Dealing with acoustical variability in speech at birth

Cécile Issard

UNIVERSITÉ PARIS DESCARTES

École doctorale 261 : Cognition, Comportements, Conduites Humaines

Laboratoire Psychologie de la Perception

Dealing with acoustical variability in speech at birth

Par Cécile Issard

Thèse de doctorat de Neurosciences Cognitives

Dirigée par Judit Gervain

Présentée et soutenue publiquement le 29 novembre 2018

Devant un jury composé de :

Judit Gervain	Directrice de thèse - CNRS & Université Paris Descartes
Anne-Lise Giraud	Rapporteuse - Université de Genève
Fabrice Wallois	Rapporteur - Université Jules Verne de Picardie
Gábor Háden	Examinateur - Hungarian Academy of Sciences
Arlette Streri	Examinatrice - Université Paris Descartes

Titre : Construction d'une représentation stable de la parole chez le nouveau-né humain

Résumé :

La parole représente probablement le son naturel le plus important pour l'homme. Les sons de parole étant variables, réagir préférentiellement aux sons de parole implique de répondre à la gamme de paramètres acoustiques qu'ils peuvent prendre. En effet, nous avons tous une voix différente, nous parlons avec des intonations différentes et nous avons peut-être un accent étranger, mais ceux qui nous écoutent perçoivent toujours les mêmes mots et les mêmes phrases. Cela implique que nous avons extrait des unités linguistiques invariantes à partir de sons variables. De même, les enfants apprennent leur langue maternelle à partir de différents locuteurs qui parlent avec des rythmes différents et des intonations différentes d'un moment à l'autre. Cela implique que, dès le début de leur vie, les humains sont capables de former directement des objets invariants à partir du son variable. Pour cela, le code neural doit être flexible envers les paramètres acoustiques que les sons de la parole peuvent prendre, conformément à l'idée que les étages supérieurs du système auditif sont sensibles à la présence d'entités auditives abstraites (ici la parole), plutôt qu'à des paramètres spectro-temporels absolus. Une question clé est donc de savoir comment les humains parviennent à extraire ces représentations invariantes des sons de parole dès le début de leur vie. Cette thèse vise à découvrir comment celles-ci sont construites chez le nouveau-né humain en mesurant les réponses hémodynamiques et électrophysiologiques du nouveau-né à la parole naturelle et à la parole modifiée temporellement ou spectralement.

Lors d'une première expérience, nous avons présenté de la parole normale ainsi que de la parole modérément compressée (60% de la durée initiale) ou fortement compressé dans le temps (30% de sa durée initiale) dans la langue maternelle des participants. Nous avons enregistré la réponse hémodynamique à ces stimuli sur les cortex frontal, temporal et pariétal à l'aide de la NIRS. Les résultats ne montrent aucune différence entre la parole normale et la parole compressée à 60 %, mais des réponses différentielles entre la parole normale et la parole compressée à 30 % ainsi qu'entre la parole compressée à 60 % et la parole compressée à 30 % dans un ensemble de régions frontales, temporales, et temporo-pariétales, de la même manière que le cerveau adulte. Ceci montre que le cerveau du nouveau-né répond à la parole de manière stable sur une gamme d'échelles de temps similaire à celle observée précédemment chez l'adulte. Dans une deuxième série d'expériences, nous nous sommes demandé si cette capacité s'appuie sur l'expérience prénatale de la structure rythmique de la langue maternelle. Nous avons reproduit la même expérience dans deux langues inconnues, l'une rythmiquement similaire à la langue maternelle (l'espagnol) et l'autre rythmiquement différente (anglais). En anglais, seule la parole compressée à 30% évoque des réponses significatives dans une région temporo-pariétale également activée pour le français, mais le schéma exact d'activations est différent de celui du français. Cela confirme que la parole compressée à 30% est

traitée différemment de la parole normale et de la parole compressée à 60%. Cela montre également que l'expérience prénatale façonne le traitement de la parole à la naissance. En particulier, une expérience prénatale de la structure prosodique ou phonologique de la langue pourrait aider les nourrissons à coder la parole de manière stable en fournissant des repères auditifs dans le signal.

En conclusion, les résultats présentés dans cette thèse soutiennet l'idée que la parole est codée comme un objet auditif abstrait dès les premières étapes du traitement auditif. Ce code auditif est en outre modulé par un traitement linguistique de plus haut niveau, intégrant la connaissance de la langue maternelle de l'auditeur. Ces connaissances sont probablement acquises à partir de la vie intra-utérine, ce qui permet un codage stable de la parole, adapté à l'environnement linguistique de l'auditeur dès la naissance.

Mots-clés : Variabilité acoustique, perception de la parole, nouveau-nés, spectroscopie de proche infrarouge, électro-encéphalographie.

Abstract:

Speech probably represents the most important natural sound for humans. As speech sounds are variable, responding preferentially to speech sounds implies responding to the range of acoustical parameters that they can take. Indeed, we all have a different voice, we speak with different melodies, and we might have a foreign accent, but those who listen to us still all perceive the same words and phrases. This implies that we have extracted invariant linguistic units from variable sounds. Similarly, infants learn their native language from various speakers who speak with different speech rates and voice qualities from moment to moment. This means that, from the beginning of their life, humans are able to directly form invariant objects from the raw, variable sound. This implies that the auditory code should be flexible towards the broad range parameters than speech sounds can take, consistent with the idea that the higher stations of the auditory system are sensitive to the presence of abstract auditory entities (in our case speech), rather than absolute spectro-temporal parameters. Therefore a key question is how humans manage to extract these invariant representations of speech sounds from the beginning of their lives. The present thesis aims to uncover how these invariant representations of speech are built in human newborns by measuring newborns' hemodynamic and electrophysiological responses to natural speech, and temporally or spectrally modified speech.

In a first experiment, we presented normal speech as well as moderately (60% of initial duration) or highly time-compressed (30% of its initial duration) speech in the participants' native language (French). We recorded the hemodynamic response to these stimuli over the frontal, temporal and parietal cortices using NIRS. The results show no difference between normal and 60%-compressed speech, but differential responses between normal and 30%-compressed speech as well as be-

tween 60%- and 30%-compressed speech in a set of frontal, temporal, and temporo-parietal regions, similarly to the adult brain. This provides evidence that the newborn brain responds to speech in a stable manner over a range of time-scales that is similar to previous findings in adults. In a second set of experiments, we asked whether this ability relies on prenatal experience with the native language's rhythmic structure. We replicated the same experiment in two unfamiliar languages, one that is rhythmically similar to the native language (Spanish), and one that is rhythmically different (English). In English, only 30%-compressed speech evoked significant responses in a temporo-parietal region also activated for French, but the exact pattern of activations was different from those for French. This confirms that 30%-compressed speech is processed differently than normal and 60%-compressed speech. This also shows that prenatal experience shapes speech processing at birth. In particular, prenatal experience with the prosodic or phonological structure of the language might help infants encode speech in a stable way by providing auditory landmarks in the signal.

To conclude, the results presented in this thesis support the idea that speech is encoded as an abstract auditory object from the first stages of auditory processing. This auditory code is further modulated by higher level linguistic processing, integrating knowledge of the listener's native language. This knowledge is likely acquired from intra-uterine life, enabling a stable encoding of speech, adapted to the listener's linguistic environment from birth.

Keywords: Acoustical variability, speech perception, newborns, near-infrared spectroscopy, electro-encephalography.

Dédicace

Il paraît que personne ne lit les thèses. Pour ma part j'en ai lues plusieurs. Pour m'inspirer, pour trouver une revue de la littérature complète, ou tout simplement parce-que les travaux qu'elles rapportaient n'avaient pas été publiés ailleurs. Je dédis cette thèse à mes lecteurs potentiels, en espérant que ce manuscrit vous apporte quelque chose.

Acknowledgement

Une thèse est toujours un travail collectif. Je souhaite remercier toutes les personnes qui, de près ou de loin, ont contribué à ce projet.

Judit tout d'abord, merci de m'avoir poussée à défendre mes idées et de m'avoir appris à les exprimer clairement, dans un cadre théorique solide et cohérent.

Je remercie également ceux qui m'ont tranmis leurs expertises techniques ou ont répondu à mes questions bizarres : Renske Huffmeijer, Laura Dugué, Lionel Granjon, Nawal Abboub, Fabrice Wallois, Mehdi Mahmouzadeh, Adalan Aarabi, Chris Angeloni.

J'ai eu la chance de connaitre des laboratoires où règnait une atmosphère d'émulation. Thierry, je me suis souvent sentie comme un ovni dans ton équipe, mais j'admire toujours la cohésion qui y règne. Arlette, tu m'as transmis ton intérêt pour les nouveau-nés, et égayé mes déjeuners avec tes histoires abracadabrantes. Alejandrina, les quelques mois à travailler pour toi ont été une véritable bouffée d'oxygène. Maria, les quelques semaines passées dans votre laboratoire demeurent comme une parenthèse lumineuse dans ces quatre années, grâce à laquelle j'ai ensuite pu voir mon travail sous un autre angle. Merci pour votre soutien sans faille.

Merci aussi à tous ceux qui, par un regard, une écoute, un mot ou un geste, m'ont retenue ici: Elena Koulagina, Lucie Martin, Laurianne Cabrera, Carline Bernard, Léo Nishibayashi, Sophie Bouton, Maxine Dos Santos, Katie Von Holzen, Viviane Huet, Mélanie Hoareau, Caterina Marino, Vincent Forma, Arielle Veenemans, Sergiu Popescu, Lisa Jacquey, Daphné Rimsky-Robert, Solène Le Bars, ainsi que tous mes amis.

