We analyzed the entrained activity in response to the flickering stimulation during the duration of the picture presentation (1000 ms) delayed by 200 ms to compensate for the conduction time of the stimulation to the visual cortices (i.e. 200-1200 ms after stimulus onset). For each electrode and each trial, we computed the power spectrum of the signal in this time-window, using the fast Fourier transform algorithm as implemented in MATLAB. For each subject, we thereafter averaged the power spectra across trials, for each condition of interest depending on the stimulation frequency (Congruent-10Hz, Incongruent-10Hz and Congruent-15Hz, Incongruent-15Hz). From the averaged power spectra, we computed a normalized power measurement at both 10Hz and

15Hz, for each electrode: the power at the frequency of interest was divided by the power averaged over surrounding frequency bins (N=3 frequency bins before and after the frequency bin of interest). We restricted this analysis to an occipito-temporal cluster, where frequency-tagged visual responses are classically maximal (de Heering & Rossion, 2015; Norcia et al., 2015). We averaged normalized power measurements over this region of interest and submitted these values to a 2x2x2 ANOVA with condition (Congruent vs Incongruent), stimulation frequency (10Hz vs 15Hz) and analysis frequency (Stimulation frequency vs Non-Stimulation Frequency) as within-subject factors.

Time-frequency analyses We additionally inspected the pattern of oscillatory power over a broader range of frequencies, from 8Hz to 60Hz. A time-frequency wavelet transform was applied to each trial and each electrode, using a family of complex Morlet wavelets, as implemented in the EEGLAB toolbox. The number of cycles in each Morlet wavelet was set as increasing linearly with frequency, form 4 at the lowest to 15 at highest. This computation resulted in an estimate of oscillatory power with a time resolution of 12ms, and frequency resolution of 1Hz. The time-frequency decomposition was then baseline corrected from 200ms to 20ms before stimulation onset, and finally averaged across trials for each condition (Congruent and Incongruent). We thereafter restricted the analyses to the period of visual stimulation here again delayed by 200 ms to compensate for the conduction time (200-1200ms after picture onset), and we performed a clustering and randomization procedure to identify significant differences between Congruent and Incongruent trials, similar to the procedure previously described. The α-value was set to 0.05, and the cluster statistics were compared to those of 5000 random permutations.

4.2.1.5.2 Auditory responses

ERP analyses For each infant, we sorted the auditory trials (i.e. locked to the onset of a triplet) by condition (AAB, ABA or ABB). Infants with less than 10 trials in one of the conditions were rejected from further analyses. On average, we obtained 42 artifact-free trials per subject and per condition (41/42/42 for AAB, ABA and ABB). We first merged the three conditions to identify the main auditory components evoked by the triplets, independently of the condition, but also late components that may indicate an anticipation of the expected label. Indeed, we identified a stable late component starting from roughly one second after the onset of the triplet until the onset of the image: it consisted of a strong negativity over central electrodes, and positivity over occipital electrodes. We thereafter computed the evoked responses per condition in each infant and averaged the voltage across the negative cluster of electrodes, over a time-period corresponding to the peak of the main effect (1000-1500ms after word onset). These values were submitted to a three-way ANOVA, with Structure (AAB, ABA or ABB) as a within subject factor.

4.2.2 Results

4.2.2.1 Congruency effects

Cluster-based analysis did not reveal any significant effect neither for the N290 component ($F(1, 31) < 1$), nor for the P400 component ($F(1, 31) < 1$). However, the clustering and permutation analysis of the late time-window indicated a significant negative cluster ($p_{corr} = 0.039$) for the Incongruent-Congruent contrast, comprising 15 electrodes over right central regions, and spreading in time from 1.02 seconds until 1.8 seconds (Figure 4.2-A).

We thereafter inspected the brain responses elicited by the flickering background during the image display. As expected, there was a significant main stimulation effect ($F(1, 31) = 88.03$, $p < 10^{-9}$) corresponding to a stronger power at the stimulation frequency (Figure 4.3-A-B). The entrainment was also stronger at 10 Hz than 15 Hz ($F(1, 31) = 13.5$, $p < 10^{-3}$). Finally, the experimental condition (congruent vs incongruent pairing) was modulating the steady-state response as revealed by the significant Condition x Analysis Frequency interaction ($F(1, 31) = 5.99$, $p = 0.02$): the entrained activity was weaker in the Incongruent condition compared to the Congruent condition.

Inspection of brain responses over a broader frequency range additionally revealed a significant negative component ($p_{corr} = 0.026$) in the low gamma band spanning over 54 left and right temporal electrodes, for the Incongruent-Congruent contrast. This component extended in time between 200 ms and 472 ms, and in frequency between 21Hz and 46Hz (Figure 4.3-C).

4.2.2.2 An effect related to the triplet structure

A significant effect related to the triplet structure was observed ($F(2, 76) = 5.05$, $p = 0.008$) (Figure 4.2-B). Post-hoc statistical analyses revealed a stronger central negativity for AAB compared to ABB triplets ($F(1, 38) = 4.54$, $p = 0.04$), and to ABA triplets ($F(1, 38) = 6.75$, $p = 0.01$). On the contrary, the comparison between AAB and ABA triplets did not reveal any significant difference ($F(1, 38) = 1.87$, $p > 0.1$).

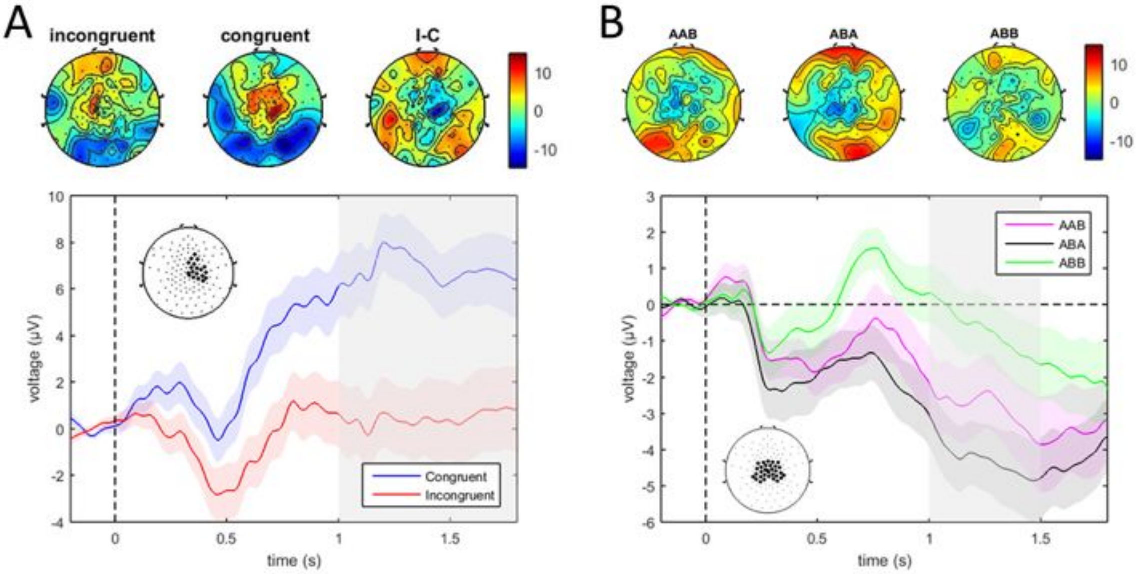


Figure 4.2: **ERP responses, Experiment 1. A**: Grand average brain responses evoked by the appearance of the visual item (averaged across the congruent and incongruent conditions) for the significant spatio-temporal negative cluster evidenced by clustering and permutation analysis. **B**: Grand average brain responses evoked by the triplets (averaged across the different structure conditions, AAB, ABA and ABB) for the late central cluster showing a strong negativity on the main response.

4.2.3 Discussion

During the test phase, we identified three neural events related to infants' learning that are easier to understand if we unstack them from the latest one: first (or last), a stronger positivity was recorded over the right central electrodes for congruent trials relative to incongruent trials. This difference became significant after the image disappeared (Figure 4.2-A). We relate this response to the infants' surprise when incongruent pairings occurred. Second, during the image presentation, the strong oscillatory activity driven by the image flickering background was increased for congruent relative to incongruent trials (Figure 4.3-A-B), revealing that infants were expecting a precise image. This interpretation is confirmed by the difference recorded during the waiting period between the triplets which predicted a precise image (AAB and ABA triplets), and the ABB triplets, for which infants cannot form any expectation (Figure 4.2-B). We will successively discuss these effects and our interpretation.

Surprise effect Our main initial hypothesis was that if infants were able to learn the pairing between the triplet structure and an image, we should therefore observe an orientation of attention to an unexpected association. We observed a significant sustained central negativity starting around 700 ms second after image onset in incongruent trials. The conservative cluster analysis approach we used estimated this response to be significant only after 1 sec (Figure 4.2-A). This late ERP component belongs to the large family of late responses, which have been related to higher-level functions such as orientation of attention to a novel event (Dehaene-Lambertz & Dehaene, 1994), recognition memory, and updates in memory representations (de Haan et al., 2003), access to consciousness (Kouider et al., 2013), and surprise (Kouider et al., 2015). As these responses exhibit functional characteristics that are similar to the adult P300 (Basirat et al., 2014; Kouider et al., 2013), they have been proposed as indexing equivalent processes. Their slow latency and sustained activity of have been related to the task difficulty for this age and to the weak myelination of long-distance fiber tracks.

Frequency-tagging Using a flickering background, during the 1-second visual display, we successfully elicited a strong stimulus-driven oscillatory activity at the stimulation frequency in early visual cortices. This oscillatory activity decreased for incongruent pairings, suggesting that the processing of the low-level features of the stimuli (brightness changes over time) was impaired (Figure 4.3-A-B). This phenomenon could either reflect an impaired entrainment due a surprise signal after the detection of the incongruency, or, more interestingly, it could reflect an enhanced entrainment when the visual stimulation fits with infants' expectations (i.e. fast oscillations or slow oscillations). In any case, this early brain response modulation reflects top-down connections from higher-level areas down to early sensory areas. It additionally suggests that infants were preparing for a precise image after experiencing the training structures. Note that the use of a flickering background allowed us to derive a more sensitive measure for early sensory responses than traditional ERPs as we did not find any difference in the analyses of the early visual ERP. The higher sensitivity of the frequency tagging approach was already underscored by de Heering et al. (2015).

Time-frequency decomposition This analysis provided additional evidence for congruency effects on early brain responses. An early gamma-band enhancement was observed in the temporo-frontal regions, relative to baseline, for the congruent condition (Figure 4.3-C). Gamma-band activity is known to play an important role for cross-modal perceptual integration. An early gamma activity enhancement has been reported in symbol-to-sound matching tasks, when sounds were congruent with a visual symbol, both in adults (Widmann et al., 2007; Yuval-Greenberg & Deouell, 2007; Schneider et al., 2008), and 11 month-old infants (Asano et al., 2015). This early gamma effect might therefore reflect a top-down integrative process between the abstract representation of the triplet structure held in working memory and its associated label. These results show that gamma-band oscillatory activity is functional as early as 5 month after birth, and this high-frequency activity already seems to support associative learning.

Waiting period Observing early neural events related to congruency during the image presentation suggests that infants may be prepared during the waiting period separating the sound from the image. Using NIRS, Emberson et al. (Emberson et al., 2015) reported that 4-month-old infants habituated to a consistent tone-smiley association, and displayed a visual response in rare trials in which the smiley was omitted. This omission response clearly demonstrates that top-down activations can prepare sensory regions to receive an expected stimulus. Here the grand average ERP response to AAB and ABA trials revealed a large central negativity which was absent in response to the ABB trials (Figure 4.2-B). Although we cannot rule out the possibility that this different might be related to a novelty effect after the new ABB structure, it could alternatively reflect the preparation of the visual regions to receive a precise input. In any case, this difference reveals that infants were processing differently the training structures from the new structure, demonstrating that learning was not limited to the detection of an immediate repetition in the triplet (i.e: they did not confound AAB and ABB triplets) and that their representation was not limited to one type of structure, and opposing it to any other structure.

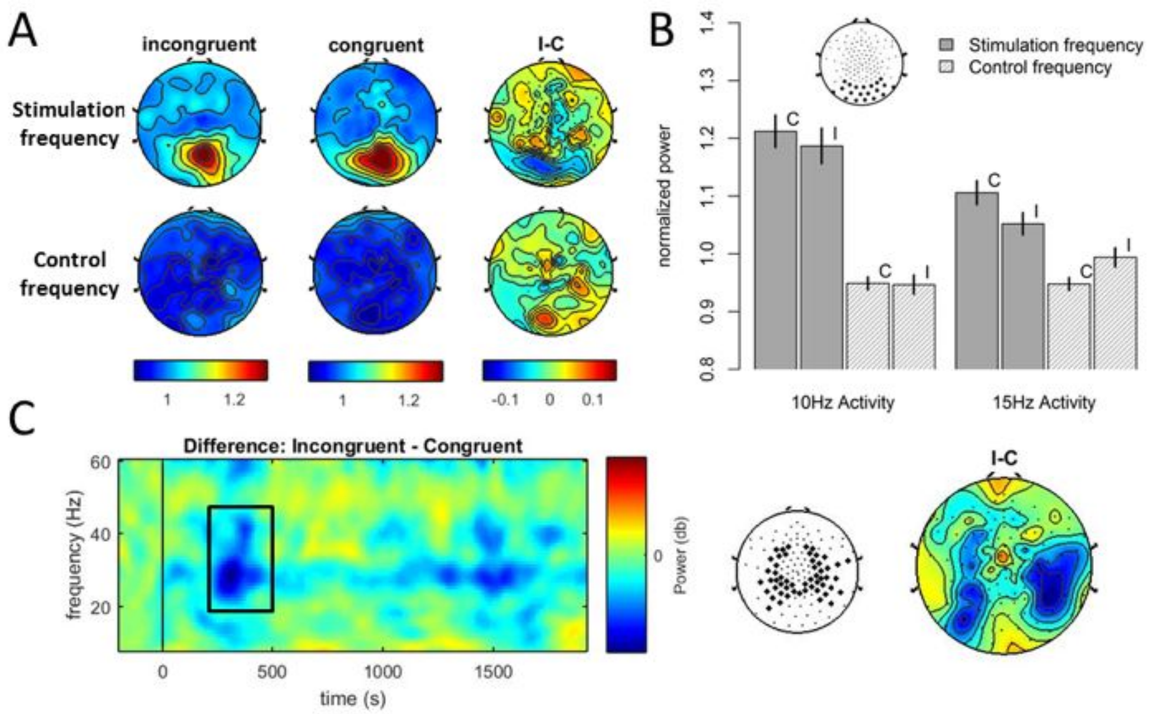


Figure 4.3: **Frequency tagging and time-frequency responses, Experiment 1.** **A**: Topographic plots of the normalized power of the oscillatory activity averaged over infants, at both stimulation frequency and control frequency. Separate columns show responses to either congruent and incongruent pairings, or the difference between incongruent and congruent conditions. **B**: Bar plots illustrating normalized power averaged over occipito-temporal electrodes, at both 10 Hz and 15Hz, in response to incongruent (I) or congruent (C) pairings. Bar colors indicate whether the oscillatory activity corresponds to the stimulation frequency (dark grey), or a control frequency (light grey). **C**: Time-frequency and topographic plots of the early gamma-band effect. The time-frequency plot represents the difference between responses to incongruent and congruent pairings, averaged over significant electrodes, as evidenced by the clustering and permutation analysis. The topographic plot represents the difference between incongruent and congruent trials, averaged over the significant time-frequency cluster.