Je remercie également ma famille qui, par son ignorance totale des sciences cognitives, m'a obligée à savoir vulgariser sans (trop) trahir les résultats. Vous avez été de supers cobayes pour le recrutement de familles à la maternité !

Enfin, Sandro, mon premier conseiller et soutien. Merci d'avoir été toujours présent et un parfait petit elfe de maison dans les dernières semaines.

A tous, merci d'avoir été là.

Contents

List of Figures

List of Tables

Publications

The present work has led to the following publications:

Issard, C.& Gervain, J. (submitted). Adaptation to time-compressed speech in the newborn brain in an unfamiliar language.

Issard, C.& Gervain, J. (2018). Variability of the Hemodynamic Response in Infants: Influence of Experimental Design and Stimulus Complexity. *Developmental Cognitive Neuroscience.*

Issard, C. & Gervain, J. (2017). Adult-like perception of time-compressed speech at birth: An fNIRS study. *Developmental Cognitive Neuroscience.*

Part I

General Introduction

Speech probably represents the most important natural sound for humans. As speech sounds are variable, understanding speech sounds implies responding to the range of acoustical parameters that they can take. Indeed, we all have a different voice, speak faster or slower, the melody of our voice changes all the time, and we might have a foreign accent. Although a word is never pronounced twice the same way, and the sound corresponding to this word is never the same, those who listen to us perceive it as the same word containing the same syllables. We recognize words and other linguistic units although they sound very different, implying that we have extracted invariant linguistic representations from variable sounds. Similarly, babies learn their native language from numerous speakers who speak very differently. They hear many voices around them, but they easily learn their native language without explicit training in the first years of life. This means that, as adults, they are able to recognize speech and its linguistic units in various acoustical forms. This is even more impressive than in adults, as infants cannot look for something that resembles a word they already know, but have to directly form an invariant object from the raw, variable sound. This means that the auditory code needs to be flexible towards the broad range parameters than speech sounds can take, consistent with the idea that the higher stations of the auditory system are sensitive to the presence of abstract auditory entities (in our case speech), rather than absolute spectro-temporal parameters. Therefore a key question is how humans manage to extract these invariant representations of speech sounds from the beginning of their life. The present thesis investigates how these invariant representations of speech emerge in human newborns. More specifically, we aim to understand whether and how the newborn brain encodes speech so as to create invariant linguistic representations.

The human brain goes through dramatic changes during development. The auditory coding schemes observed in adult organisms might have emerged from more basic ones present earlier in development. A given function develops based on the already existing material. Thus newborns' cerebral functional networks constitute the building blocks of the mature systems, in this case speech perception. The newborn brain, therefore, represents an opportunity to understand the initial state of speech encoding.

We seek to establish whether or not newborns can produce stable responses to variable speech sounds, and if yes under what conditions. Intra-uterine life offers an important amount of sensory stimulation. We, therefore, aim to understand how prenatal experience with speech shapes perception, in particular the ability to extract invariance. Specifically, the prosodic structure of language, which is conveyed to the fetus by the uterine environment, might help infants encode speech in a stable way, flagging landmarks in the signal that cue the boundaries of linguistic units. We will address these issues experimentally by testing newborn infants' brain responses to temporally or spectrally distorted utterances in the native language (French), in a prosodically similar unfamiliar language (Spanish) and a prosodically different unfamiliar language (English) using near-infrared spectroscopy (NIRS) and electroencephalograny (EEG).

This thesis is organized as follows. In the General Introduction, we will first describe what a speech sound is (Chapter 1). We will then review how temporal and spectral variability of speech are managed in human adults. We will then focus on the neonatal population, explaining why it is an interesting one to study sensory processing and speech perception, reviewing the development of auditory perception at birth, and arguing that the numerous speech perception capacities displayed by newborns suggest that they are already able to deal with the acoustical variability of speech. We will finally describe the two methodologies that we used in the following experimental chapters, namely NIRS and EEG (Chapter 2). Three experimental chapters will then report the NIRS (Chapters 3 and 4), and EEG (Chapter 5) studies conducted. The implications of the obtained results and future perspectives are then discussed in the General Discussion.

Chapter 1

Theoretical background

1.1 What is a speech sound?

1.1.1 Spectral characteristics of speech sounds

Speech is a complex sound created by a sound source, i.e. a carrier, dynamically modulated by the vocal tract. This carrier, together with the modulations, determines the spectral structure of speech sounds at a given time point. Speech sounds might have one of two types of spectra: the harmonic and the continuous spectrum. Therefore speech is a mix of harmonic sounds with periodic fluctuations, often comprised between 50 and 500 Hz, and non-harmonic sounds with aperiodic fluctuations, often above 1 kHz (Rosen, 1992).

The first type of spectral structure of speech sounds is the continuous, broadband one. In this type of sounds, the carrier is the breath exhaled from the lungs without vibration of the vocal folds. When passing through the constrictions of the vocal tract the flow of air becomes aperiodic, i.e. turbulent, and produces a continuous, noise-like spectrum over a broad frequency band (Dudley, 1940). This category of sounds is called unvoiced and characterizes certain types of consonants. The aperiodic spectrum of unvoiced consonants has the highest energy at high frequencies (Rosen, 1992).

The second type of spectral structure of speech sounds is the harmonic one. This type of spectrum is produced by the vibration of the vocal folds. When individuals speak, they tighten their vocal folds across the larynx in a cyclic way. This leads to the vibration of the vocal folds when breath is exhaled from the lungs. The periodicity of the vibration determines the spectrum of the sound: the frequency of vibration of the the vocal folds corresponds to the fundamental frequency (F0), i.e. the lowest frequency of the spectrum, followed by its harmonics (also called formants), i.e. sine waves at each multiple of the fundamental frequency. The energy decreases with harmonics (i.e. the lowest harmonics have the most energy). The fundamental frequency is the main acoustic cue for perceived pitch (Moore, 2012), whereas formants characterize phonemic category. Speech sounds presenting this type of spectrum are called voiced sounds. They include

vowels and certain types of consonants (Chiba and Kajiyama, 1941).

This basic sound, the carrier, is dynamically modulated by the vocal tract. The vocal tract acts like a filter modulating the sound source (Fant, 1971). As the sound travels through the vocal tract, it is filtered by its cavities. By actively modifying the shape of their vocal tract and the size of the different cavities, speakers give rise to spectral peaks in the sound, suppress some frequencies, or introduce noise-like aperiodic sounds. Each cavity amplifies some frequencies, called its resonance frequencies, and reduces others depending on its shape and size. For voiced sounds, the vocal tract modulates the amplitude of the harmonics. As a result, some harmonics are more prominent than others in the speech spectrum.

For unvoiced, noise-like speech sounds, the vocal tract acts like a band pass. The final sound is aperiodic within a restricted frequency band centered at the resonance frequency of the vocal tract. The frequency band within which the spectrum is continuous depends on the location of constriction (Heinz and Stevens, 1961). If the constriction occurs at the back of the vocal tract (i.e. at or close to the vocal folds), the sound passes through several cavities and is therefore more modulated than if it is generated by a constriction at the front of the vocal tract (i.e. at the teeth or lips) and doesn't travel through vocal cavities. The configuration of the vocal tract can also suppress some harmonics from the speech spectrum by shunting the sound to the nasal cavity. These suppressed frequencies are called zero poles or anti-formants (Fujimura, 1962). Noise-like, turbulent sounds can also be introduced on top of a harmonic voiced sound by blocking the air and suddenly releasing the air flow. In the case of unvoiced sounds, the air flow can be blocked at different locations of the vocal tract, creating a continuous spectrum in a frequency band corresponding to the resonance frequency of the cavity where blocking occurs. A continuous spectrum at low frequencies can also be introduced by shunting the air flow to the nasal cavity (Fujimura, 1962).

Therefore, a main characteristic of speech is that it is a modulated signal. In the spectral dimension by the vibration frequency of the vocal folds and the configuration of the vocal tract at each time point.

1.1.2 Temporal characteristics of speech sounds

Speech intensity and spectral content vary over time in a characteristic way. The temporal structure of speech sounds can be decomposed into its amplitude envelope and its temporal fine structure (Rosen, 1992; Moore, 2008). The amplitude envelope, also known as amplitude modulations (AM) is defined by the slow fluctuations of the amplitude of the acoustic wave. Regardless of spectral content, the amplitude of a sound can vary, superimposing variations of intensity over time on top of the vibration frequency (Figure 1.1.2). In speech, modulations of amplitude over time are comprised between 2 and 120 Hz roughly (Rosen, 1992). The temporal fine structure can be defined as the instantaneous frequency, varying around a center frequency (Moore, 2008; Moon and Hong, 2014). Regardless of intensity, the instantaneous carrier frequency of the sound varies over time around

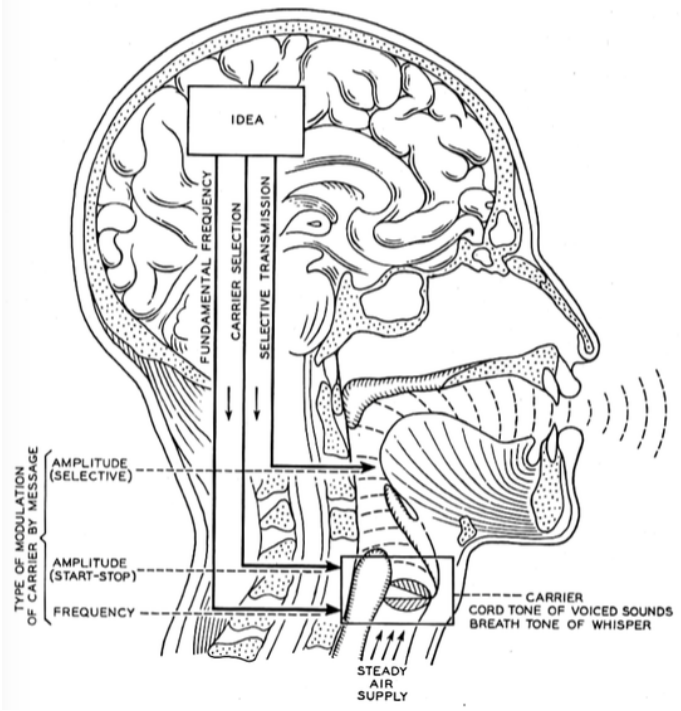


Figure 1.1: Speech as a modulated carrier. Adapted from Dudley et al. (1940).

a given central frequency. Temporal fine structure of speech is comprised between 600 Hz and 10 kHz (Rosen, 1992). They give speech sounds their spectral shape and so contain formant patterns. This temporal fine structure is modulated, as instantaneous frequency oscillates around a central value.