In a nutshell, 5 month-old infants were able to associate arbitrary visual labels with abstract auditory patterns in a few trials. They not only extract and simultaneously represent multiple auditory structures based on syllable repetitions, but they can chain a second operation and attach a label to this abstract categorization.

Eventually, it is worth mentioning that Experiment 1 does not strictly demonstrate generalization from training vocabulary to test vocabulary. Indeed, during test, 75% of the AAB and ABA trials are congruent, thus allowing infants to learn regularities only based on the test vocabulary. Nevertheless, this hypothesis remain highly unlikely as first, triplets were constructed such that all syllables could be either in A, or B position of the AAB and ABA structures, thus decreasing the co-occurrence probabilities between syllables, and second, we report on average only 15% of triplet repetition across all test duration. We control for this possible confound in Experiment 2.

4.3 Experiment 2

Experiment 2 intends to replicate and extend the results obtained in Experiment 1. As infants might be preparing early sensory regions to operate on the forthcoming label during the waiting period, we defined two different labels, belonging to two different modalities (visual and auditory) in order to target two different sensory systems (Figure 4.1-B). We also changed the two training structures to ABA and ABB and as in experiment 1, introduced the third structure (AAB) during the test phase. Finally, we better controlled for generalization from the training vocabulary to the test vocabulary: during the test phase, we defined some critical trials for which we used the test vocabulary, with 50% of congruent trials and 50% of incongruent trials, so that infants could not learn any vocabulary specific dependency (Table 4.1). Besides, only those critical trials were considered in the congruency analyses. All remaining trials were congruent, using the training vocabulary.

4.3.1 Methods

4.3.1.1 Participants

A total of 57 healthy infants were recruited to participate in this experiment. From these 57 infants, 2 infants were excluded from further analyses due to technical problems, and one infant refused to wear the net. In order to minimize the attrition rate, brain responses to the auditory (*"Schtroumpf"*) and to the visual (the red fish) labels were processed separately. Infants showing less than 10 trials in the congruent or the incongruent condition were rejected from further analyses. We thus report on a group of 34 infants (20/14 girls/boys, 20 weeks ± 2 weeks, range 18-24 weeks) for the analyses of the visual label, and 32 infants (19/13 girls/boys, 20 weeks ± 2 weeks, range 18-24 weeks) for those of the auditory label. Finally, we report on a group of 49 infants (31/18 girls/boys, 20 weeks ± 2 weeks, range 18-24 weeks) for the analyses of the anticipatory responses. The experiment was approved by the regional committee for biomedical research and parents provided their written informed consent before the experiment.

4.3.1.2 Stimuli and procedure

The red fish was the same as in experiment 1 and always presented over a 10-Hz flickering background. The auditory label consisted of a nonce monosyllabic word (*"Schtroumpf"*) recorded from a female native speaker of French in an infant-directed speech register. The duration of this auditory word (0.955 s) roughly matched the duration of the visual label (1 s) (Figure 4.1-B).

The structure of the experimental paradigm was identical to Experiment 1 (Figure 4.1-B): during a training phase of 36 trials (3 blocks of 12 trials), infants were first introduced with the consistent dependency between each structure (ABB and ABA triplets) and their associated label (the 10 Hz-fish and the *"Schtroumpf"*). The order of presentation of the two structures, the training vocabulary and the association between a given structure and the visual/auditory label were counterbalanced across subjects.

In the test phase, triplets conformed either to ABB, ABA or AAB structures. In order to test for generalization, a subset (50%) of critical ABB and ABA trials was defined, for which the vocabulary was different from the training phase. Crucially, for a half of these critical ABB and ABA trials, the associated label was congruent with the previously learned association (e.g. ABB words followed by the 10Hz-fish image and ABA words followed by the *"Schtroumpf"* word), while for the remaining half of these critical trials, the associated label was incongruent with the learned association (e.g. ABB words followed by the *"Schtroumpf"* word and ABA words followed by the 10Hz-fish label). The remaining non-critical trials (50% of ABB and ABA trials) aimed at reinforcing the triplet-label dependency. For those trials, the association between triplets and auditory or visual label was therefore systematically consistent with the associations defined during the training phase, and the training vocabulary was used. In total, 75% of ABB and ABA trials conformed to a congruent structure-label association, and 25% of ABB and ABA trials conformed to an incongruent structure-label association. Finally, triplets conforming to the novel AAB structure were randomly associated

Structure	**Vocabulary**	**Condition**
ABB	50% training vocabulary	50% congruent pairings
	50% test vocabulary	25% congruent pairings 25% incongruent pairings
ABA	50% training vocabulary	50% congruent pairings
	50% test vocabulary	25% congruent pairings 25% incongruent pairings
AAB	50% training vocabulary	25% auditory label 25% visual label
	50% test vocabulary	25% auditory label 25% visual label

Table 4.1: **Trials types' distribution during test phase.** For each triplet structure, the table describes the proportion of trials that involved the training or test vocabulary, and the proportion of congruent and incongruent pairings (for ABB and ABA trials), or the proportion of auditory and visual pairings (for AAB trials)

with both labels, 50% visual - 50% auditory. The new vocabulary was used in 50% of the AAB trials, and the training vocabulary was used in the remaining 50% of the AAB trials (Table 4.1).

The two labels and the three structures were presented with equal frequency. The experiment was stopped when infants became fussy or inattentive.

4.3.1.3 EEG recordings and data processing

The recording apparatus, as well as the preprocessing parameters were identical to those used for Experiment 1. Congruency effects were separately analyzed for the visual and auditory labels.

4.3.1.4 Statistical analyses

The statistical approach was similar to experiment 1: for each modality (auditory and visual) we computed the ERP and time-frequency responses, and inspected congruency effects.

4.3.1.4.1 Neural responses to the visual label

Brain responses to the image in critical trials were sorted by condition (Congruent or Incongruent) in each infant. Infants with less than 10 trials in one of the two conditions were rejected from further analyses. We recorded 15/16 artifact-free trials on average per subject in the Congruent/Incongruent conditions. **(**ERP) and **(**time-frequency) analyses were identical to experiment 1.

Frequency tagging analyses The activity at 10Hz and 15Hz (as a control condition) were retrieved using the same computations than in experiment 1. Note however, that only one stimulation frequency (10Hz) was used here, the 15Hz activity corresponding to the spontaneous brain activity in this frequency range. Peak measurements over the same occipito-temporal region were thereafter submitted to a 2x2 ANOVA with condition (Congruent vs Incongruent) and analysis frequency (10Hz vs 15Hz) as within-subjects factors. We additionally compared entrained brain activity elicited by 10Hz stimulation between Experiment 1 and Experiment 2. The 10 Hz and 15 Hz brain responses during 10 Hz stimulation were computed from the main visual response (all conditions merged together), and submitted to a 2x2 ANOVA, with analysis frequency (10Hz vs 15Hz) as a within-subject factor, and experiment as a between subject factor (Experiment 1 vs. Experiment 2).

4.3.1.4.2 Neural responses to the auditory label

Brain responses to the auditory nonce word (*"Schtroumpf"*) in critical trials were sorted by condition (Congruent or Incongruent) in each infant. Infants with less than 10 trials in one of the two conditions were rejected from further analyses. On average we obtained 16/16 artifact-free trials per subject in the Congruent/Incongruent conditions.

ERP analyses We used the same approach as described for visual responses in experiment 1 to analyze auditory evoked responses. We first inspected the main brain response to the auditory word, collapsing all conditions together. We observed 3 temporal windows of interest: 150-450 ms displaying a positive activation over the temporo-inferior frontal regions in both hemispheres synchronous with a median posterior activation, 600-900 ms during which the positive activation moved towards median-frontal areas whereas the posterior negativity expanded laterally, and a late component (1000-1800 ms) with a more distributed topography. We analyzed the two early frontal components using a cluster-based approach, averaging the voltage over these clusters and time-windows of interest for each infant and submitting these values to a two-way ANOVA with Condition (Congruent vs Incongruent) as a within subject factor. The late component was analyzed using a clustering analysis coupled with a randomization procedure (Maris & Oostenveld, 2007), as described in Experiment 1.

Time-frequency analyses The analysis procedure for assessing congruency effects over time-frequency responses was identical to the one used for Experiment 1.

4.3.1.4.3 Neural responses to the triplet structures

ERP analyses In experiment 2, we aimed at testing infants' anticipatory abilities. We hypothesized that the central negativity observed in experiment 1 during the waiting period before the occurrence of the image for the labeled structures was related to some neural preparation of the visual regions to receive a given image. In experiment 2 we used two labels belonging to two different sensory modalities in order to decode from this waiting period neural preparation in the auditory regions in response to the structure associated to the nonce word, or in the visual regions in response to the structure associated to the red fish. We therefore sorted brain responses to the auditory triplets depending on whether they were associated to an auditory or a visual label (Audio or Visual conditions). On average, we report 49 artifact-free trials per subject and per condition (49/49 for Audio and Visual respectively). We first merged the two conditions to inspect the main response during the waiting period. As the elicited topography appeared rather distributed over the scalp, and since we had no strong a priori expectation on the location or timing for the auditory or visual preparation responses, we used the unsupervised randomization and clustering procedure during the whole waiting period (1000-1700 ms), setting the α-value to 0.1.

4.3.2 Results

4.3.2.1 Congruency effects

Cluster-based analysis of early sensory responses to both auditory and visual labels did not reveal any congruency effect (in the case of the visual label: $F(1,33) < 1$ for both the N290 and P400 components; in the case of the auditory label: $F(1,31) < 1$ for the 150-450 ms time-window, and $F(1,31) = 1.8$, $p = 0.19$ for the 500-950 ms time window). However, using the clustering and randomization procedure to analyze the late brain responses, we observed significant congruency effects for both auditory and visual labels. The visual incongruency yielded a significant effect ($p_{corr} = 0.045$) including a right centro-temporal negative cluster (17 electrodes) and a left fronto-central positive cluster (12 electrodes), and extending from 1.268 seconds to 1.8 seconds after stimulus onset (Figure 4.4-A). This dipolar effect was due to a more negative right centro-temporal activity, and a more positive left frontal activity in the incongruent condition. The auditory incongruency yielded a significant negative effect ($p_{corr} = 0.048$) over a cluster of right posterior electrodes (13 electrodes), spanning between 1 second and 1.8 seconds after stimulus onset, with more negative activity in response to incongruent trials (Figure 4.4-B).

In the case of triplet-image pairing trials, we additionally analyzed the entrainment response to the flickering background (Figure 4.5-A-B). We observed a main effect of the stimulation frequency ($F(1, 33) = 13.12$, $p < 10^{-3}$). The ANOVA additionally revealed a marginal condition effect ($F(1, 33) = 4.13$, $p = 0.05$), with more power in the incongruent than in the congruent condition. Crucially, there was no significant Condition x Analysis Frequency interaction ($F(1, 33) < 1$) as evidenced in experiment 1. Besides, a main effect of experiment was observed during the 10 Hz stimulation between experiment 1 and experiment 2 ($F(1, 64) = 33.37$, $p < 10^{-6}$), due to a power increase in experiment 1 relative to experiment 2 (Figure 4.5-C). The significant Experiment x Analysis Frequency interaction ($F(1, 64) = 31.30$, $p < 10^{-6}$), indicates that this stronger power in Experiment 1 was mainly driven by the entrained activity at 10 Hz, and not the control spontaneous activity at 15 Hz. We also captured the main Analysis Frequency effect ($F(1, 64) = 90.23$, $p < 10^{-12}$) over the two Experiments, confirming that the flickering background efficiently induced an enhanced activity at stimulation frequency.

Finally, analyzing the time-frequency decomposition of the brain responses did not reveal any significant effects.

4.3.2.2 Anticipatory effects

The analysis of the waiting period using the conservative randomization and clustering procedure did not reveal any significant effect.