Temporal modulations in speech have characteristic rates and magnitudes, both for amplitude and frequency modulations. These characteristic values are visible in speech AM and FM spectra, defined as the amount of modulation as a function of modulation rate. The FM spectrum of speech follows a low-pass shape, with strong modulation below 8 Hz (Houtgast and Steeneken, 1985; Sheft et al., 2012; Varnet et al., 2017). The FM spectrum shows no difference between languages (Varnet et al., 2017). The AM spectrum shows a band-pass shape, with a peak of maximal amplitude or power between 4 and 5 Hz across numerous languages from different linguistic classes (Ding et al., 2017). However, small differences between languages arise in well-controlled semi-read speech segments, with a peak at about 5 Hz for syllable-timed languages and at about 4 Hz for stress-timed languages (Varnet et al., 2017).

More complex models show that speech temporal structure comprises several temporal integration windows at different time scales (Selkirk, 1986; Nespor and Vogel, 2007). Peaks in the amplitude envelope are clustered at different time scales in a hierarchical way: they form small groups within small time-windows, and these small groups form larger ones when considered in larger time windows. This property has been found in a large variety of speech sounds (Kello et al., 2017).

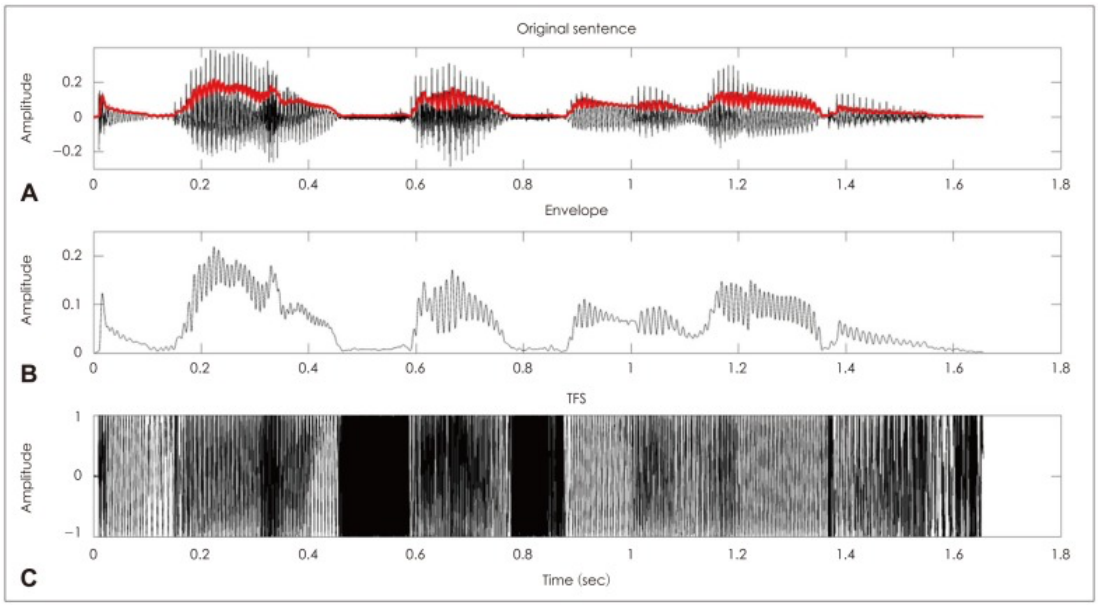


Figure 1.2: Schematic representation of temporal modulations in speech. A: Waveform of the original sentence (black), with its envelope superimposed (red). B: Envelope of the sentence shown in A. C: Temporal fine structure of the sentence shown in A. Adapted from Moon & Hong (2014).

The largest temporal window or slowest modulation scale is comprised between 1 and 4 Hz and corresponds to the phrasal level. Within these modulations, faster ones can be observed going from 4 to 8 Hz, corresponding to the syllable rate. This modulation rate has been identified as prominent across several languages, as visible in their speech modulation spectra (Ding et al., 2017; Varnet et al., 2017). This intermediate level finally contains smaller time windows of 20 to 30 ms, corresponding to modulations between 30 and 50 Hz. These rapid amplitude modulations correspond to the phonemic rate (Goswami and Leong, 2013). These models may differ in terms of the levels that they take into account: some models focus on the syllable (Ding et al., 2017), whereas others add the prosodic foot to take the stress-pattern into account (Goswami and Leong, 2013).

At the phonological level, speech temporal structure can be characterized by speech rhythm. Languages can be divided into three rhythmic classes: syllable-timed (e.g. French and Spanish), stress-timed (e.g. English), and mora-timed (Japanese). This hypothesis was originally made following impressionistic differences when listening to languages, with the idea that speech temporal structure is organized in isochronous units, namely syllables, interstress intervals, and morae (Lloyd James, 1940; Pike, 1945). No acoustical evidence has supported the isochrony hypothesis. Instead, more recently, phonological accounts for these three categories have been provided. Relying on consonant and vowel durations, languages can be classified according to the ratio between the percentage of the

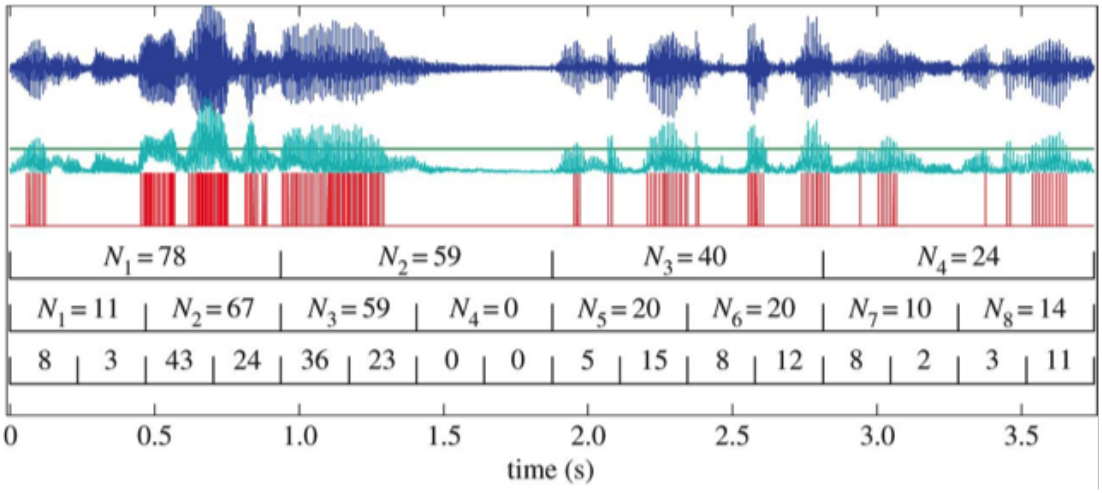


Figure 1.3: Top (blue): speech waveform. Middle (green): amplitude envelope. Bottom (red): peak event series. N: Event counts inside time windows (represented by brackets). Adapted from Kello et al. (2017).

sentence duration taken up by vowels, and the standard deviation of vocalic or consonantal intervals within sentences (Ramus et al., 2000). However, some languages remain difficult to classify in this typology (Grabe and Low, 2002). These metrics have also been shown to be highly sensitive to speaking style and speaker, more than to language and rhythmic class (Arvaniti, 2012), consistently with cross-linguistic differences in the AM spectra only for well-controlled speaking style (Varnet et al., 2017).

1.1.3 Speech sounds are variable

The acoustical characteristics of speech described in the two previous sections all vary in a significant way from one situation to another. Some acoustic variation carries linguistic significance, other differences are linked to meta-linguistic factors (e.g. speaker identity), yet others are random. Listeners need to be sensitive to relevant variation, while disregard irrelevant, random variability in the signal.

Linguistically, variability in the spectral domain allows to differentiate certain linguistic units. Spectral shape is one of the dimensions that define a phoneme. Therefore, spectral modulations introduced by the speaker's vocal tract on a similar carrier differentiate phonemes (e.g. /u/ vs. /a/, see figure 1.1.3). In some languages, variation in pitch is used as a prosodical cue: higher pitch indicates stress, which allows listeners to discriminate words with different lexical stress. High-pitched syllables can also signal the boundaries between units at all levels of the sentence, e.g feet, words, or clauses (Nespor and Vogel, 2007). Variation in the F0 range within sentences can be used to emphasize a word or a clause (Vaissière, 1983). F0 contour also indicates whether a sentence is declarative or interrogative.