4.3.3 Discussion

As in experiment 1, an incongruent pairing yielded a late and slow response over central electrodes (Figure 4.4). We thus confirm that 5-month-old infants can extract and represent multiple auditory structures. We additionally show that

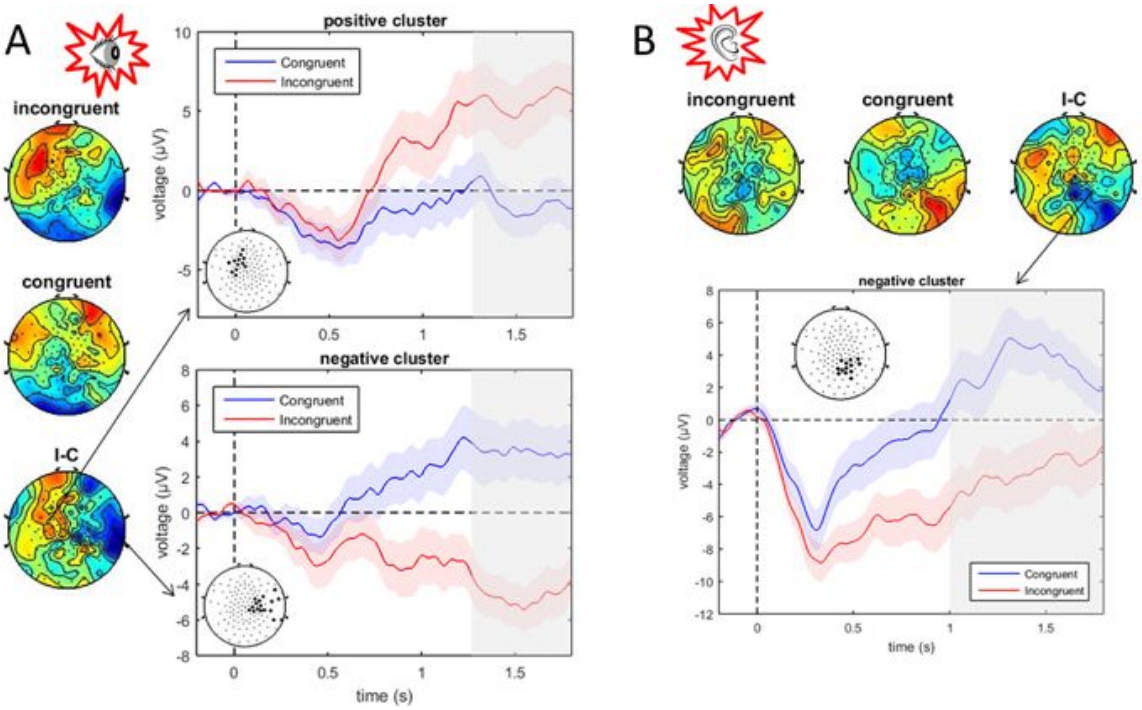


Figure 4.4: **ERP responses, Experiment 2.** **A**: Grand average brain responses evoked by the appearance of the visual item (averaged across the congruent and incongruent conditions) for the significant spatio-temporal dipole evidenced by clustering and permutation analysis. The time-course of both positive and negative clusters are depicted. **B**: Grand average brain responses evoked by the appearance of the auditory item (averaged across the congruent and incongruent conditions) for the significant spatio-temporal negative cluster evidenced by clustering and permutation analysis.

infants can pair these representations with either a visual or an auditory label. However, several differences in the results of both experiments might suggest that infants were not similarly engaged in the task probably due to differences in task difficulties.

First, although we observe a clear steady-state evoked response at 10 Hz in early visual cortices during the presentation of the flickering visual stimulation, this entrained activity was weaker in experiment 2 relatively to experiment 1 (Figure 4.5-C) suggesting a different attentional engagement in the visual displays in both experiments. In Experiment 1, infants crucially have to encode the specific identity of the image in order to perform the task, enhancing the neural processing of the different visual features through attention. On the contrary, in Experiment 2, infants only need to note whether it was an image or a sound. The features of the image are then irrelevant to the task, and visual attention can be lowered. In adults' studies (S. T. Morgan et al., 1996; Muller et al., 2006; Kim et al., 2007), the strength of steady-state visual evoked responses increases with the amount of visual attention needed to make sense of the input. Young infants also adapt their attentional load to the complexity of the task (Kidd et al., 2012, 2014). Thus in line with these studies, our results suggest that 5 month-olds already benefit from top-down mechanisms allowing them to optimize the neural resources they allocate to a given task.

This different strategy may additionally explain why we did not replicate the power enhancement at the flickering frequency for congruent trials in experiment 1. We observed instead a global power enhancement of the signal generalizing to the 15 Hz (which was not entrained) for incongruent pairings (Figure 4.5-A-B). This global response can be explained by a reallocation of attention toward the visual stimulus when it was unexpected enhancing the signal to noise ratio across all frequencies.

The time-frequency analysis did not reveal any significant effect, neither in response to auditory word, nor in response to the image. This difference could be explained by the same task discrepancy between the two experiments as discussed above, or it could alternatively be due to a higher variability in the case of experiment 2. Indeed, while in Experiment 1 infants were presented with on average 22 trials in each condition (Congruent and Incongruent), in Experiment 2, responses to congruent and incongruent trials had to be analyzed separately for visual and auditory labels. Infants therefore only experienced 15/16 trials on average in each condition (Congruent and Incongruent) for each modality (auditory and visual labels). Pediatric research typically comes with low numbers of trials, and a few trials more can represent a significant improvement in terms of variability reduction and statistical power.

Finally, we were unable to evidence any correlate of active neural preparation to the occurrence of the forthcoming auditory or visual label during the waiting period, after the occurrence of the triplet structure. Although we cannot rule-out the fact that infants might not actively anticipate the auditory of visual labels, this failure might be alternatively due to the fact that the analysis procedure we used was too conservative for evidencing such subtle processes. Some different analyses might be better fitted for evidencing such anticipatory activations.

4.4 General discussion

All together, these two experiments provide strong evidence that 5 month-old infants can simultaneously represent multiple auditory structures, and pair these representations with arbitrary visual or auditory objects. These results extend previous findings on abstract learning, as we demonstrate that young infants can form abstract representations of speech structures reliable enough to be associated with an arbitrary label. While the original Marcus et al. (1999) design only allowed testing dishabituation performances from one triplet structure to another, we go beyond discrimination skills as we test infants' ability to represent each structure, as a specific abstract entity, that can be used as an input for further analyses in order to discover the systematic pairing with an arbitrary label. We additionally show that infants can distinguish the two paired triplet structures from a third one, attesting that infants' performances did not reflect a mere repetition detection mechanism or edge bias. This study additionally complements previous behavioral results suggesting that preverbal infants have difficulties simultaneously representing more than one triplet structure (Kovacs & Mehler, 2009; Kovacs, 2014). In these studies, infants had to associate each structure with the left or right side of the screen. The authors used anticipatory fixations as a correlate of their learning performances. This task is particularly challenging for participants both in terms of spatial knowledge and executive control. First, infants have to form representations for the left and right sides of a monitor, which is not trivial. Then they have to associate each triplet structure to one of these spatial locations; and finally they must have robust executive control to orient their eyes towards the appropriate spatial location. Although infants can perform each of these tasks separately, their failure might reflect the complexity of integrating the different cognitive processes rather than their inability to represent multiple abstract structures. Using electrophysiology, we circumvent executive control issues as we measure surprise

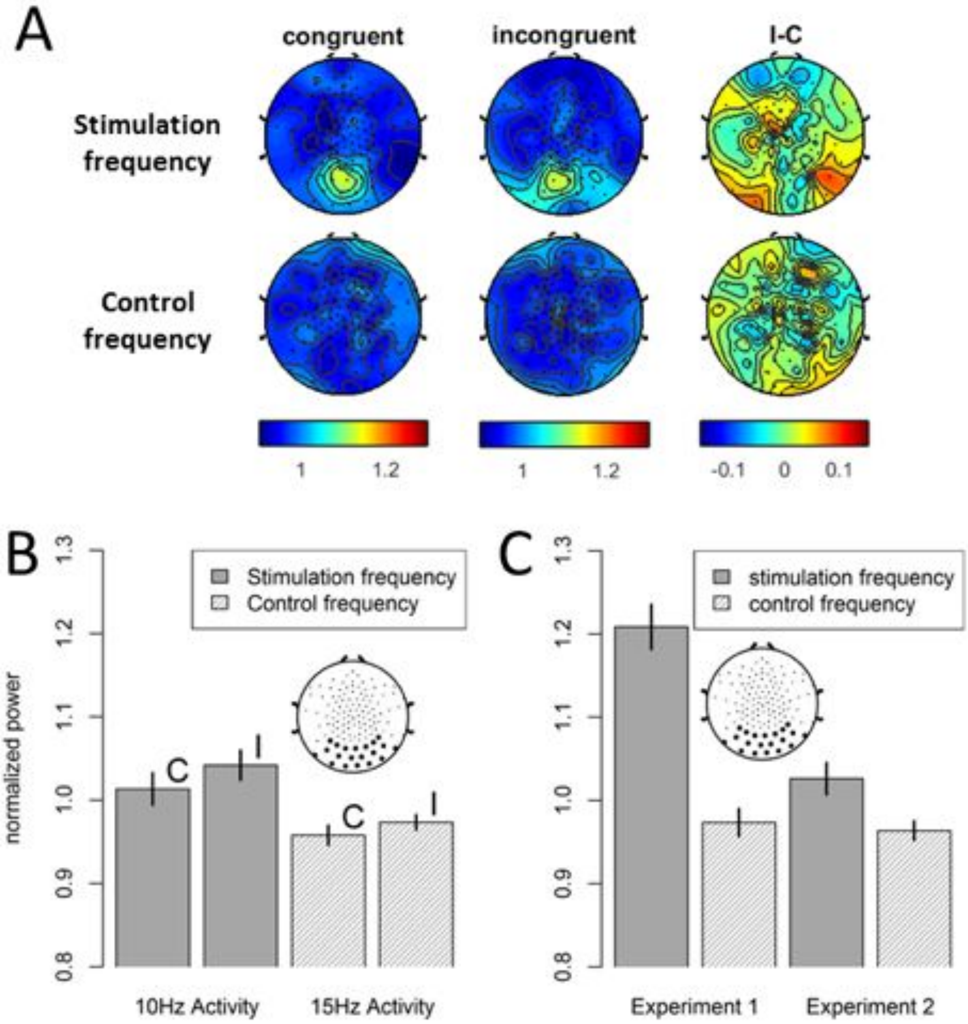


Figure 4.5: **Frequency tagging responses, Experiment 2.** **A**: Topographic plots of the normalized power of the oscillatory activity averaged over infants, at both stimulation frequency and control frequency. Separate columns show responses to either congruent and incongruent pairings, or the difference between incongruent and congruent conditions. **B**: Bar plots illustrating normalized power averaged over occipito-temporal electrodes, at both 10 Hz (stimulation frequency, dark grey) and 15Hz (control frequency, light grey), in response to incongruent (I) or congruent (C) pairings. **C**:Bar plots illustrating the strong difference in SSVEP power between Experiment 1 and Experiment 2, averaged over occipito-temporal electrodes. The evoked oscillatory activity by the stimulation at 10 Hz is much stronger in Experiment 1, compared to Experiment 2.

responses directly from brain activity. Besides, it might be easier to associate the abstract representation to a specific label, rather than a spatial location. The two labels might even act as cues for categorizing the two abstract structures.

These results also echo previous research on non-human cognition, asking whether abstraction is a human specific ability. A wide range of animals, from bees to apes, can be trained to master some abstract representations of information. Extensive work has investigated the acquisition of abstract representations for same and different relations among many species. Using identity match-to-sample protocol, it has been evidenced that honey bees (Giurfa et al., 2001), pigeons (Wright et al., 1988), rats (Mumby, 2001), infant chimpanzees (Oden et al., 1988) or dolphins (Herman & Gordon, 1974) can learn to compare between a memorized sample and a perceived test stimuli, and adapt their behavior depending on whether sample and test are similar or different. Although these results demonstrate some abstraction abilities, it does not require actually encoding an abstract representation of same or different (Premack, 1983). In relational match-to-sampling procedures, however, subjects have to compare more than individual items. They are instead presented with a set of multiple objects representative of an internal relation of same or different, and participants are trained to select another array of stimuli that matches (or not) the retained relation, irrespective of the identity of the stimuli. Primates (Wasserman et al., 2001), as well as several bird species (Smirnova et al., 2015; Zentall et al., 2002; Pepperberg, 1987; Martinho & Kacelnik, 2016) have succeeded in solving such problems, suggesting that they can actually represent and operate on the concepts of same and different. However, careful analyses of participants' performances revealed that, unlike humans, animals seemed to rely on a perceptual measure of entropy rather than on a genuine representation of same and different (Young & Wasserman, 1997; Fagot et al., 2001). On the contrary, in the present study, infants cannot categorize the different triplet structures based on either the identity of the constitutive syllables, as they are constantly changed, or a measure of entropy, as all triplets contain exactly two different syllables, one being repeated. It is particularly the case when AAB and ABB structures are compared as the same number of immediate repetitions and differences are present in both structures. Besides, such protocols rely on either extensive training of the animals or very specific learning mechanism in the case of avian filial imprinting (Martinho & Kacelnik, 2016), while infants only need a few habituation trials to capture the abstract pattern, and generalize their knowledge to new stimuli. Recent experiments even suggest that infants can readily derive abstract knowledge from very few exemplars (Gerken & Knight, 2015; Gerken et al., 2015).

Sequence learning abilities in non-human animals has also been investigated across various species, with mixed results. Birds (Gentner et al., 2006; Van Heijningen et al., 2009; Chen et al., 2015; Spierings & ten Cate, 2016), rats (Murphy et al., 2008), and primates (Hauser & Glynn, 2009; Fitch & Hauser, 2004) can be trained to discriminate two structures which are related to differences in reinforcement. However, in many experiments, generalization of the response to novel instantiations of the structures suggests that non-human animals rather solve this discrimination task based on the perceptual features of the sequences, instead of detecting the abstract structure governing the sequence. Although they appear to possess abstract abilities, it seems like animals represent the input at a completely different level to what is observed in infants, or adults. In a comparative study on the neural representations of abstract sequences in humans adults and non-human primates, Wang et al. (L. Wang et al., 2015) report that the human brain seems to benefit from brain regions where the different abstract features of an auditory sequence are integrated into a single neural representation, whereas in the macaque brain these abstract features are encoded in distinct areas. Interestingly, these integrative regions of the human brain are also involved in various symbolic computations such as language processing, musical syntax processing or mathematical calculation. Abstract representations in humans therefore appear closely connected to language and symbolic reasoning, and the infant brain architecture might already allow an easy access to the symbolic system, supporting the discovery of the complex structure of language. Although we cannot evidence symbolic computations with the present results, we nonetheless evidence unified representations of the abstract structures, as robust mental variables. Infants' ability to associate each abstract representation with an arbitrary label suggests that they possess the foundations for the emergence of symbolic reference.