Spectral variability can also carry information about the speaker. Formant

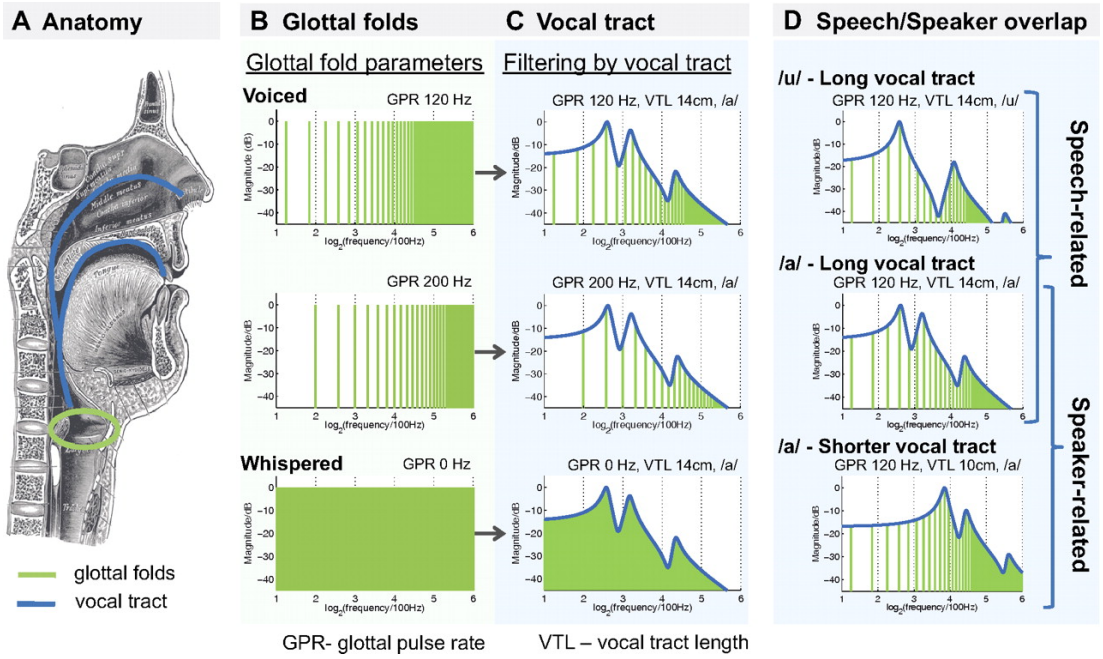


Figure 1.4: Spectral variations in the speech signal. Adapted from von Kriegstein et al. (2010).

ratio theories had initially stated that a given speech sound could be characterized regardless of its pitch: speech sounds would be relative patterns between formants, not absolute formant frequencies (Potter and Steinberg, 1950). Males and females have vocal tracts of different lengths. The longer vocal tract in males shifts the F0 the spectral content of speech, with male producing speech with a lower pitch than female speakers, who themselves produce speech with a lower pitch than children (e.g. Atkinson, 1978; Peterson and Barney, 1952). Accordingly, different spectral shapes between women and men have been measured for several types of consonants (Hagiwara, 1995; Schwartz, 1968). Regarding vowels, F0 and mean formants frequency vary between speakers of the same gender, speakers being distributed around an F1-F2 ratio for each vowel measured (Johnson et al., 1993). Finally, within speakers, F1 frequency may vary of +- 3 s.d. between acoustical realizations of the same vowel (Peterson and Barney, 1952). The scale factors of each vowel-formant couple were stable across languages (Fant, 1966, 1975), providing evidence for a range of variability that can be tolerated around the characteristic spectral shape.

Finally, F0 varies as a function of the speaking style and the emotion conveyed by the speaker. Emotions are associated with different mean F0 and different variability in F0. Sadness is reflected by reduced F0 and F0 range, whereas joy, anger, and fear are reflected by increased F0 and F0 range (Banse and Scherer, 1996). Although the mean F0 remains stable over speaking styles, vowel formants are shifted or spread compared to clear speech (Picheny et al., 1986).

Speech also varies significantly in the temporal domain. As for the spectrum, variability in the temporal domain can also differentiate units of different meaning. Vowel duration is reduced in function as compared to content words (Picheny et al., 1986). Some languages use variation of duration as phonologically contrastive, i.e. to define different phonemes (e.g. /a/ vs /a:/ in Finno-Ugric languages). Variation in the duration of speech sounds is also an important prosodic cue: longer syllables can indicate stress, which allows listeners to discriminate words in the case of lexical stress, or signal the boundaries between prosodic units (Nespor and Vogel, 2007). Speakers can also emphasize some words by varying their duration or inserting pauses before an important one (Vaissière, 1983), which is reflected in speech rate.

Speech rate varies between speakers (Johnson et al., 1993), and across time. Various speech rates have been reported in the literature, from 130 words per minute (Keitel et al., 2018) to 160 (Giordano et al., 2017) and 210 words per minute (Di Liberto et al., 2015). Speech rate is also modulated by the emotions conveyed by the speaker. It is lower in fear than in joy and anger (Scherer et al., 1991).

This variability between utterances is context-dependent, varying with speaking style. In clear speech (i.e. when speakers aim to be more intelligible), speech rate is reduced by both inserting pauses between words and lengthening individual speech sounds as compared to conversational speech. Conversely, in conversational speech, vowels are reduced, and stop bursts are often not released, leading to shorter speech sounds (Picheny et al., 1986). Comparing adult-directed speech (ADS) to infant-directed speech (IDS), IDS has a slower rate. In IDS, temporal modulations are shifted towards lower modulation rates as compared to ADS. Even temporal structure can vary: in English, IDS presents a lower synchronization between syllable-rate and phonemic-rate AM, but greater synchronization between stress-rate and syllable-rate AM in IDS. This effect is modulated by the age of the targeted infant, highlighting the fact that speech sounds varies with context, which includes the listener (Fernald et al., 1989; Leong et al., 2014; Soderstrom, 2007).

Finally, as compared to conversational speech, clear speech shows a bigger difference in power (perceived as loudness) between consonants and vowels (increase in RMS amplitude of consonants, particularly stop consonants) (Picheny et al., 1986). Variations in intensity are also used to indicate prosodic stress.

To sum up, the acoustical properties of speech constantly vary, both in the spectral and the temporal domains. The same variations carry both linguistic (e.g. word stress), and paralinguistic information (e.g. speaker identity or emotion). Variability is therefore a key feature of speech sounds, which contrasts with the invariant perception of speech that listeners experience everyday.

1.2 A stable perception of a variable sound

Despite the important variability described above, human listeners perceive speech in a stable way. They instantaneously recognize linguistic units such as phonemes, syllables and words. They perform this task effortlessly and are often not even conscious of the amount of variability between acoustical realizations of a given syllable or word. There is no direct relationship between the sound of a linguistic unit and its identity, but humans easily map the two. The efficiency of the human brain in this ability contrasts with the poor performance of machines in automatic speech recognition: for your cell phone to understand what you say, you need to articulate clearly and speak in a stereotyped way. This exemplifies that adapting to the acoustical variability in speech is a complex problem, and that to date we don't understand how the human brain accomplishes this so efficiently. As speech sounds can be described in terms of their spectral and their temporal properties, we review the available data on this problem for these two sources of variability separately.

1.2.1 Adaptation to spectral variability

A great challenge for speech perception comes from the fact that spectral cues can carry both linguistic (e.g. phoneme identity) and paralinguistic information (e.g. gender or emotions) (see section 1.1). Listeners must tease apart spectral variability introduced by speakers and contexts to extract invariant linguistic units.

At a behavioral level, adaptation to spectral variability of speech is performed effortlessly (e.g. Peterson and Barney, 1952). According to vocal tract normalization theories, listeners perceptually evaluate vowels on a talker-specific coordinate system (Joos, 1948). Hence, speaker normalization would be a necessary step achieved prior to the identification of specific speech sounds. In an MEG study, two vowel categories (/a/ vs /o/) were presented to participants, either shifting f0 trajectory (inducing a change in overall voice pitch) or modifying mean formant frequency (shift of the vowel spectrum without modifying spacing of harmonics nor formant pattern). Change in pitch elicited an N1m component in Heschl's gyrus bilaterally, whereas change in vowel or mean formant frequency elicited an N1m in the planum temporale bilaterally, just behind Heschl's gyrus (posterior STS). The early timing of this difference supports the claim that spectral normalization occurs prior to speech recognition (Andermann et al., 2017). In an identification task with lists of words said by multiple speakers, increasing variability (through the number of words and speakers) impeded the performance more when participants attended to the voice than when they attended to the words (Mullennix and Pisoni, 1990). Consistently, adult participants better discriminated vowels while voice changes than they discriminate voices while the syllable changes. Looking at the ERPs, the P3 component was significantly smaller in amplitude and peaked significantly earlier in the vowel task than in the talker task, suggesting that, consistently with the behavioral results, discriminating vowels in the context of several voices required less resources than discriminating voices in the context of

several vowels (Kaganovich et al., 2006). Taken together, these results suggest that when listeners hear speech, they automatically normalize speaker variability at early stages of speech processing.

Other studies suggested that invariant representations could be extracted at higher levels of the speech processing stream. In an fMRI adaptation study, the left posterior medial temporal gyrus showed a rebound of activity when the word changed after adaptation to a word uttered by several speakers, meaning that an invariant representation of the word had already been coded in this region. No rebound of activity was observed when the speaker changed after adaptation to the same speaker pronouncing different words, meaning that this region was insensitive to speaker variability. This effect was not present when stimuli were pseudo-words, indicating that spectral normalization performed in the left medial temporal cortex (a region involved in auditory word form recognition) is specific to language processing (Chandrasekaran et al., 2011). In another study, the cortical activation pattern associated to discriminating several vowels uttered by several speakers partially overlapped with the one associated with discriminating speakers uttering different vowels, providing empirical support for speaker normalization at higher-level linguistic stages (Formisano et al., 2008).

At a behavioral level, variations both in F0 and in spectral shape (i.e. formant spacing) slow down vowel and word recognition (Magnuson and Nusbaum, 2007). fMRI studies revealed an increased activity in the right Heschl's gyrus when participants discriminated shift in F0 despite changes of speaker (Kreitewolf et al., 2014). In another study, varying syllables f0 was associated with increased activity in Heschl's gyrus in both hemispheres, whether the task performed by the participants was linguistic (categorizing the syllables) or not (judging loudness). On the contrary, variations in spectral shape (i.e. mean formant frequency) activated the left STG/STS, and this activation was modulated by the task that participants performed: this region was more activated when participants had to categorize the syllables than when they had to judge the syllable loudness (von Kriegstein et al., 2010). Moreover, the same region was more activated when similar spectral manipulations were applied to speech sounds than when these spectral manipulations were applied to animal vocalizations (von Kriegstein et al., 2007). These results suggest that variability in spectral shape (i.e. mean formant frequency) is managed concurrently to speech recognition. All these results provide evidence that listeners deal with spectral variability in speech across different stages, with pitch normalization at early auditory stages prior to speech sound identification, and spectral shape normalization at later stages, integrated with linguistic processing. Several studies investigated how frequency modulations are encoded in humans. Although investigating the perception of spectral variability was not the main purpose of these studies, frequency modulations are dynamic changes around a center frequency. For this reason studies investigating the encoding of frequency modulations give insightful informations to study the encoding of spectral variability, and build hypotheses about how this variability is managed. These studies used electro- or magneto-encephalography that provide information about neural

encoding at the population level (see section 2.2.1). In a first study, human participants listened to a tone modulated in amplitude at 37 Hz and in frequency between 0.3 and 8 Hz. An auditory steady-state response (aSSR, Picton et al. 2003) was observed in the auditory cortex at the AM frequency (i.e. 37 Hz) with spectral sidebands (i.e. secondary peaks) at AM+/- FM frequencies (e.g. 37.8 Hz if the tone FM rate was 0.8 Hz). Moreover, for slower FM (< 5 Hz), the phase of the aSSR at the AM frequency tracked the stimulus carrier frequency change (Luo et al., 2006). This encoding of frequency modulations through phase-locking was confirmed by subsequent studies. In particular, using complex tones frequency-modulated at 3 Hz, an aSSR was observed at the same frequency, and these delta oscillations were phase-locked to the tone FM (Henry and Obleser, 2012). Finally, using a larger range of FM, a last study replicated the aSSR at the FM frequency up to 8 Hz, but not above, providing further evidence for two distinct encoding mechanisms for FM (Millman et al., 2010). These studies show that frequency modulations are encoded by entrainment of neural activity at the FM rate, as shown by an aSSR at the corresponding frequency. Regarding spectral variability in speech, it is possible that a similar mechanism applies. Neural populations of the auditory cortex synchronize at the rate of the FM, this way adjusting their dynamic range to the range of the stimulus spectral variations.

1.2.2 Adaptation to temporal variability

As for spectral cues, temporal cues in speech can carry both linguistic (e.g. phoneme identity) and paralinguistic information (e.g. gender or emotions) (see section 1.1). Listeners must as well tease apart temporal variability linked to meta-linguistic informations or randomness to extract invariant linguistic units.

Adaptation to temporal variability of speech has been extensively studied using time-compressed speech. Behaviorally, successful adaptation to time-compressed speech has been observed in adults and older children, in tasks such as word comprehension (Dupoux and Green, 1997; Orchik and Oelschlaeger, 1977; Guiraud et al., 2013), sentence comprehension (e.g. Ahissar et al., 2001; Peelle et al., 2004), or reporting syllables (Mehler et al., 1993; Pallier et al., 1998; Sebastián-Gallés et al., 2000). In these latter studies, participants had to report the words or syllables they perceived in sentences compressed to 38%-40% of their initial duration, after having been habituated with compressed speech or after having received no habituation. Specifically, using 40% compression, Mehler et al. (1993) presented French or English sentences to French or English monolinguals and to French-English bilinguals. Participants reported higher numbers of words when they were initially habituated to and then tested in their native language. A follow-up study showed that Spanish-Catalan bilinguals adapted to Spanish or Catalan sentences compressed at 38% after habituation in the other language (i.e. habituation in Spanish before test in Catalan, and habituation in Catalan before test in Spanish). In two subsequent studies, English monolingual adults tested with 40% compressed English sentences benefited from habituation to time-compressed speech in Dutch,

which is rhythmically similar to English, and Spanish monolinguals tested with 38% compressed Spanish sentences benefited from habituation with Catalan and Greek, languages that shares rhythmic properties with Spanish (Pallier et al., 1998; Sebastián-Gallés et al., 2000). These studies show that adults are able to adapt to moderately compressed speech in their native language, as well as in unfamiliar languages, if those belong to the same rhythmic class as their native language.

Neuroimaging studies shed light on the cerebral networks supporting this ability. In an fMRI study, (Peelle et al., 2004) presented syntactically simple or complex sentences compressed to 80%, 65%, or 50% of their normal duration. Time-compressed sentences produced activation in the anterior cingulate, the striatum, the premotor cortex, and portions of temporal cortex, regardless of syntactic complexity. Others studies found that some brain regions, e.g. Heschl's gyrus (Nourski et al., 2009) and the neighboring sectors of the superior temporal gyrus (Vagharchakian et al., 2012), showed a pattern of activity that followed the temporal envelop of compressed speech, even when linguistic comprehension broke down, e.g. at 20% compression rate. Other brain areas, such as the anterior part of the superior temporal sulcus, by contrast, showed a constant response, not locked to the compression rate of the speech signal for levels of compression that were intelligible (40%, 60%, 80% and 100% compression), but ceased to respond for compression levels that were no longer understandable, i.e. 20% (Vagharchakian et al., 2012). This variety of response patterns is explained by different temporal receptive windows across brain regions: regions such as early auditory areas have short temporal receptive window, integrating temporal events over short timescales, whereas higher regions such as anterior and posterior STG, IFG and supra-marginal gyrus integrate information over longer time windows and collapse above a certain amount of information contained in their temporal receptive window, i.e. when speech is too compressed (Davidesco et al., 2018; Lerner et al., 2014). In children (from 8 to 13 years old), responses to fast speech occurred in the left-premotor, primary motor regions, and Broca's area, as well as in right inferior frontal and anterior superior temporal cortices (Guiraud et al., 2018). In summary, adaptation to speech temporal variability involves a distributed cerebral network that includes both sensory and higher-level cortical regions.

On the neurophysiological side, several studies have investigated how adaptation to speech rate is accomplished. The role of the theta rhythm (4-8 Hz) for adaptation to speech rate was first suggested by behavioral results. When silent intervals are introduced within sentences compressed to 33% of their initial duration, comprehension is restored if silences occur at a periodical rate that restores the theta-syllable rhythm – 80ms silences following 40 ms of time-compressed speech, aligning with the typical 120 ms duration of syllables (Ghitza and Greenberg, 2009). Accordingly, when adults listen to time-compressed speech, oscillatory activity is phase-locked to the envelope of the sound in the theta band. Speech compressed to 50% or 33% of its initial duration elicited an increase in power shared by EEG and acoustic signals in the frequency band corresponding to the syllabic rate of each condition. Moreover, increased syllable rate in time-compressed speech led

to an increase in power in the theta band, and a decrease in power in the gamma band (27-28 Hz). Phase-locking occurred for normal speech as well as for the two compression rates (Pefkou et al., 2017). These results were partly replicated in typically developing children (from 8 to 13 years old) who listened to normal and naturally fast speech while oscillatory activity was recorded through MEG. For both normal and fast speech, phase-coherence between oscillatory activity and the envelope of the sentences was increased in the theta band (4-7 Hz) as compared to baseline (when no sound was presented). However, no increase in power was observed in this group of children (Guiraud et al., 2018). Looking at a broader frequency band, oscillatory activity was phase-locked to the time-compressed speech envelope in a broad 0-20 Hz band (Ahissar et al., 2001). Phase-locking to the envelope of time-compressed sentences was also observed in the high-gamma band (70Hz and above) across compression levels from 20 to 75% of initial duration, i.e. even when participants didn't understand the sentences (Nourski et al., 2009). All these results were observed whether or not the participant understands the sentences, although phase-locking correlated with comprehension in the broad 0-20 Hz range (Ahissar et al., 2001). These results highlight the fact that different brain regions respond differently to variations in speech rate: early auditory regions track envelope at all compression rates (Davidesco et al., 2018; Nourski et al., 2009), whereas regions further along the auditory hierarchy track the envelope only during comprehension (Davidesco et al., 2018). These results also stress the importance of envelope-tracking by cortical activity as a prerequisite to adapt (as reflected in adults by comprehension) to time-compressed speech, but also shows that envelope tracking is in itself insufficient to explain comprehension. To conclude, adaptation to speech rate variability is done by following the envelope of speech sounds and phase-locking oscillations in the theta and high-gamma bands, which correspond to two important rhythmic scales of speech sounds (i.e. syllabic and phonemic scales), although other mechanisms are also necessary to provide a full account of the perception of time-compressed speech.