Our results further shed light on the neural architecture of the developing brain supporting these fast and complex computations. The reported modulations of early sensory activity in response to congruent or incongruent pairings evidence sophisticated structural organization, with top-down connections from higher-level areas. As suggested by previous research (Emberson et al., 2015; Kouider et al., 2015), infants already seem to benefit from efficient long-distance connections mediating either surprise responses from the top of the processing hierarchy or expectation-based activations. Such neural strategies might readily allow efficient cross-modal learning.

CHAPTER 5

DISCUSSION AND PERSPECTIVES

Throughout the experimental work reported in this thesis, we tackled a number of technical and theoretical issues in the field of cognitive development. In Chapter 1, we complemented the current methodological framework of infant brain imaging, assessing the reliability of the 10/20 electrode placement system, and proposing an anatomical infant brain atlas. In Chapter 3 and Chapter 4, we evidenced amazingly fast and powerful abilities for encoding the features and structure of short syllable sequences. We discuss here the methodological and conceptual contributions of this thesis.

5.1 Relevance of developmental EEG

It should be stressed that the early cognitive abilities reported in this thesis could be evidenced thanks to the use of EEG recordings combined with cognitive experimental paradigms. While, historically, research in cognitive development has been restricted to the realm of observable behavior, electrophysiology and the recent proliferation of refined signal processing methods has allowed to inspect directly the **living infant brain** at work. EEG measurements provide two main advantages that we were able to exploit to probe early cognitive abilities.

5.1.1 Innovative paradigms

First, the experimental procedure allows the development of innovative experimental paradigms. Indeed, the electroencephalogram is a non-invasive and painless recording of cerebral activity, particularly well suited for studying very young infants for whom the demands of other imaging techniques may be too great. The procedure makes no demands on verbal or motor abilities, it is relatively less perturbed by brief head movements and it can be set up within a short time period, making fewer demands on the attention span of infants. The responses of interest are then extracted directly from brain activity, often contrasting two different experimental conditions, and inspecting response differences. Unlike classical behavioral methods for which the understanding of a cognitive process depends on the measurement of a final behavioral outcome, imaging techniques provide a chance to observe directly the **brain processes** at play during a given cognitive task, bypassing slow head turn or eye movement or sucking responses. The trials of the experimental design can therefore be shorten as no behavioral outcome is expected from the participant, and the number of trials can in turn be increased. More interestingly, the brain processes can be inspected in great details online, during the learning phase itself. That is how, in Chapter 3 using highly time-controlled stimulation material, we could evidence brain correlates for segmenting a continuous speech stream in 8 month-old infants. We could even observe a gradual increase of activity related to word extraction throughout the familiarization stream. These fast learning responses are unfortunately out of the reach of behavioral methods, and only electrophysiology allows a detailed inspection of **learning correlates**.

Besides, the classical habituation/dishabituation paradigms widely used in behavioral experiments only provide a limited window onto infant cognitive processes as they only evidence discrimination responses. On the contrary, since EEG responses are not restricted to a small array of behavioral outcomes, the experimental designs can **move beyond** habituation/dishabituation processes and capture more than discrimination responses. That's why in Chapter 4, we were able to implicitly teach and test the acquisition of a systematic association between several sequence structures and the forthcoming sensory labels. We therefore indirectly tested participants' abilities to represent and manipulate simultaneously multiple sequence structures. Such a paradigm would have been much more challenging to set up based only on behavioral measurements. Using anticipatory eye movements and similar triplet structures, a few studies have attempted to test infants' abilities to maintain and manipulate several mental representations (Kovacs & Mehler, 2009; Kovacs, 2014), but the results revealed that most infants managed to perform the task for only one of the taught mental representations. Together with our results, it suggests that the cognitive efforts for producing the expected behavioral responses was too demanding. We could circumvent these constraints using electrophysiology.

5.1.2 Innovative processing techniques

Second, the multiple dimensions (e.g. time, sensors, frequencies…) of the EEG signal offer almost **limitless possibilities** for analyzing cerebral responses. Depending on the working hypotheses, innovative methods can be implemented, in order to evidence a given cognitive or brain mechanism. ERPs are typically used to evidence different neural responses underlying different experimental conditions, but several measures other than voltage amplitudes can be extracted and interpreted from the EEG signal. Depending on the experimental constraints, these methods can prove to be more sensitive than typical averaged voltages. In the present thesis we particularly relied on the frequency content of the cerebral activity. Using **frequency tagging** approaches in both Chapter 3 and Chapter 4, we chose to constrain the oscillatory cerebral activity at a precise stimulation frequency, in order to retrieve strong steady-state responses. This technique is particularly suited for studying the infant brain since it is relatively **robust** to artifacts, hence providing high signal-to-noise ratios. Besides, compared to ERP responses for which the timings of the expected responses are often unknown, this technique is **objective** and **predictive**: the response appears exactly at the stimulation frequency defined by the experimenter. We were indeed able to extract clear entrained brain responses both in auditory regions (Chapter 3) and in visual regions (Chapter 4). Moreover the high sensitivity of this measure allowed us to evidence subtle modulations of the oscillatory power: in Chapter 3 we could detect a progressive power enhancement at word frequency during the familiarization stream, and in Chapter 4, we observed a power decrease at stimulation frequency in response to incongruent pairings.

Time-frequency signal decompositions additionally allow the inspection of oscillatory spontaneous or evoked activity over time across a broader frequency spectrum. In the past decades, adult electrophysiology revealed that several cognitive processes involved (non-entrained) cortical rhythms in various frequency bands. Experimental and modeling neurophysiologists have put a lot of efforts in elucidating the underlying cellular and circuit mechanisms, stressing the fundamental role of synaptic inhibition for rhythmogenesis in the mature brain. However, the infant brain circuitry remains immature for quite an extended period, especially for inhibitory interactions: parvalbumin inhibitory interneurons develop very slowly until at least 10 months of post-term age (Honig et al., 1996), and GABA, the principal inhibitory transmitter in the adult brain, acts as an excitatory transmitter in early postnatal stage (Ben-Ari et al., 1997). Therefore, it is likely that the infant brain displays somewhat different cortical dynamics, and interpretations from adult literature should be taken with great **caution**. Nevertheless, promising perspectives of research are open to understand the **ontogeny of cortical rhythms**, and its functional implications for temporal coordination between remote areas of the brain or sensory integration for example. In both Chapter 3 and Chapter 4, we were able to evidence modulations of beta or gamma band activity depending on the experimental conditions, indicating that the oscillatory content of the signal already supports higher-level cognitive functions.

Several other options such as connectivity measures, decoding techniques, brain-behavior correlations or source modelling still remain to be explored to probe both cognitive processes and brain mechanisms. In the case of source modelling, it would be interesting to use the infant atlas presented in Chapter 2 and reconstruct cortical activity from sensor activity in order to localize the recorded entrained or evoked neural responses. However, at the moment, there is no satisfactory infant head model for solving the inverse problem.

5.1.3 Limitations

Finally, we cannot discuss developmental EEG without addressing the limiting practical issues we have been facing throughout this thesis. When working with young subjects, participants' tolerance and task commitment are highly restricted, and despite the tricks that the investigator can set up, the attrition rate for infant electrophysiology remains higher than for behavioral or adult research. The number of trials is therefore much weaker than what can be recorded in adults. Moreover, the immaturity of some brain processes can induce additional variability in the time-locking of the electrical components, therefore decreasing the voltage amplitude of the averaged evoked responses. Besides, we evidenced in Chapter 2 a substantial inter-subject variability in the correlations between sensors and the underlying anatomical structures. This variability, although modest in relation to the diffusion of the electrical components, still induces additional variability in sensors from one subject to the next. These various issues result in **decreased statistical power**.

Regarding statistical analysis of EEG data, another main limitation actually comes as a drawback from the extreme richness of the EEG recordings. Indeed, we have to deal with the **multiple comparisons problem**. This problem originates from the fact that the effect of interest is evaluated at an extremely large number of sensors, (frequency,) time points. Due to this large number of statistical comparisons, it is not possible to control for the false positive error rate when performing simple repeated pairwise statistical comparisons. The obtained probabilities must be somehow corrected. The classical parametric Bonferroni correction is a very conservative option when the number of samples is large, as it does not take into account the spatio-temporal redundancy of the signal. In the present thesis, we chose to use a non-parametric statistical testing procedure (Maris & Oostenveld, 2007), allowing a **straightforward** extraction of the clusters of electrodes and time-windows of significant effects. However, although we were able to evidence significant differences, this procedure also appeared to be strongly affected by the

low signal-to-noise ratio of ERP responses. Because of the slow, long-lasting dynamics of infant ERPs we had to increase the sensitivity of the test by setting the alpha threshold to 0.1 (which does not affect the false alarm rate of the test). Still, the reported effects were not particularly strong, despite the rather high numbers of participants in the different studies. Alternatively, the multi-dimensionality of the signal can be reduced by preselecting temporal windows or clusters of sensors, based on the working hypotheses, or the existing literature. However, this solution is quite limited when it comes to exploring distributed electrical components, or **new** and **undocumented** effects. Statistical analyses therefore remain challenging for developmental neuroscientists, and additional efforts should be allocated to set up a more systematic methodological framework for statistics in the field of infant brain imaging.

Despite these limitations, we aimed at adopting a **rigorous approach**, and the effects reported in this thesis were obtained using rather conservative procedures. In the future we intend to explore more sensitive approaches such as decoding (King & Dehaene, 2014) or spatial filtering (Schurger et al., 2013), and investigate whether these methods can provide additional insights.

5.2 Sequence encoding: a hierarchy of cognitive processes

Beside these methodological considerations, the experimental studies reported in this thesis addressed a series of increasingly complex cognitive processes which might play a fundamental role in the acquisition of the hidden structure of language. We propose that the encoding of linguistic structures can cover **four separable mental processes**.

5.2.1 Surface computations

First, as reviewed in Chapter 1 of this thesis, infants are born with **innate skills** for processing the rapid, sequential surface attributes of the speech stream. The cortical circuitry and the functional architecture of the human infant brain set a number of constraints on early perceptual abilities. Regarding basic acoustic processing of the environment, both imaging and behavioral studies have demonstrated over the past decades initial **perceptual biases** for preferentially attending to speech (Mehler et al., 1988; Peña et al., 2003). Similarly, a **phoneme-sensitive network** appears to be functional at the very onset of cortical organization, allowing the perception of subtle acoustic differences between the phonemes of speech

(Mahmoudzadeh et al., 2013). Interestingly, the functional organization of the pSTG might initially support the invariant representations of speech sounds underlying early categorical perception (Dehaene-Lambertz & Baillet, 1998). These initial biases for perceiving or neglecting acoustic changes are crucial for encoding the speech signal as a temporal sequence of linguistic units.

Similarly, infants readily seem gifted with a **toolkit of mechanisms** for processing the structural properties of speech sequences. Several behavioral and some imaging studies have evidenced very early sensitivity to the **statistical dependencies** between **adjacent** units of a sequence (Saffran et al., 1996; Teinonen et al., 2009). Additional studies have further documented implicit statistical learning abilities between **non-adjacent** units (Gómez, 2002; Gómez & Maye, 2005; Friederici et al., 2011; Marchetto & Bonatti, 2014), and in Chapter 3 of this thesis, we complemented this work by providing electrophysiological evidence that as soon as 8 months of age, infant can track online dependencies between non-adjacent units of a continuous speech stream. Beside these statistical learning abilities, human infants show some perceptual biases allowing the detection of some particular features of a given sequence. For example, from birth on, infants are particularly sensitive to **repetitions** as they can discriminate between sequences comprising adjacent repetitions, from sequences without any adjacent repetitions (Gervain, Macagno, et al., 2008; Gervain et al., 2012). We capitalized on this sensitivity in Chapter 4. Another perceptual bias which plays an important role in sequence processing relates to the enhanced perceptual saliency of sequence-initial and sequence-final positions. Sequence encoding in both infants (Ferry et al., 2016) and adults (Endress et al., 2005) is biased towards the **edges** of linguistic segmental or supra-segmental units.

These computational abilities and perceptual biases are particularly robust, they develop very early in life, and they can operate over different perceptual modalities. Thus, we can reasonably speculate that they reflect basic processing mechanisms that derive from the **innate constraints** of the perceptual systems. However, these early computational abilities and perceptual biases are certainly not sufficient for acquiring the underlying structure of language, as animal studies revealed that we share these mechanisms with **other animals**. Categorical perception has been evidenced in chinchillas (Kuhl & Miller, 1978), monkeys (Morse & Snowdon, 1975) or quails (Kluender et al., 1987). Non-human primates have been shown to possess powerful statistical learning abilities (Hauser et al., 2001; Newport et al., 2004), and repetition detection was demonstrated in a variety of animal species (Murphy et al., 2008; Giurfa et al., 2001). Although these early abilities for processing online the acoustic or structural properties of the speech input play an important role in natural language learning, these competences did not develop throughout evolution specifically for the language faculty. Conversely, the languages of the world rather capitalized on perceptual discontinuities for deriving the phonemic code, or on edge biases by placing linguistic regularities at the edges of morphological constituents.

5.2.2 Structural generalization

Second, using these powerful perceptual and computational abilities, infants crucially need to integrate the structural properties of the input and **project generalizations** in order to process novel linguistic elements that have never been experienced. Generalization is typically evidenced using behavioral habituation /dishabituation procedures in which the infant is first familiarized with a given structure implemented over a specific set of elements, and then presented with familiar and novel structures built out of a novel set of elements. Infants have generalized if they display different behavioral responses to familiar and novel sequence structures. From this perspective, generalization corresponds to a learning mechanism in which infants habituate to the abstract features of the input, allowing them to discriminate between the learned structure and a novel one. Besides, this generalization step can be **dissociated** from basic sensitivity to statistical regularities described above. Indeed, in a statistical learning task, Peña et al. (2002) demonstrated that adult participants could detect long-distance statistical dependencies, and still fail to recognize these regularities when implemented with different elements. Therefore, although participants can detect and process the statistical contingencies contained in a speech stream, another type of computation is needed to trigger the emergence of structural generalization.

The precise mechanisms that govern these abstraction abilities are still unclear, but a growing field of developmental research is exploring a “Bayesian” or “probabilistic” approach to account for generalization. **Hierarchical Bayesian models** (Kemp et al., 2007; Tenenbaum et al., 2011) address this ability to infer abstract knowledge by positing the existence of increasingly abstract levels of representations, in which the higher levels constrain lower-level learning. In such models, each level of the hierarchy represents a hypothesis space for explaining the observable data. Given observations at the lowest level, statistical inference can be used to update the likelihood of hypotheses at the higher levels, hence capturing the notion of generalization, or **“overhypothesis”** (R. M. Martin & Goodman, 1960). Thanks to this hierarchical organization, knowledge can be acquired at levels quite remote from the experienced data. Hierarchical Bayesian models thus provide an interesting framework for studying knowledge acquisition, as they make direct predictions about what is generalized, given a set of observations. Besides, it seems to account for several behavioral outcomes in infant learning paradigms (Frank & Tenenbaum, 2011; Dewar & Xu, 2010). As a future direction, we intend to explore whether the infant brain directly implements such Bayesian learning mechanisms. In this Bayesian framework, we speculate that, as infants experience the successive triplet words in Chapter 4, they should progressively overhypothesize to converge towards the abstract structures AAB, ABB or ABA. The successive syllables of a given word should in turn become more and more predictable as the hidden structure is progressively discovered. Since the amplitude of the neural responses correlates with the predictability of the input (Winkler, 2007), we expect the evoked response to each of the experienced syllables to be modulated by the progressive refinements of higher-level hypothesis space, with a gradual decrease of its amplitude throughout the experiment.

The Bayesian framework therefore provides a new perspective onto knowledge development as it states that behaviors which could be considered as innate given their extreme precocity, might actually just be acquired extremely fast thanks to this hierarchical organization of knowledge and the **"blessing of abstraction"**. However, this approach does not address how the hypothesis space itself emerges at each level of the hierarchy. Besides, given the "appropriate" set of hypotheses, this model can be fit virtually to any data. Any account of induction must rely on some **initial knowledge**, and it is still unclear how these initial assumption can be selected, and whether Bayesian models of learning can be fully viable based on minimal assumptions.

5.2.3 Recoding

Third, human abilities for sequence encoding entail more than a mere discrimination between a familiar and a novel sequence structure. Humans easily move beyond habituation mechanisms, as they can **recode** the linguistic input into a **mental schema** that captures the relationships between the successive stimuli or categories of stimuli. Although generalization, as evidenced by dishabituation responses to a novel structure, reveals some form of abstraction in preverbal infants as discussed previously, it does not address the encoding format of the abstract structure. Thus, generalization does not necessarily imply a unified representation of the "algebraic pattern" governing the input sequence. As discussed by McClelland & Plaut (McClelland & Plaut, 1999), it could simply reflect passive habituation to some of its abstract features. In Chapter 4 of the present thesis, we proposed an experimental paradigm that, we believe, is better suited for addressing this question. We posit that if infants can encode the abstract structures into unified schema, these generic representations should be available to the infant's mind for, first, categorizing between different structures, and second, for chaining successive mental operations involving these mental representations. The results presented in Chapter 4 demonstrate that by 5 months of age infants can already form and manipulate such mental representations for minimal repetition-based linguistic structures such as AAB, ABB or ABA.

This recoding process allows to **reduce the complexity** of the world into a discrete space of mental variables that can further be manipulated, combined or transformed. In Chapter 4, we were able to evidence early correlates of such encoding, setting the foundations for the upcoming development of rich combinatorial systems of knowledge such as language, mathematics or music. Although we could not precisely assess the encoding format of the abstract structures from the experimental results reported in Chapter 4, we were able to evidence that all three structures elicited different neural representations. These results imply that infants were able to **combine** different features of the input triplets (such as information related to the presence of adjacent repetitions and information related to the ordinal position of the repetition) in order to form generic mental representations flexible enough to be maintained throughout the experiment and to be used to differentiate each structure from the two others. More interestingly, the generated representations could be used as input for a pairing task, demonstrating that preverbal infant can already **chain multiple complex mental operations** involving abstract representations.

The ability to flexibly compress the complexity of the sensory input into generic mental representations might explain the **specificity** of the human mind among the animal kingdom. Several studies have evidenced impressive abstraction abilities in a variety of animal species, from bees (Giurfa et al., 2001) to apes (Oden et al., 1988). However, very few report convincing evidence for the encoding of complex patterns which involve a combination of different abstract features (but see (Spierings & ten Cate, 2016)). In the case of structure learning in non-human animals, several results could be accounted for by perceptual strategies (Corballis, 2009). Besides, animal studies involve extensive training procedures while human adults and infants can capture these patterns within only a few trials. Thus, the human brain appears to possess a particularly efficient, if not unique, competence for extracting and combining abstract sequence information into a unified representation (L. Wang et al., 2015), therefore expanding the complexity spectrum of structures that it can capture.

5.2.4 Entering a symbolic system

Fourth, the hallmark of human cognition is the ability to share, learn and use a **rich communicative system of symbols**. Language acquisition therefore consists in discovering the **mapping** between the recoded speech sequences onto conventionally defined linguistic symbols and rules. The combinatorial power of the brain is then boosted by the discovery of symbolic referents, not only by sharing with conspecifics, but also by allowing to operate directly on the symbols using the syntax of the referential system. Moreover, during language acquisition, verbal symbols can actually mediate the access to the mental representations. A series of word learning experiments have demonstrated that communicative labels serve as **"invitations"** to form abstract categories of objects, directing infants' attention to commonalities among the presented objects (Waxman & Braun, 2005; Waxman & Markow, 1995; Fulkerson & Waxman, 2007; Ferry et al., 2010; Ferguson & Lew-Williams, 2016). Symbolic reference therefore starts with a repeated perceptual association between an arbitrary signifier and the signified object or concept (Nieder, 2009). This pairing then **switches** from such indexical stage to a more powerful symbolic stage, as a **bidirectional** relation between the signifier and the signified is recognized. Indeed, the symbol activates a representation of the object, and the object activates a representation of a symbol (Ngon & Peperkamp, 2016). Csibra & Gergely (2006) suggest that such referential assignment could be mediated by ostensive communication cues. Still, the mechanisms underlying this referential switch remain unknown.

Throughout the experiments reported in Chapter 4, we demonstrated that infants were able to pair mental representations of the abstract structures to a visual or verbal label. The acquisition of verbal or written symbols for objects and concepts is primarily rooted in this ability to **pair** any abstract mental representation to an arbitrary auditory or visual element. As a future perspective, we intend to explore the nature of the generated association. Since infants manage to capture the referential nature of speech very early on, we can speculate that the generated structure-item association is not limited to a temporal dependency between the two elements. Infants might have already access to rudimentary symbolic reference, and they may interpret that the sensory item actually **represents** its associated linguistic structure.

In this perspective, we designed a **follow-up experiment** in order to test whether the structure-item association entails more than a pure temporal dependency between the two elements. Similarly to Experiment 1, infants are presented with trisyllabic nonce words instantiating one of two simple repetition-based structures, consistently followed by a visual item. Each structure is associated to a specific visual item. After the training phase, we introduce 30% of test trials in which the temporal dependency is **reversed**: first, the cartoon picture is presented followed by the auditory trisyllabic word (Figure 5.1). Crucially, item-structure associa-

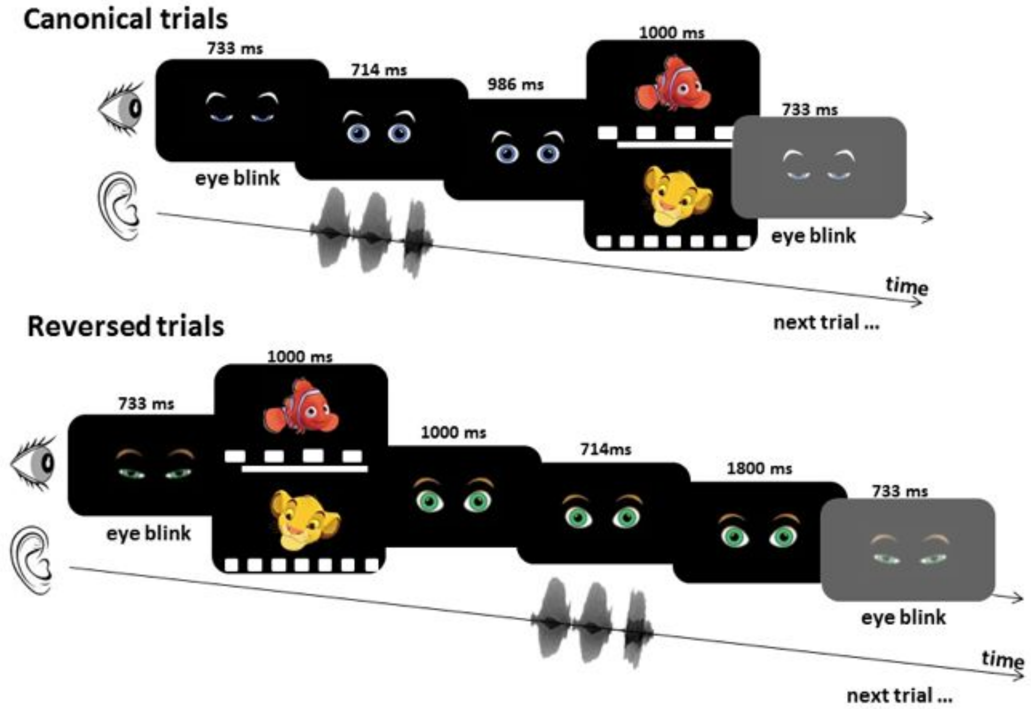


Figure 5.1: **Experimental paradigm of the follow-up study.** The canonical trials are similar to the ones presented in Experiment 1 Chapter 4 (4.2), with only congruent pairings. In the reversed trials, the visual item is presented first, followed by the triplet structure, either congruent (50%) or incongruent (50%) reversed pairings.

tions in these reversed trials were completely balanced, so that infants could not learn any statistical dependency from the image to the auditory structure. The remaining 70% of trials are presented in the canonical order, with only congruent pairings, to reinforce the structure-item dependency. Building on Experiment 1 and Experiment 2, we expect infants to learn the structure-item association from the canonical trials, and we hypothesize that if participants appreciate the cartoon pictures as symbols for the abstract structures, structure-item associations should be bidirectional. Congruent and incongruent pairings in reversed trials should therefore elicit distinct neural responses.

5.3 The neural mechanisms

Eventually, the findings reported in this thesis complement our understanding of the developing brain.

In both of these studies we successfully used fast rhythmic stimulation to probe non-trivial cognitive processes. The recorded evoked steady state responses demonstrate that, although immature, the infant brain already allows **precise temporal coding** of the auditory and visual streams of external stimulation. In the first EEG experiment, the rhythmic familiarization stream elicited strong auditory evoked steady state responses, and cerebral oscillations were measured as correlates for continuous speech segmentation. In particular, oscillations at syllabic and word frequencies appeared to index the extraction of the perceptual units of the familiarization stream. In adults, oscillatory mechanisms in the auditory cortex have already been hypothesized to account for parsing the speech signal into representational units relevant for language processing (Giraud & Poeppel, 2012; Poeppel, 2003). The present results suggest that such oscillatory models for speech perception could already be viable in preverbal infants, raising of course several conceptual questions about the biological constrains on the auditory cortex circuitry to allow chunking at the appropriate temporal granularities. However, further investigations should be held to explore the pattern of oscillatory responses in the immature brain, and assess their functional role in speech processing.

Moreover, the recorded steady state responses, low-level entrained activity by definition, also informed us on higher-level processes: progressive extraction of words in Chapter 3 and detection of incongruent structure-item parings in Chapter 4. In the case of speech segmentation (Chapter 3), the progressive enhancement of oscillatory power throughout the 2-minutes familiarization stream suggests that auditory regions entrained by rhythmic stimulation receive input from the neural machinery performing statistical computations over the surface features of the stimulation. Statistical computations could even possibly be performed locally in sensory areas. On the contrary, in Chapter 4, it is highly unlikely that the complex processes relative to the detection of incongruent structure-item pairings occurred locally in early visual areas, as they involve abstract computations over linguistic input and cross-modal integration. The recorded modulations of early visual responses therefore indicate some **top-down connections** from remote higher-level brain regions down to early sensory areas. Although immature, the infant brain already seems to possess a wealth of functional long-range fiber tracts (Dubois et al., 2015), allowing information transfer between processing modules for abstract knowledge and early sensory areas. However, the functional role of these long-distance connections down to early sensory regions remains to be understood.

Finally, in addition to these low-level sensory responses, the typical ERP analyses revealed that the experimental conditions also affected late components in both of the reported studies. These responses, recorded around 1 second after stimulus onset are classically referred to as **Late Slow Waves** (de Haan et al., 2003; Csibra et al., 2008). The LSW is involved in a variety of experimental paradigms, suggesting that it reflects generic amodal processes, associated with higher-order computations such as attention, working memory or conscious perception. It has been proposed as an equivalent of the adult P300 (Kouider et al., 2013), supported by a large-scale distributed network of long-distance cortico-cortical connections between sensory regions and associative fronto-parietal regions. Although present from birth on (Dubois et al., 2015), these long fiber tracts remain weakly myelinated during the first year of life (Yakovlev & Lecours, 1967), explaining the reported delayed latencies. Our results are compatible with this interpretation, as the reported LSWs were recorded in response to higher-order computations: rule violation (Chapter 3), or incongruent structure-item pairings (Chapter 4), likely involving associative areas for memory storage and abstraction abilities. The analogy with the adult P300 additionally raises fundamental questions about infant cognition, and the interplay between language learning and conscious processing.