1.3 Dealing with acoustical variability in speech at birth

1.3.1 Why newborns: Looking at the auditory foundations of speech perception

Birth represents a particular moment in the course of sensory development. Sensory systems have already started their development and are functional in utero (Abdala and Keefe, 2012; Eggermont and Moore, 2012), but they have never faced faces and visual objects, experienced the intensity of daylight, smelled other odors than the amniotic fluid, or heard unfiltered sounds, without maternal physiological background noise. Sensory signals are greatly modified by maternal tissues that attenuate the light and high-frequency sounds (Querleu et al., 1988). Therefore,

at birth, the physical properties of sensory stimulations suddenly change. Despite transnatal continuity in sensory development, newborns face the external world for the first time, and as such have little experience with broadcast sensory signals. They experience a transition from a dark, attenuated environment to a bright one with louder and spectrally richer sounds. For this reason, they offer a window on the core of sensory systems. It has been suggested that the auditory system is adapted to the statistical structure of natural sounds (Mizrahi et al., 2014; Theunissen and Elie, 2014). Speech is a particularly significant one in the newborn's environment, given the importance of vocal communication in our species. If this hypothesis is true, adaptation of the auditory system to speech should be observable in the absence of linguistic expertise, at the first contact with broadcast speech, i.e. at birth.

The goal here is not to talk about innate capacities as opposed to acquired skills, but rather to describe a set of tools that newborns have to face the external world and with which they will start to build their cognitive skills during the first months of life. Among these available tools, auditory processing is of great importance as it constrains the information fed into the linguistic system (Friederici, 2002). It is thus essential to understand auditory perception in this transitional state in order to understand the building blocks of speech perception. Development provides the opportunity to observe how cognitive functions emerge as an interplay of different biological and experiential factors.

The majority of studies that investigated the processing of variability in speech, especially the perception of temporally and spectrally distorded sounds, involved adults and children (see section 1.2). These participants have high linguistic expertise, which makes it difficult to disentangle the relative contributions of auditory and linguistic linguistic knowledge. Infants are not yet linguistic experts, but they already have intensive experience with speech sounds and important linguistic capacities by the end of the first year of life (Gervain and Mehler, 2010). Newborns represent a temporal window before language acquisition and considerable experience with broadcast sensory signals influence sensory perception. Thus, investigating newborn's speech perception abilities are essential to a better understanding of the auditory foundations of speech processing. The present work focuses on this transitional period, testing infants between 1 and 4 days of life.

1.3.2 Auditory processing at birth

How are newborns equipped to perceive the auditory world when they face it for the first time? Humans are born with a relatively well-developed auditory system. The first behavioral signs of fetal audition, measured by motions or heat rate after a sound is presented close to the mother's abdomen, can be observed around 26 weeks of gestation (Birnholz and Benacerraf, 1983). Anatomically, the cochlea is operational around 24 weeks of gestation. However, the outer and middle ear are still immature at birth, filtering and amplifying sounds differently than the adult ear. The auditory nervous system is also immature at birth, the auditory nerve

showing less precise phase-locking to temporal modulations (Moore and Linthicum, 2007; Abdala and Keefe, 2012).

Despite this physiological immaturity, numerous studies provide evidence for well developed functional responses to sounds. By the last month of gestation, fetuses discriminate speech from piano melodies, irrespective of spectral content, indicating that they perceive temporal structure of complex sounds (Granier-Deferre et al., 2011). Consistently, omission of the downbeat in sequences of musical instruments provokes a mismatch response with two peaks, around 200 and 400 ms, showing that newborns build expectations about the temporal structure of a sound sequence (Winkler et al., 2009). Regarding spectral processing, a pure tone with a different pitch than the preceding ones elicits a mismatch negativity (MMN) response in newborns aged 1 to 4 days (Alho et al., 1990). This result was then replicated with more complex sounds. Newborns present a different brain response to musical sounds of different timbre but the same pitch, than to a sound with a different pitch than the preceding ones (Háden et al., 2009). These results show that newborns are capable of extracting and processing pitch cues in complex sounds. Newborns can even detect complex spectro-temporal structures. Indeed, blocks of stimulation presenting sounds in which the temporal structure scaled relative to the center frequency (a characteristic feature of natural environmental sounds), or sounds that lacked this structure elicited a larger hemodynamic response in the left frontal and left temporal cortices than blocks with an alternation of the two types of structure, indicating that newborns discriminate the two types of spectro-temporal structure (Gervain et al., 2016). This provides evidence that newborns process spectral and temporal features in a integrated way to detect the complex spectro-temporal structures present in natural sounds.

Taken together, these results provide evidence for substantial auditory capacities at birth, laying the foundations for sophisticated speech perception abilities. Given the fact that audition is functional during the last trimester of gestation, and that low-frequency components of speech are transmitted through the womb (Querleu et al., 1988), this opens the possibility of prenatal experience shaping speech perception at birth. In particular, if low-frequency modulations of speech are transmitted through the womb, it is possible that fetuses are familiarized with speech temporal envelope during the last trimester of gestation.

1.3.3 Suggestive evidence for stable perception of speech at birth

Despite their lack of experience with broadcast speech, newborns show prelinguistic capacities that suggest that they are capable of extracting an invariant representation of speech from variable sounds.

A first line of evidence is the preference for speech sounds from birth. Behaviorally, newborns make more headturns to unfiltered speech than to heartbeat (a familiar stimuli from their intra-uterine life) or filtered speech, either low-pass as heard in the womb, high-pass or band-pass (Ecklund-Flores and Turkewitz, 1996).

They also adjust their sucking rate to hear speech sounds rather than sine waves tracking the fundamental and the first three formants. These stimuli retained the duration, pitch contour, amplitude envelope, relative formant amplitude, and relative intensity of their speech counterparts (Vouloumanos and Werker, 2007). These results suggest that newborn participants prefer to hear natural speech sounds, a preference that cannot be explained by familiarity with the stimuli or the simple spectro-temporal properties of the sound. On the neural side, the right temporal and right frontal cortex respond more to speech sounds than to other vocal sounds such as human emotional sounds or monkey calls (Cristia et al., 2014). Together, all these results suggest that readily from birth humans would be equipped with an auditory module dedicated to speech sounds, to process them with auditory and cognitive mechanisms.

A second set of studies investigated speech-specific processing comparing it to backward speech. Backward speech preserves the spectral content of speech but disturbs its temporal structure, sounding nothing like speech while being a good acoustical control. At 3 months, forward speech produced a larger activation than backward speech in the left angular gyrus and left mesial parietal lobe. No greater activation was found for backward speech, adding evidence for a preferential processing of speech sounds from the first months of life (Dehaene-Lambertz et al., 2002). Here, the difference between natural speech sounds and spectral analogs appeared in cortical regions outside of STS and STG, which are typically activated in studies comparing speech to other categories of natural sounds. Different results have been found in the neonatal brain. In a NIRS study, forward speech elicited a larger response than backward speech and silence in the temporal cortex, with larger responses in the left hemisphere for forward speech (Pena et al., 2003). Other studies investigated the role of prenatal experience. In these studies, participants listened to their native language or a foreign language, played either forward or backward. Forward speech evoked a larger activity than backward speech in the left temporal area for the native, but not the foreign language (Sato et al., 2012). This result was replicated and extended in an independent laboratory: when the experiment presented forward and backward speech in the native and a foreign language, forward speech evoked a larger response than backward speech in bilateral fronto-temporal regions in the native language, but not in the foreign language. But when the same foreign language was presented with a whistled surrogate language, forward speech triggered a larger response than backward speech in a similar but smaller region, and only in the left hemisphere. No difference was observed between forward and backward for the surrogate whistled language (May et al., 2018). However, another similar study found the reverse pattern with speech low-pass filtered at 400 Hz to mimic what is heard in the womb. Backward speech elicited a larger response than forward speech in the bilateral temporal cortices and in the right frontal cortex for the foreign language, while the native language triggered a similar response independently of directionality in both the left and right temporal cortices (May et al., 2011). This suggests that the neonatal brain preferentially responds to sounds with the canonical temporal structure of speech,

and that this capacity modulated by context and prenatal experience.

A final set of studies have provided evidence for extraction of invariant speech patterns across speakers. In an ERP study, newborns were presented with sequences of four syllables recorded from four different speakers. Among the four syllables, the first three were always from the same phonemic category, i.e. /ta/ or /pa/, and the fourth one were either from the same category as the first three (standard trials) or from the other category (deviant trials). In half of the trials, the four syllables were spoken by the same speaker, and in the other half by four different speakers. Newborn participants presented a mismatch response to the phonemic change (/ta/ vs /pa/), whether or not the speaker remained the same (Dehaene-Lambertz and Pena, 2001). This result was partially replicated in preterm newborns. In two EEG and NIRS experiments, participants listened to series of four syllables with the last one being either the same as the first three (standard blocks), or from another phonemic category (/ba/ vs /ga/) or spoken by a different speaker (deviant blocks). Interestingly, similar results were observed in ERP data, with a mismatch response only to the deviant phoneme, but time-frequency analysis of the EEG signal also revealed a response to the change of voice, although smaller and in different clusters of electrodes than the change of phoneme (Mahmoudzadeh et al., 2017). NIRS results mirrored the time-frequency results, with a larger hemodynamic response to deviant phoneme as compared to the standard blocks in large portion of the frontal lobe bilaterally, and smaller response that quickly faded away only in one channel of the right frontal lobe to the change of voice (Mahmoudzadeh et al., 2013). Taken together, these results suggest that the neonatal brain identifies speech sounds, and perform pre-linguistic operations on it despite acoustical variability due to multiple speakers.