To conclude, thanks to the temporal properties of high-density EEG, we were able to describe complex neural processing in the infant brain enlarging our perspective on early competencies. The possibilities offered by non-invasive brain imaging open a new domain to understand the computational properties of the infant brain.

REFERENCES

Abrams, D. A., Nicol, T., Zecker, S. G., & Kraus, N. (2006). Auditory brainstem timing predicts cerebral asymmetry for speech. *J Neurosci*, *26*(43), 11131–11137.

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001a). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences*, *98*(23), 13367–13372.

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, a., Mahncke, H., & Merzenich, M. M. (2001b, nov). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(23), 13367–72.

doi: 10.1073/pnas.201400998

Akiyama, L. F., Richards, T. R., Imada, T., Dager, S. R., Wroblewski, L., & Kuhl, P. K. (2013). Age-Specific Average Head Template for Typically Developing 6-Month-Old Infants. *PLoS One*, *8*(9), e73821.

Altaye, M., Holland, S. K., Wilke, M., & Gaser, C. (2008). Infant brain probability templates for MRI segmentation and normalization. *NeuroImage*, *43*(4), 721–730.

doi: http://dx.doi.org/10.1016/j.neuroimage.2008.07.060

Altmann, G. T. M., & Dienes, Z. (1999). Rule learning by seven-month-old infants and neural networks. *Science*, *284*(5416), 875.

Armitage, S. E., Baldwin, B. A., & Vince, M. A. (1980). The fetal sound environment of sheep. *Science*, *208*(4448), 1173–1174.

Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H., & Thierry, G. (2015). Sound symbolism scaffolds language development in preverbal infants. *Cortex*, *63*, 196–205.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Development*, *52*(4), 1135.

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996). Models of word segmentation in fluent maternal speech to infants. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 117–134.

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of experimental child psychology*, *64*(1), 3–26. doi: 10.1006/jecp.1996.2332

Barkovich, A. J. (2000). *Concepts of myelin and myelination in neuroradiology* (Vol. 21) (No. 6). doi: 10.1016/S0006-3223(01)01323-3

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.

doi: 10.1016/j.cognition.2014.03.013

Beauchamp, M. S., Beurlot, M. R., Fava, E., Nath, A. R., Parikh, N. A., Saad, Z. S., ... Oghalai, J. S. (2011). The developmental trajectory of brain-scalp distance from birth through childhood: implications for functional neuroimaging. *PLoS One*, *6*(9), e24981.

doi: 10.1371/journal.pone.0024981PONE-D-11-12012[pii]

Ben-Ari, Y., Khazipov, R., Leinekugel, X., Caillard, O., & Gaiarsa, J. L. (1997). *GABA(A), NMDA and AMPA receptors: A developmentally regulated 'menage a trois'* (Vol. 20) (No. 11). doi: 10.1016/S0166-2236(97)01147-8

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–3258.

Bergelson, E., & Swingley, D. (2013). The acquisition of abstract words by young infants. *Cognition*, *127*(3), 391–397.

Bernal, S., Dehaene-Lambertz, G., Millotte, S., & Christophe, A. (2010). Two-year-olds compute syntactic structure on-line. *Dev Sci*, *13*(1), 69–76.

Bernal, S., Lidz, J., Millotte, S. S., & Christophe, A. (2007). Syntax constrains the acquisition of verb meaning. *Language Learning and Development*, *3*(4), 325–341.

Bertoncini, J., Bijeljac-Babic, R., Blumstein, S. E., & Mehler, J. (1987). Discrimination in neonates of very short CVs. *The Journal of the Acoustical Society of America*, *82*(1), 31.

doi: 10.1121/1.395570

Bertoncini, J., & Mehler, J. (1981). Syllables as units in infant speech perception. *Infant Behavior and Development*, *4*, 247–260.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of experimental psychology: human perception and performance*, *14*(3), 345.

Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., ... Murphy, D. G. M. (2011). Early specialization for voice and emotion processing in the infant brain. *Current Biology*, *21*(14), 1220–1224.

doi: 10.1016/j.cub.2011.06.009

Blume, W. T., Buza, R. C., & Okazaki, H. (1974). Anatomic correlates of the ten-twenty electrode placemnt system in infants. *Electroencephalography and Clinical Neurolophysiology*, *36*, 303–307.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me familiar names help launch babies into speech-stream segmentation. *Psychological Science*, *16*(4), 298–304.

Bosch, L., & Sebastian-Galles, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *2*(1), 29–49.

Bosch, L., & Sebastian-Galles, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and speech*, *46*(2-3), 217–243.

Brannon, E. M. (2002). The development of ordinal numerical knowledge in infancy. *Cognition*, *83*(3), 223–240.

Brusini, P., Dehaene-Lambertz, G., Dutat, M., Goffinet, F., & Christophe, A. (2016). ERP evidence for on-line syntactic computations in 2-year-olds. *Developmental Cognitive Neuroscience*, *19*, 164–173.

Brusini, P., Dehaene-Lambertz, G., van Heugten, M., de Carvalho, A., Goffinet, F., Fiévet, A.-C., & Christophe, A. (2016). Ambiguous function words do not prevent 18-month-olds from building accurate syntactic category expectations: An ERP study. *Neuropsychologia*. doi: 10.1016/j.neuropsychologia.2016.08.015

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.

Cauvet, E., Limissuri, R., Millotte, S., Skoruppa, K., Cabrol, D., & Christophe, A. (2014). Function words constrain on-line recognition of verbs and nouns in French 18-month-olds. *Language Learning and Development*, *10*(1), 1–18.

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature neuroscience*, *13*(11), 1428–1432.

Chen, J., van Rossum, D., & Ten Cate, C. (2015). Artificial grammar learning in zebra finches and human adults: XYX versus XXY. *Animal cognition*, *18*(1), 151–164.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature neuroscience*, *1*(5), 351–353.

Chi, J. G., Dooling, E. C., & Gilles, F. H. (1977). Gyral development of the human brain. *Annals of Neurology*, *1*(1), 86–93.

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.

Chomsky, N. (1980). Rules and representations. *Behavioral and brain sciences*, *3*(01), 1–15.

Christiansen, M. H., & Curtin, S. (1999). Transfer of learning: rule acquisition or statistical learning? *Trends Cogn Sci*, *3*, 289–290.

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

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.

Chu, C. J., Leahy, J., Pathmanathan, J., Kramer, M. A., & Cash, S. S. (2014). The maturation of cortical sleep rhythms and networks over early development. *Clinical Neurophysiology*, *125*(7), 1360–1370. doi: 10.1016/j.clinph.2013.11.028

Cointepas, Y., Mangin, J. F., Garnero, L., Poline, J. B., & Benali, H. (2001). BrainVISA: Software Platform for Analysis of Multi-modality Brain Data. *NeuroImage*, *13*(6), S98.

Colombo, J., & Bundy, R. S. (1981). A method for the measurement of infant auditory selectivity. *Infant Behavior and Development*, *4*, 219–223.

Corballis, M. C. (2009). Do rats learn rules? *Animal Behaviour*, *78*(4). doi: 10.1016/j.anbehav.2009.05.001

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. *Processes of Change in Brain and Cognitive Development*, 249–274. doi: 10.1.1.103.4994

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends Cogn Sci*, *13*(4), 148–153.

Csibra, G., Kushnerenko, E., & Grossmann, T. (2008). Electrophysiological Methods in Studying Infant Cognitive Development. In C. A. Nelson (Ed.), *Handbook of developmental cognitive neuroscience, 2nd edition* (pp. 247–262). Cambridge, Massachusetts: MIT Press.

Dawson, C., & Gerken, L. (2009). From domain-generality to domain-sensitivity: 4-month-olds learn an abstract repetition rule in music that 7-month-olds do not. *Cognition*, *111*(3), 378–382.

Dawson, C. R., & Gerken, L. (2012). Can rational models be good accounts of developmental change? The case of language development at two time scales. *Rational constructivism in cognitive development*, 95–124.

De Hevia, M. D., & Spelke, E. S. (2010). Number-space mapping in human infants. *Psychological Science*, *21*(5), 653–660.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176.

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, *9*(2), 133–150.

de Haan, M., Johnson, M. H., & Halit, H. (2003). Development of face-sensitive event-related potentials during infancy: a review. *International Journal of Psychophysiology*, *51*(1), 45–58.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport*, *8*(4), 919–924.

Dehaene-Lambertz, G., & Baillet, S. (1998). A phonological representation in the infant brain. *Neuroreport*, *9*(8), 1885–1888.

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.

doi: 10.1126/science.1077066

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Meriaux, S., Roche, A., Sigman, M., & Dehaene, S. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proc Natl Acad Sci U S A*, *103*(38), 14240–14245.

doi: 0606302103[pii]10.1073/pnas.0606302103

Dehaene-Lambertz, G., Montavont, A., Jobert, A., Allirol, L., Dubois, J., Hertz-Pannier, L., & Dehaene, S. (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., & Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport*, *12*(14), 3155–3158.

de Heering, A., & Rossion, B. (2015). Rapid categorization of natural face images in the infant right hemisphere. *Elife*, *4*, e06564.

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.

Dewar, K. M., & Xu, F. (2010). Induction, overhypothesis, and the origin of abstract knowledge: Evidence from 9-month-old infants. *Psychological Science*, *21*, 1871–1877. doi: 10.1177/0956797610388810

Dubois, J., Benders, M., Cachia, A., Lazeyras, F., Ha-Vinh Leuchter, R., Sizonenko, S. V., . . . Huppi, P. S. (2008). Mapping the early cortical folding process in the preterm newborn brain. *Cerebral Cortex*, *18*(6), 1444–1454.

Dubois, J., Benders, M., Lazeyras, F., Borradori-Tolsa, C., Leuchter, R. H. V., Mangin, J. F., & Hüppi, P. S. (2010). Structural asymmetries of perisylvian regions in the preterm newborn. *NeuroImage*, *52*(1), 32–42.

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. doi: 10.1093/cercor/bhn097

Dubois, J., Poupon, C., Thirion, B., Simonnet, H., Kulikova, S., Leroy, F., . . . Dehaene-Lambertz, G. (2015). Exploring the early organization and maturation of linguistic pathways in the human infant brain. *Cerebral Cortex*, bhv082.

Dutoit, T. (1997). *An introduction to text-to-speech synthesis* (Vol. 3). Springer.

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & psychophysics*, *16*(3), 513–521.

Eimas, P. D., Miller, J. L., & Jusczyk, P. W. (1987). On infant speech perception and the acquisition of language.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–306.

Emberson, L. L., Richards, J. E., & Aslin, R. N. (2015). Top-down modulation in the infant brain: Learning-induced expectations rapidly affect the sensory cortex at 6 months. *Proceedings of the National Academy of Sciences*, *112*(31), 9585–9590.

Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General*, *134*(3), 406–419. doi: 10.1037/0096-3445.134.3.406

Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, *18*(3), 254–260.

Evans, A. C., Collins, D. L., Neelin, P., MacDonald, D., Kamber, M., & Marrett, T. S. (1994). Three-dimensional correlative imaging: applications in human brain mapping. *Functional neuroimaging: technical foundations*, 145–162.

Fagot, J., Wasserman, E. A., & Young, M. E. (2001). Discriminating the relation between relations: the role of entropy in abstract conceptualization by baboons (Papio papio) and humans (Homo sapiens). *Journal of Experimental Psychology: Animal Behavior Processes*, *27*(4), 316.

Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences*, *105*(29), 9926–9930.

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.

Ferguson, B., & Lew-Williams, C. (2016). Communicative signals support abstract rule learning by 7-month-old infants. *Scientific reports*, *6*.

Ferguson, B., & Waxman, S. R. (2015). Visual abstract rule learning by 3-and 4-month-old infants. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, 692–697.

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development*, *64*(3), 657–674.

Ferry, A. L., Fló, A., Brusini, P., Cattarossi, L., Macagno, F., Nespor, M., & Mehler, J. (2016). On the edge of language acquisition: Inherent constraints on encoding multisyllabic sequences in the neonate brain. *Developmental Science*, *19*(3), 488–503. doi: 10.1111/desc.12323

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3- and 4-month-old infants: an advantage of words over tones. *Child Development*, *81*(2), 472–479.

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, *303*(5656), 377–380.

Floccia, C., Nazzi, T., & Bertoncini, J. (2000). Unfamiliar voice discrimination for short stimuli in newborns. *Developmental Science*, *3*(3), 333–343.

Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., & Collins, D. L. (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, *54*(1), 313–327.

doi: http://dx.doi.org/10.1016/j.neuroimage.2010.07.033

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. *Nat Commun*, *5*.

Forget, J., Buiatti, M., & Dehaene, S. (2009). Temporal integration in visual word recognition. *J Cogn Neurosci*, *22*(5), 1054–1068.

doi: 10.1162/jocn.2009.21300

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.

doi: 10.1111/j.1467-7687.2008.00794.x

Frank, M. C., & Tenenbaum, J. B. (2011). Three ideal observer models for rule learning in simple languages. *Cognition*, *120*(3), 360–371. doi: 10.1016/j.cognition.2010.10.005

Friederici, A. D., Friedrich, M., & Christophe, A. (2007). Brain Responses in 4-Month-Old Infants Are Already Language Specific. *Current Biology*, *17*(14), 1208–1211. doi: 10.1016/j.cub.2007.06.011

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.

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6-and 12-month-olds. *Cognition*, *105*(1), 218–228.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, *440*(7088), 1204–1207.

Gerken, L., Dawson, C., Chatila, R., & Tenenbaum, J. (2015). Surprise! Infants consider possible bases of generalization for a single input example. *Developmental Science*, *18*(1), 80–89.

Gerken, L., Jusczyk, P. W., & Mandel, D. R. (1994). When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition*, *51*(3), 237–265.

Gerken, L., & Knight, S. (2015). Infants generalize from just (the right) four words. *Cognition*, *143*, 187–192.