1.3.4 The aims of the current thesis

To what extent is this preferential processing for speech robust to acoustical variability and distortions? Does the capacity of adult listeners to normalize acoustical variability in speech come from their linguistic expertise or are they part of these initial speech- and vocalizations-specific auditory mechanisms? The present work aims at answering these questions by looking at the processing of spectral and temporal variability of speech before the emergence of considerable linguistic expertise, namely in the neonatal brain.

The aim of this work is to establish how newborns build a stable representation of speech despite its acoustical variability. As described in section 1.1.3, this variability is manifested both in the temporal and the spectral domains. It has been shown that adult listeners are able to normalize speech variability up to a certain level, both in the spectral and temporal domains. The speech perception abilities shown by newborns in the context of several speakers suggest that their auditory system is able to normalize this variability to build a stable auditory representation of speech. However, how the newborn's brain achieve to manage this variability has never been explicitly investigated. This work therefore aims to

fill several objectives:

First to determine whether or not newborns can normalize acoustical variability in speech and build a stable representation of it. If this is the case, determine if this is possible for temporal variability, spectral variability, or both.

Second to elucidate the neural mechanisms that support this capacity in newborns, and map them in the brain. We here aim to determine the building blocks of these circuits, i.e. the brain regions already involved at birth, and possibly cortical regions activated in newborns but not in adults. The existing models of electrophysiological mechanisms to encode speech are based on adult data, and don't make predictions about the developmental origins of these mechanisms. We here aim to explore the neurophysiological tools that newborns have at their disposal to manage speech spectral and temporal variability.

Chapter 2

Techniques used in this thesis

2.1 NIRS

This section is adapted from Issard and Gervain (2018).

2.1.1 Basic principles of NIRS

NIRS is an optical method that can be used to monitor brain activity at rest, in response to sensory stimulation, or in a cognitive task, from birth to adulthood in healthy as well as in clinical populations. NIRS takes advantage of the differential near-infrared light absorption properties of oxy- and deoxyhemoglobin (HbO and HbR, respectively) to detect concentration changes of these two chromophores in blood, i.e. the hemodynamic response.

At the physiological level, NIRS relies on specific properties of the neurovascular coupling in the brain. Increased neural activity triggers an increase in oxygen delivery to the active region a few seconds later. As a result, there is an initial decrease in HbO followed by a strong increase, while HbR concentration decreases (after a possible brief initial overshoot). Upon continued stimulation, HbO levels remain high and then return to baseline, sometimes after an undershoot. In parallel, HbR levels remain low and then return to baseline after a possible overshoot. This pattern of variations in HbO and HbR concentrations over time defines the Hemodynamic Response Function (HRF) . A typical newborn infant hemodynamic response is depicted in Figure 2.1.1. Importantly, more oxygen is delivered to the active brain regions than what is taken up by the tissues, which is why it is possible to measure regional changes in blood oxygenation.

At the physical level, NIRS takes advantage of the optical properties of biological tissues in the red and near-infrared range. In the near-infrared window, i.e. light whose wavelength is above 700 nm, head tissues are almost transparent to light, and HbO and HbR have well separated absorption coefficients. Using these optical properties, NIRS measures the difference in light intensity between a source emitting near-infrared light at several wavelengths and light detectors placed at a systematic distance. Typically two wavelengths are used, one on each side of the

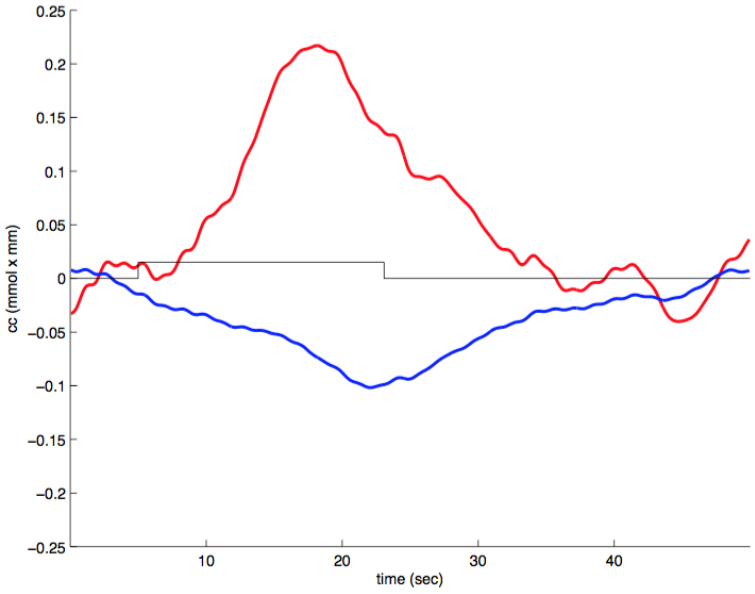


Figure 2.1: A typical hemodynamic response as observed in a newborn participant in our laboratory. Red: HbO, Blue: HbR, Rectangle: stimulation.

isosbestic point of the absorption spectra of HbR and HbO, e.g. one <780 nm and one >830 nm (Scholkmann et al., 2014). A paired source and detector form a measurement channel. In turbid media, photons travel across a broad distribution of random paths due to the effects of scattering (Fukui et al., 2003; Okada and Delpy, 2003a,b). A large proportion of the light detected by the detectors travels through a donut shaped trajectory between a source and the detectors coupled with it. Light that travels in other directions does not reach the detectors, it is scattered. The measured light intensities can be converted into concentrations of HbO and HbR using the Beer-Lambert law, modified to take into account specificities of light travelling through a biological, thus optically non-ideal, non-homogeneous medium, such as the scatter and the non-linear, random path of photons. As most commercially available NIRS systems used in cognitive developmental research are unable to quantify the scatter and the actual paths of photons, these systems can only measure relative concentrations of HbO and HbR, i.e. concentration change relative to a previous time point. Absolute concentrations of HbO and HbR can be measured with NIRS systems able to quantify the scatter and the pathlength (for a review of such commercially available systems, see Wolf et al., 2007). Absolute oxygenation levels are highly informative for clinical purposes, while relative concentrations are sufficient in most experimental settings, where differences between two conditions or a condition and baseline are of interest.

The ideal source-detector separation is variable as a function of the brain region tested, the age of participants and other factors. There is a trade-off between spatial resolution and penetration depth: the farther the optodes, the deeper the penetration, but the lower the spatial resolution and the higher the noise. In

Figure 2.2: Pictures of different fNIRS headgears and their respective approximative channel locations. A: The UCL system (adapted from Lloyd-Fox et al., 2017). B: The Hitachi ETG-4000 system (adapted from May et al., 2011). C: The NIRx NIRScout system as used in our laboratory.

adults, the penetration into the cortex is about 0.5 cm with a source-detector distance of 3 cm (Strangman et al., 2002). In infants, several source-detector distances have been tested (between about 2-5cm), providing a penetration depth into the cortex greater than in adults, up to several centimeters in preterm newborns. Multi-distance channel set-ups, with short-distance channels of less than 2 centimeters, can also be used to record HbO and HbR concentration changes in non-brain tissues. This signal can then be subtracted from the signal obtained from regular channels to improve the signal-to-noise ratio (Emberson et al., 2016). Like functional Magnetic Resonance Imaging (fMRI), fNIRS records the hemodynamic response as a proxy for neural activity in a localized brain region. But unlike fMRI, which detects only HbR, fNIRS records both HbO and HbR, providing more information about the hemodynamic response (Steinbrink et al., 2006).

2.1.2 Why use fNIRS in infant research?

The ease of use of NIRS is particularly relevant for developmental research. Sources and detectors are mounted on a flexible cap that can be easily placed on the participant's head (see Figure 2.1.2 for pictures of commercially available set-ups). In addition, NIRS does not require a shielded room. Furthermore, NIRS is relatively motion-tolerant compared to fMRI, allowing a behavioral paradigm to be used simultaneously. NIRS requires no tracer substance or gel. It is completely silent and uses no magnetic field or radio pulses. It is thus perfectly safe and non-invasive. It is also possible to take the set-up outside the laboratory to reach specific populations and study cognitive development under ecological conditions, such as under-nourishment or recurrent infections in rural Africa (Lloyd-Fox et al., 2016).

fNIRS can be used with developmental populations for whom fMRI is complicated or impossible. Newborns can be tested directly in their bassinet, rendering experiments much easier (Benavides-Varela et al., 2012; Issard and Gervain, 2017; Pena et al., 2003; Rossi et al., 2012). The relatively high motion-tolerance of fNIRS is relevant for testing infants and children, as well as clinical populations likely to produce involuntary movements such as epileptic patients. NIRS also allows longer, even continuous monitoring. The NIRS cap can be left on patients' heads for several hours to capture seizures or other activity whose onset is unpredictable (Wallois et al., 2010). Another clinical population for whom fNIRS is of particular interest is patients with cochlear implant. Cochlear implant users cannot be placed in an MRI because of the metallic parts of the implant. They

cannot be tested with EEG either as the implant interferes with the EEG signal. fNIRS has proven useful to monitor activity in the auditory cortex with children following cochlear implantation (Sevy et al., 2010). fNIRS is thus a useful alternative to fMRI when the age or the pathology of the tested developmental population is incompatible with MRI or with EEG.