Gervain, J., Berent, I., & Werker, J. F. (2012). Binding at Birth: The Newborn Brain Detects Identity Relations and Sequential Position in Speech. *Journal of Cognitive Neuroscience*, *24*(3), 564–574.

doi: 10.1162/jocn_a_00157

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences*, *105*(37), 14222–14227.

Gervain, J., Nespor, M., Mazuka, R., Horie, R., & Mehler, J. (2008). Bootstrapping word order in prelexical infants: A Japanese-Italian cross-linguistic study. *Cognitive Psychology*, *57*(1), 56–74.

Gilmore, J. H., Lin, W., Prastawa, M. W., Looney, C. B., Vetsa, Y. S., Knickmeyer, R. C., … Gerig, G. (2007). Regional gray matter growth, sexual dimorphism, and cerebral asymmetry in the neonatal brain. *J Neurosci*, *27*(6), 1255–1260.

Giraud, A.-L., & Poeppel, D. (2012). Speech Perception from a Neurophysiological Perspective BT - Springer Handbook of Auditory Research. In *Springer handbook of auditory research* (Vol. 43, pp. 225–260).

Giurfa, M., Zhang, S., Jenett, A., Menzel, R., & Srinivasan, M. V. (2001). The concepts of 'sameness' and 'difference' in an insect. *Nature*, *410*(6831), 930–933.

Glasel, 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*.

doi: S1053-8119(11)00624-0[pii]10.1016/j.neuroimage.2011.06.016

Gliga, T., Volein, A., & Csibra, G. (2010). Verbal labels modulate perceptual object processing in 1-year-old children. *Journal of cognitive neuroscience*, *22*, 2781–2789. doi: 10.1162/jocn.2010.21427

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436.

Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109–135.

Gomez, R. L., Gerken, L., & Schvaneveldt, R. W. (2000). The basis of transfer in artificial grammar learning. *Memory & Cognition*, *28*(2), 253–263.

Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science*, *7*(5), 567–580.

doi: 10.1111/j.1467-7687.2004.00381.x

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. *Dev Sci*, *15*(6), 885–894.

doi: 10.1111/j.1467-7687.2012.01186.x

Goodsitt, J. V., Morgan, J. L., & Kuhl, P. K. (1993). Perceptual strategies in prelingual speech segmentation. *J Child Lang*, *20*(2), 229–252.

Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language*, *51*(4), 548–567.

Grieve, P. G., Emerson, R. G., Fifer, W. P., Isler, J. R., & Stark, R. I. (2003). Spatial correlation of the infant and adult electroencephalogram. *Clinical Neurophysiology*, *114*(9), 1594–1608.

doi: http://dx.doi.org/10.1016/S1388-2457(03)00122-6

Grieve, P. G., Emerson, R. G., Isler, J. R., & Stark, R. I. (2004). Quantitative analysis of spatial sampling error in the infant and adult electroencephalogram. *NeuroImage*, *21*(4), 1260–1274.

doi: http://dx.doi.org/10.1016/j.neuroimage.2003.11.028

Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron*, *65*(6), 852–858.

Gunz, P., Neubauer, S., Golovanova, L., Doronichev, V., Maureille, B., & Hublin, J.-J. (2012). A uniquely modern human pattern of endocranial development. Insights from a new cranial reconstruction of the Neandertal newborn from Mezmaiskaya. *Journal of Human Evolution*, *62*(2), 300–313.

doi: http://dx.doi.org/10.1016/j.jhevol.2011.11.013

Habas, P. A., Scott, J. A., Roosta, A., Rajagopalan, V., Kim, K., Rousseau, F., . . . Studholme, C. (2012). Early folding patterns and asymmetries of the normal human brain detected from in utero MRI. *Cereb Cortex*, *22*(1), 13–25.

doi: bhr053[pii]10.1093/cercor/bhr053

Hallé, P. A., Durand, C., & de Boysson-Bardies, B. (2008). Do 11-month-old French infants process articles? *Language and speech*, *51*(1-2), 23–44.

Hauser, M. D., & Glynn, D. (2009). Can free-ranging rhesus monkeys (Macaca mulatta) extract artificially created rules comprised of natural vocalizations? *Journal of Comparative Psychology*, *123*(2), 161.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*(3). doi: 10.1016/S0010-0277(00)00132-3

Hay, J. F., Pelucchi, B., Estes, K. G., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology*, *63*(2), 93–106.

Hellstöm, B., Karlsson, B., & Müssbichler, H. (1963). Electrode placement in EEG of infants and its anatomical relationship studied radiographically. *Electroencephalography and clinical neurophysiology*, *15*(1), 115–117.

Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc Natl Acad Sci U S A*, *109*(49), 20095–20100.

doi: 1213390109[pii]10.1073/pnas.1213390109

Herman, L. M., & Gordon, J. A. (1974). Auditory delayed matching in the bottlenose dolphin. *Journal of the Experimental Analysis of Behavior*, *21*(1), 19–26.

Hirsh-Pasek, K., Nelson, D. G. K., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, *26*(3), 269–286.

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.

doi: 10.1016/j.cognition.2010.03.006

Hochmann, J.-R., Mody, S., & Carey, S. (2016). Infants' representations of same and different in match-and non-match-to-sample. *Cognitive Psychology*, *86*, 87–111.

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.

Homan, R. W., Herman, J., & Purdy, P. (1987). Cerebral location of international 10-20 system electrode placement. *Electroencephalography and clinical neurophysiology*, *66*(4), 376–382.

Honig, L. S., Herrmann, K., & Shatz, C. J. (1996). Developmental changes revealed by immunohistochemical markers in human cerebral cortex. *Cereb Cortex*, *6*(6), 794–806.

Izard, V., Dehaene-Lambertz, G., & Dehaene, S. (2008). Distinct cerebral pathways for object identity and number in human infants. *PLoS Biol*, *6*(2), e11.

Jasper, H. H. (1958). The ten twenty electrode system of the international federation. *Electroencephalography and clinical neurophysiology*, *10*, 371–375.

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.

doi: 10.1111/j.1467-7687.2009.00886.x

Johnson, S. P., Fernandes, K. J., Frank, M. C., Kirkham, N., Marcus, G., Rabagliati, H., & Slemmer, J. A. (2009). Abstract rule learning for visual sequences in 8-and 11-month-olds. *Infancy*, *14*(1), 2–18.

Jurcak, V., Tsuzuki, D., & Dan, I. (2007). 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. *NeuroImage*, *34*(4), 1600–1611.

doi: http://dx.doi.org/10.1016/j.neuroimage.2006.09.024

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*(3), 675–687.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, *32*(3), 402–420.

Kaufman, J., Csibra, G., & Johnson, M. H. (2005). Oscillatory activity in the infant brain reflects object maintenance. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(42), 15271–15274.

doi: 10.1073/pnas.0507626102

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307–321. doi: 10.1111/j.1467-7687.2007.00585.x

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party". *J Neurosci*, *30*(2), 620–628.

doi: 10.1523/JNEUROSCI.3631-09.2010

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One*, *7*(5), e36399.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795–1804.

Kim, Y. J., Grabowecky, M., Paller, K. A., Muthu, K., & Suzuki, S. (2007). Attention induces synchronization-based response gain in steady-state visual evoked potentials. *Nature neuroscience*, *10*(1), 117–125.

King, J. R., & Dehaene, S. (2014). A model of subjective report and objective discrimination as categorical decisions in a vast representational space. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *369*(March), 20130204. doi: 10.1098/rstb.2013.0204

Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science*, *237*(4819), 1195–1197.

Kouider, S., Long, B., Le Stanc, L., Charron, S., Fievet, A.-C., Barbosa, L. S., & Gelskov, S. V. (2015). Neural dynamics of prediction and surprise in infants. *Nature communications*, *6*.

Kouider, S., Stahlhut, C., Gelskov, S. V., Barbosa, L. S., Dutat, M., De Gardelle, V., . . . Dehaene-Lambertz, G. (2013). A neural marker of perceptual consciousness in infants. *Science*, *340*(6130), 376–380.

Kovacs, A. M. (2014). Extracting regularities from noise: Do infants encode patterns based on same and different relations? *Language Learning*, *64*(s2), 65–85.

Kovacs, A. M., & Mehler, J. (2009). Flexible learning of multiple speech structures in bilingual infants. *Science*, *325*(5940), 611–612.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., … Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*(5326), 684–686.

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *The Journal of the Acoustical Society of America*, *63*(3), 905–917.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21.

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096–9101.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608.

Kuklisova-Murgasova, M., Aljabar, P., Srinivasan, L., Counsell, S. J., Doria, V., Serag, A., … Rueckert, D. (2011). A dynamic 4D probabilistic atlas of the developing brain. *NeuroImage*, *54*(4), 2750–2763.

doi: http://dx.doi.org/10.1016/j.neuroimage.2010.10.019

Lany, J., & Saffran, J. R. (2010). From statistics to meaning: infants' acquisition of lexical categories. *Psychological Science*, *21*(2), 284–291.

Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, *20*(2), 215–225.

Le Troter, A., Auzias, G., & Coulon, O. (2012). Automatic sulcal line extraction on cortical surfaces using geodesic path density maps. *NeuroImage*, *61*(4), 941–949.

doi: http://dx.doi.org/10.1016/j.neuroimage.2012.04.021

Legerstee, M., Barna, J., & DiAdamo, C. (2000). Precursors to the development of intention at 6 months: Understanding people and their actions. *Developmental psychology*, *36*(5), 627.

Leroy, F., Glasel, H., Dubois, J., Hertz-Pannier, L., Thirion, B., Mangin, J. F., & Dehaene-Lambertz, G. (2011). Early Maturation of the Linguistic Dorsal Pathway in Human Infants. *J Neurosci*, *31*(4), 1500–1506.

Leroy, F., Mangin, J. F., Rousseau, F., Glasel, H., Hertz-Pannier, L., Dubois, J., & Dehaene-Lambertz, G. (2011). Atlas-free surface reconstruction of the cortical grey-white interface in infants. *PLoS One*, *6*(11), e27128.

doi: 10.1371/journal.pone.0027128PONE-D-10-05498[pii]

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review*, *74*(6), 431.

Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex*, *15*(10), 1621–1631. doi: 10.1093/cercor/bhi040

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. *Cereb Cortex*.

Luck, S. J. (2005). An Introduction to the Event-Related Potential Technique. *Monographs of the Society for Research in Child Development*, *78*(3), 388.

doi: 10.1118/1.4736938

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*(6), 1001–1010.

Maddieson, I., & Disner, S. F. (1984). *Patterns of sounds*. Cambridge university press.

Mahmoudzadeh, M., Dehaene-Lambertz, G., Fournier, M., Kongolo, G., Godjil, S., Dubois, J., … Wallois, F. (2013). Syllabic discrimination in premature human infants prior to complete formation of cortical layers. *PNAS*.

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science*, 314–317.

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*.

Marcus, G. F. (1999a). Reply to Christiansen and Curtin. *Trends Cogn Sci*, *3*(8), 290–291.

Marcus, G. F. (1999b). Reply to Seidenberg and Elman. *Trends Cogn Sci*, *3*(8), 289.

Marcus, G. F. (2000). Pabiku and Ga Ti Ga Two Mechanisms Infants Use to Learn About the World. *Current Directions in Psychological Science*, *9*(5), 145–147.

Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (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.

doi: 10.1126/science.283.5398.77

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.

doi: http://dx.doi.org/10.1016/j.jneumeth.2007.03.024

Marno, H., Farroni, T., Dos Santos, Y. V., Ekramnia, M., Nespor, M., & Mehler, J. (2015). Can you see what I am talking about? Human speech triggers referential expectation in four-month-old infants. *Scientific reports*, *5*.

Martin, A., Onishi, K. H., & Vouloumanos, A. (2012). Understanding the abstract role of speech in communication at 12months. *Cognition*, *123*(1), 50–60.

Martin, R. M., & Goodman, N. (1960). *Fact, Fiction, & Forecast.* (Vol. 25).

doi: 10.2307/2964684

Martinho, A., & Kacelnik, A. (2016). Ducklings imprint on the relational concept of 'same or different'. *Science*, *353*(6296), 286–288.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*(1), 122–134.

Maye, J., Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111.

McClelland, J. L., & Plaut, D. C. (1999). Does generalization in infant learning implicate abstract algebra-like rules? *Trends Cogn Sci*, *3*(5), 166–168.

Mehler, J., Bertoncini, J., Barrière, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. *Perception*, *7*(5), 491–497.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178.

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*(6174), 1006–1010.

Meyer, L., Obleser, J., & Friederici, A. D. (2013). Left parietal alpha enhancement during working memory-intensive sentence processing. *Cortex*, *49*(3), 711–721.

doi: 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.

doi: 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.

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & psychophysics*, *18*(5), 331–340.

Molnar, M., Gervain, J., & Carreiras, M. (2014). Within-rhythm Class Native Language Discrimination Abilities of Basque-Spanish Monolingual and Bilingual Infants at 3.5 Months of Age. *Infancy*, *19*(3), 326–337.

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.

Morgan, S. T., Hansen, J. C., & Hillyard, S. A. (1996). Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proceedings of the National Academy of Sciences*, *93*(10), 4770–4774.

Morse, P. A. (1972). The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology*, *14*(3), 477–492.

Morse, P. A., & Snowdon, C. T. (1975). An investigation of categorical speech discrimination by rhesus monkeys. *Perception & psychophysics*, *17*(1), 9–16.

Mueller, J. L., Friederici, A. D., & Mannel, C. (2012). Auditory perception at the root of language learning. *Proc Natl Acad Sci U S A*, *109*(39), 15953–15958.

doi: 10.1073/pnas.1204319109

Muller, M. M., Andersen, S., Trujillo, N. J., Valdes-Sosa, P., Malinowski, P., & Hillyard, S. A. (2006). Feature-selective attention enhances color signals in early visual areas of the human brain. *Proceedings of the National Academy of Sciences*, *103*(38), 14250–14254.

Mumby, D. G. (2001). Perspectives on object-recognition memory following hippocampal damage: lessons from studies in rats. *Behavioural brain research*, *127*(1), 159–181.

Murphy, R. A., Mondragon, E., & Murphy, V. A. (2008). Rule learning by rats. *Science*, *319*(5871), 1849–1851.

Näätäneiv, R., Lehtokoski, A., Lennest, M., Luuki, A., Alliki, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*, 432–434.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, *43*(1), 1–19.

Nelson, C. A., & DeRegnier, R. (1992). Neural correlates of attention and memory in the first year of life. *Developmental Neuropsychology*, *8*(2-3), 119–134.

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, *49*(2), 85–117. doi: 10.1016/j.cogpsych.2003.12.002

Ngon, C., Martin, A., Dupoux, E., Cabrol, D., Dutat, M., & Peperkamp, S. (2013). (Non)words, (non)words, (non)words: evidence for a protolexicon during the first year of life. *Developmental Science*, *16*(1), 24–34.

doi: 10.1111/j.1467-7687.2012.01189.x

Ngon, C., & Peperkamp, S. (2016). What infants know about the unsaid: Phonological categorization in the absence of auditory input. *Cognition*, *152*, 53–60. doi: 10.1016/j.cognition.2016.03.014

Nieder, A. (2009). *Prefrontal cortex and the evolution of symbolic reference* (Vol. 19) (No. 1). doi: 10.1016/j.conb.2009.04.008

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: a review. *Journal of vision*, *15*(6), 4.

Oberecker, R., & Friederici, A. D. (2006). Syntactic event-related potential components in 24-month-olds' sentence comprehension. *Neuroreport*, *17*(10), 1017–21.

doi: 10.1097/01.wnr.0000223397.12694.9a

Oberecker, R., Friedrich, M., & Friederici, A. D. (2005). Neural correlates of syntactic processing in two-year-olds. *Journal of cognitive neuroscience*, *17*(10), 1667–78.

doi: 10.1162/089892905774597236

Odabaee, M., Freeman, W. J., Colditz, P. B., Ramon, C., & Vanhatalo, S. (2013). Spatial patterning of the neonatal EEG suggests a need for a high number of electrodes. *NeuroImage*, *68*(0), 229–235.

doi: http://dx.doi.org/10.1016/j.neuroimage.2012.11.062

Oden, D. L., Thompson, R. K., & Premack, D. (1988). Spontaneous transfer of matching by infant chimpanzees (Pan troglodytes). *Journal of Experimental Psychology: Animal Behavior Processes*, *14*(2), 140.

Okamoto, M., Dan, H., Sakamoto, K., Takeo, K., Shimizu, K., Kohno, S., … Dan, I. (2004). Three-dimensional probabilistic anatomical cranio-cerebral correlation via the international 10-20 system oriented for transcranial functional brain mapping. *NeuroImage*, *21*(1), 99–111.

doi: S1053811903005366[pii]

Oshima-Takane, Y., Ariyama, J., Kobayashi, T., Katerelos, M., & Poulin-Dubois, D. (2011). Early verb learning in 20-month-old Japanese-speaking children. *Journal of Child Language*, *38*(03), 455–484.

Pallier, C., Devauchelle, A.-D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(6), 2522–2527.

doi: 10.1073/pnas.1018711108

Palva, J. M., Monto, S., Kulashekhar, S., & Palva, S. (2010). Neuronal synchrony reveals working memory networks and predicts individual memory capacity. *Proc Natl Acad Sci U S A*, *107*(16), 7580–7585.

doi: 0913113107[pii]10.1073/pnas.0913113107

Papousek, M., Bornstein, M. H., Nuzzo, C., Papousek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior and Development*, *13*(4), 539–545.

Patil, A. V., Safaie, J., Moghaddam, H. A., & Wallois, F. (2011). Experimental investigation of NIRS spatial sensitivity. *Biomed Opt Express*, *2*(6), 1478–1493.

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb Cortex*, *23*(6), 1378–1387.

doi: 10.1093/cercor/bhs118

Pena, M., Arias, D., & Dehaene-Lambertz, G. (2014). Gaze Following Is Accelerated in Healthy Preterm Infants. *Psychol Sci*.

doi: 10.1177/0956797614544307

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

Peña, M., Maki, A., Kovacic, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: an optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences*, *100*(20), 11702–11705.

Pena, M., Pittaluga, E., & Mehler, J. (2010). Language acquisition in premature and full-term infants. *Proc Natl Acad Sci U S A*, *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.

Pepperberg, I. M. (1987). Acquisition of the same/different concept by an African Grey parrot (Psittacus erithacus): Learning with respect to categories of color, shape, and material. *Animal learning & behavior*, *15*(4), 423–432.

Perrot, M., Rivière, D., & Mangin, J.-F. (2011). Cortical sulci recognition and spatial normalization. *Medical Image Analysis*, *15*(4), 529–550.

doi: http://dx.doi.org/10.1016/j.media.2011.02.008

Picton, T. W., John, M. S., Dimitrijevic, A., & Purcell, D. (2003). Human auditory steady-state responses. *International journal of audiology*, *42*(4), 177–219.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech communication*, *41*(1), 245–255.

Premack, D. (1983). Animal cognition. *Annual review of psychology*, *34*(1), 351–362.

Price, C. J. (2012). *A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading* (Vol. 62) (No. 2). doi: 10.1016/j.neuroimage.2012.04.062

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, *288*(5464), 349–351.

Regan, D. (1989). Human brain electrophysiology: evoked potentials and evoked magnetic fields in science and medicine.

Riviere, D., Papadopoulos, D., Poupon, C., Poupon, F., Coulon, O., Poline, J.-B., ... Mangin, J.-F. (2000). A structural browser for human brain mapping. *NeuroImage*, *11*(5, Supplement), S912.

doi: http://dx.doi.org/10.1016/S1053-8119(00)91840-8

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

Saffran, J. R., 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.

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.

doi: http://dx.doi.org/10.1016/S0010-0277(98)00060-2

Schneider, T. R., Debener, S., Oostenveld, R., & Engel, A. K. (2008). Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. *NeuroImage*, *42*(3), 1244–1254.

Schurger, A., Marti, S., & Dehaene, S. (2013). Reducing multi-sensor data to a single time course that reveals experimental effects. *BMC Neuroscience*, *14*, 122.

doi: 10.1186/1471-2202-14-122

Seidenberg, M. S., & Elman, J. L. (1999). Do infants learn grammar with algebra or statistics? *Science*, *284*(5413), 433.

Senkowski, D., Schneider, T. R., Foxe, J. J., & Engel, A. K. (2008, aug). Crossmodal binding through neural coherence: implications for multisensory processing. *Trends in neurosciences*, *31*(8), 401–9.

doi: 10.1016/j.tins.2008.05.002

Shady, M., Gerken, L., & Jusczyk, P. W. (1995). Some evidence of sensitivity to prosody and word order in ten-month-olds. In *Proceedings of the 19th boston university conference on language development* (Vol. 2, pp. 553–562). Somerville. MA: Cascadilla Press.

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.

doi: doi:http://dx.doi.org/10.1121/1.2198947

Shi, R., Marquis, A., Gauthier, B., Bamman, D., Magnitskaia, T., & Zaller, C. (2006). Segmentation and representation of function words in preverbal French-learning infants. In *Proceedings of the 30th annual boston university conference on language development* (Vol. 2, pp. 549–560). Cascadilla Somerville, MA.

Shi, R., & Melançon, A. (2010). Syntactic Categorization in French-Learning Infants. *Infancy*, *15*(5), 517–533.

doi: 10.1111/j.1532-7078.2009.00022.x

Shi, R., Werker, J. F., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy*, *10*(2), 187–198.

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.

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, *54*(1), 1–32.

doi: 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-mo-old infants. *Proc Natl Acad Sci U S A*, *108*(15), 6038–6043.

doi: 10.1073/pnas.1017617108

Smirnova, A., Zorina, Z., Obozova, T., & Wasserman, E. (2015). Crows spontaneously exhibit analogical reasoning. *Current biology*, *25*(2), 256–260.

Southgate, V., Csibra, G., Kaufman, J., & Johnson, M. H. (2008). Distinct processing of objects and faces in the infant brain. *Journal of cognitive neuroscience*, *20*(4), 741–749. doi: 10.1162/jocn.2008.20052

Sowell, E. R., Thompson, P. M., Rex, D., Kornsand, D., Tessner, K. D., Jernigan, T. L., & Toga, A. W. (2002). Mapping sulcal pattern asymmetry and local cortical surface gray matter distribution in vivo: maturation in perisylvian cortices. *Cerebral Cortex*, *12*(1), 17–26.

Spence, M. J., & DeCasper, A. J. (1987). Prenatal experience with low-frequency maternal-voice sounds influence neonatal perception of maternal voice samples. *Infant Behavior and Development*, *10*(2), 133–142.

Spierings, M. J., & ten Cate, C. (2016). Budgerigars and zebra finches differ in how they generalize in an artificial grammar learning experiment. *Proceedings of the National Academy of Sciences*, *113*(27), E3977–E3984.

Steinmetz, H., Fürst, G., & Meyer, B.-U. (1989). Craniocerebral topography within the international 10-20 system. *Electroencephalography and clinical neurophysiology*, *72*(6), 499–506.

Streeter, L. A. (1976). Language perception of 2-mo-old infants shows effects of both innate mechanisms and experience. *Nature*.

Tadel, F., #231, Ois, Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A User-Friendly Application for MEG/EEG Analysis. *Computational Intelligence and Neuroscience*, *2011*.

doi: 10.1155/2011/879716

Talairach, J., & Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain. 3-Dimensional Proportional System: An Approach to Cerebral Imaging.* Translated by Mark Rayport. New York: Thieme Medical Publishers Inc. Stuttgart New York: George Thieme Verlag.

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.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC neuroscience*, *10*(1), 21.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science (New York, N.Y.)*, *331*(6022), 1279–1285. doi: 10.1126/science.1192788

Thiessen, E. D. (2012). Effects of inter-and intra-modal redundancy on infants' rule learning. *Language Learning and Development*, *8*(3), 197–214.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental psychology*, *39*(4), 706.

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, *10*(2), 172–175.

Toga, A. W., & Thompson, P. M. (2003). Mapping brain asymmetry. *Nat Rev Neurosci*, *4*(1), 37–48.

Towle, V. L., Bolaños, J., Suarez, D., Tan, K., Grzeszczuk, R., Levin, D. N., … Spire, J.-P. (1993). The spatial location of EEG electrodes: locating the best-fitting sphere relative to cortical anatomy. *Electroencephalography and clinical neurophysiology*, *86*(1), 1–6.

Trehub, S. E. (1973). Infants' sensitivity to vowel and tonal contrasts. *Developmental psychology*, *9*(1), 91.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, 466–472.

Tsuzuki, D., & Dan, I. (2013). Spatial registration for functional near-infrared spectroscopy: From channel position on the scalp to cortical location in individual and group analyses. *NeuroImage*(0).

doi: http://dx.doi.org/10.1016/j.neuroimage.2013.07.025

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., … Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–289.

Van Heijningen, C. A. A., De Visser, J., Zuidema, W., & Ten Cate, C. (2009). Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the National Academy of Sciences*, *106*(48), 20538–20543.

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.

doi: doi:http://dx.doi.org/10.1121/1.3486197

Vouloumanos, A., Martin, A., & Onishi, K. H. (2014). Do 6-month-olds understand that speech can communicate? *Developmental Science*, *17*(6), 872–879.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: the privileged status of speech for young infants. *Developmental Science*, *7*(3), 270–276.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164.

Wang, L., Uhrig, L., Jarraya, B., & Dehaene, S. (2015). Representation of numerical and sequential patterns in macaque and human brains. *Current biology*, *25*(15), 1966–1974.

Wang, X. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological reviews*, 1195–1268.

doi: 10.1152/physrev.00035.2008.

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends Cogn Sci*, *7*(12), 553–559.

Wasserman, E. A., Fagot, J., & Young, M. E. (2001). Same-different conceptualization by baboons (Papio papio): The role of entropy. *Journal of Comparative Psychology*, *115*(1), 42.

Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology*, *43*(3), 217–242.

Waxman, S. R., & Braun, I. (2005). Consistent (but not variable) names as invitations to form object categories: new evidence from 12-month-old infants. *Cognition*, *95*(3), B59–B68.

Waxman, S. R., Lidz, J. L., Braun, I. E., & Lavin, T. (2009). Twenty four-month-old infants' interpretations of novel verbs and nouns in dynamic scenes. *Cognitive Psychology*, *59*(1), 67–95.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12-to 13-month-old infants. *Cognitive Psychology*, *29*(3), 257–302.

Waxman, S. R., & Markow, D. B. (1998). Object Properties and Object Kind: Twenty-One-Month-Old Infants' Extension of Novel Adjectives. *Child Development*, *69*(5), 1313–1329.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63.

Widmann, A., Gruber, T., Kujala, T., Tervaniemi, M., & Schröger, E. (2007). Binding symbols and sounds: evidence from event-related oscillatory gamma-band activity. *Cerebral Cortex*, *17*(11), 2696–2702.

Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of Psychophysiology*, *21*(3-4), 147–163. doi: 10.1027/0269-8803.21.34.147

Wright, A. A., Cook, R. G., Rivera, J. J., Sands, S. F., & Delius, J. D. (1988). Concept learning by pigeons: Matching-to-sample with trial-unique video picture stimuli. *Animal learning & behavior*, *16*(4), 436–444.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, *85*(3), 223–250.

Yakovlev, P. I., & Lecours, A. R. (1967). The myelogenetic cycles of regional maturation of the brain. *Regional development of the brain in early life*, 3–70.

Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy*, *15*(4), 420–433.

Young, M. E., & Wasserman, E. A. (1997). Entropy detection by pigeons: Response to mixed visual displays after same-different discrimination training. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*(2), 157.

Yuval-Greenberg, S., & Deouell, L. Y. (2007). What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. *The Journal of Neuroscience*, *27*(5), 1090–1096.

Zentall, T. R., Galizio, M., & Critchfield, T. S. (2002). Categorization, concept learning, and behavior analysis: An introduction. *Journal of the Experimental Analysis of Behavior*, *78*(3), 237–248.