Finally, the flexibility of optode arrangement and the compatibility of the two signals makes co-recording with EEG possible (Mahmoudzadeh et al., 2013; Telkemeyer et al., 2009). The two techniques are complementary, NIRS allows spatial localization, while EEG has high temporal resolution. Furthermore, while the hemodynamic response is slow, NIRS technology itself has a higher temporal resolution than fMRI, typically allowing a sampling frequency above 10 Hz for small and medium channel numbers, whereas fMRI typically has a sampling frequency of about 0.5-1.5 Hz. Although the major determiner of temporal resolution using hemodynamic imaging is the temporal dynamics of the physiological signal itself, which is slow, the higher sampling rate is of particular interest for the identification of the temporal dynamics of cortical networks. These investigations are based on the analysis of the peak latency and phase delay of the response. This may reveal the order in which regions are activated, showing the temporal map of activation in the involved network (Mahmoudzadeh et al., 2013). NIRS can also establish functional connectivity between regions based on the synchrony of their respective activation, e.g. by computing correlations or phase synchronization between channels (Benavides-Varela et al., 2017; Homae et al., 2011; Molavi et al., 2014). Because NIRS has a higher sampling rate than fMRI, it leads to more precise temporal measures of connectivity.

2.1.3 Limitations of the fNIRS method

There are several constraints that need to be taken into account when using NIRS. First, the hemodynamic response is slow, operating in the order of seconds, unlike the electrophysiological response, which works at the millisecond range. Both block designs and "event-related" designs can be used, but stimulation periods need to last at least a few seconds. In block designs, a series of stimuli is presented to create long stimulation periods (typically between 5-30 sec), separated by long baseline periods to let the hemodynamic response return to baseline. The length of the baseline periods between blocks depends on several factors such as the age of participants, the amount of activation produced and the type of design. With younger participants, the hemodynamic response may take longer to appear and to return to baseline (Arichi et al., 2012), so block designs with sufficiently long baseline periods are appropriate. In "event-related" designs, stimulation as short as 3 seconds have been used, separated by baseline periods as short as 4 seconds (Emberson et al., 2015; Homae et al., 2006; Taga and Asakawa, 2007). In this type of designs, the end of the hemodynamic response to a stimulus typically overlaps with the response to the next stimulus. This issue can be managed by using General Linear Model (GLM) for data analysis.

Another main limitation of fNIRS is its limited access to deep brain structures and its low spatial resolution. The penetration of NIR light into the head is relatively shallow at a standard 3 cm source-detector separation (with a penetration into the cortex of up to 0.5 cm in adults, 1.5 cm in infants and several centimeters in preterm newborns). Multiple source-detector distances and overlapping channels can be used in high density set-ups to create three-dimensional, i.e. tomographic, images in order to identify the source of the response with better precision and probe deeper areas (Liao et al., 2010), but these technologies are currently not yet sufficiently developed to be routinely used in developmental research. If fiducial measures of probe placement exist, i.e. the position of the probes with respect to external landmarks (ear, nasion etc.) or a co-ordinate system, then the measurement points can be mapped onto the participant's anatomical MRI, if available, or to an age-appropriate template (Lloyd-Fox et al., 2014; Matsui et al., 2014). The cortical area activated can thus be identified with greater precision.

2.2 EEG

2.2.1 Basic principles of EEG

Electro-encephalography (EEG) is a method that allows to measure the electric field associated with brain activity. Sensory and cognitive processes are associated with electric activity in brain cell populations. This electric activity sums up in a global electric field that can be measured non-invasively by placing several electrodes on the surface of the scalp. The measured signal is called the electroencephalogram. As EEG directly measures brain electric activity, it offers the opportunity to measure the electrophysiological states and mechanisms associated with perception and cognition with a high-temporal resolution, at a millisecond scale. Because EEG is non-invasive, it can be used to monitor brain activity in a variety of situations (e.g. at rest, in response to sensory stimulation, or in a cognitive task), from birth to adulthood, in healthy as well as in clinical populations. Finally, EEG can be used in addition to other brain imaging methods with higher spatial resolution, such as fNIRS, as the two techniques provide complementary information about brain activity (Wallois et al., 2010). For all of these reasons, EEG is the most widely used technique to measure brain electric activity in humans.

Despite its large success, EEG remains poorly understood regarding the physiological signal it measures. Because EEG measures electric potential at the scalp, it measures extracellular field potentials (also called Local Field Potentials, LFPs) integrated over large volumes of the brain. The main contributor to the electroencephalogram is synaptic currents. When neurons are active, they produce electric currents in the extracellular media. Neurotransmitters acting on synaptic receptors provoke inward flow of ions at the synapse. To compensate for this, ions are released in the extracellular media. This flow creates an electric field around the active neuron. Between 10 000 and 50 000 need to be synchronously active

Figure 2.3: EEG setup as used in our laboratory.

for their currents to sum up and reach a sufficient amplitude to be detected by EEG (Murakami and Okada, 2006). Synaptic currents being slow events, they can most easily overlap in time to substantially impact LFPs and induce a measurable signal in EEG (Buzsáki et al., 2012). Action potentials can also contribute to the electro-encephalogram, in particular because spikes are followed by long-lasting afterhyperpolarizations which longer time-scale make good candidates to contribute to slow signals observed in EEG (Buzsaki et al., 1988). Still, spikes per se are extremely fast signals that may not overlap in time over large populations, hence making little contribution to the EEG.

A key feature of brain activity measured by EEG is its oscillatory nature. Signals measured at the scalp are periodic, characterized by successive cycles in narrow band frequencies, the most commonly described being delta (1-4Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz), and gamma (30-80 Hz). These oscillations can be observed readily in the raw signal, such as the alpha rhythm over occipital electrodes when adult participants close their eyes or theta rhythm in newborns. It is necessary to decompose the signal in the Fourier domain to parse the different oscillatory components when the brain oscillatory behavior is heterogeneous. Several physiological mechanisms have been proposed to account for these oscillations. A first mechanism involves only excitatory interactions between neurons. In this scenario, excitatory inputs arriving just after a spike delay the next spike, whereas those arriving long after a spike advance the next one. These interactions being reciprocal, neurons soon fire synchronously within a connected

population. This mechanism has been shown to lead to oscillations in the theta range in hippocampal slices (Traub et al., 1992). A second possible mechanism involves reciprocal inhibitory connections. Here, either inhibitory interneurons are activated out-of-phase of each other, a first interneuron inactivating a second one, which in turns inactivates the first one; or interneurons are activated together, which mutually inhibits them until inhibition is released and they fire again in a new cycle. This scenario has been shown to lead to oscillation in the gamma range . A last scenario involves excitatory-inhibitory feedback loop between several types of both excitatory pyramidal neurons and inhibitory interneurons. Interneurons help creating oscillations by feedback and feedforward connections: when activated by excitatory neurons, interneurons inhibit their connected excitatory neurons (including the ones that activated them in the case of feedback connections) for a time-constant, leading to a through in the LFP oscillation. Then, when inhibition is released, excitatory neurons fire again, leading to a peak in the LFP oscillation. They activate the interneurons they are connected to, generating a new cycle in LFP oscillation. All these scenarios lead to secondary synaptic currents detectable by EEG. Besides these direct synaptic connections, oscillations can propagate within neuronal populations by sensing the ions released in the extra-cellular space by active neuronal populations generating LFPs, and this way synchronize to their active neighbors. This phenomenon is called ephatic coupling and contributes to amplify the power of the oscillations (see Wang, 2010, for a detailed review of the mechanisms leading to oscillations in neuronal populations). To summarize, all types of neurons contribute to LFPs, and therefore to the EEG signal. Oscillations measured by EEG reflect cyclic succession of excitation and inhibition in large synchronized neuronal populations. For this reason they are well suited to implement and give rise to cognitive functions.

Besides oscillatory activity, EEG allows to measure evoked potentials. Evoked potentials are sharp deflections in the EEG signal that can be observed when averaging responses to sensory stimulation or periods of cognitive activity (Luck, 2014). Evoked potentials typically contain several positive or negative peaks usually associated with the psychological processing steps, such as sensory processing for the negative peak around 100 ms after stimulus onset (N1) (Davis, 1939), or semantic processing for the negative peak around 400 ms after stimulus onset (N400) (Kutas and Hillyard, 1980). Contrary to oscillatory activity that can be either phase-locked to the stimulation or not, evoked potentials reflect neural activity that is time-locked to the stimulation (Cohen, 2014). Several mechanisms have been proposed for the physiological origin of evoked potentials. The "signal plus noise" hypothesis proposes that evoked responses are generated by a supplementary time-locked signal superimposed on top of oscillatory activity (Munck and Bijma, 2010). The "phase reset" hypothesis proposes that evoked potentials result from an alignment of the ongoing oscillations (Makeig et al., 2002). Finally, the "amplitude asymmetry" hypothesis proposes that the amplitude of oscillations is not centered around zero, but around a time-varying off-set called baseline-shift. Evoked responses would be the result of this baseline-shift in oscillatory amplitude,

in particular in the alpha range (Munck and Bijma, 2010).

To sum-up, EEG measures extra-cellular field potentials created by large populations of synchronized active neurons. As parallel currents in opposite directions cancel out, active neurons have to be spatially aligned within these populations and produce radial current dipoles to contribute to the EEG.

2.2.2 EEG in infants

Electro-encephalography of the human infant presents specificities that should be taken into account when interpreting data. Oscillatory as well as ERP components are far from being mature in infants, both in terms of latency and elements to be observed.

Some specificity of the infant EEG arises from the properties of the infant head. Due to a reduced head size as compared to adults (34 cm at birth, 42 at 6 months and 45 at 12 months on average; World Health Organization, 2018), a reduced number of electrodes have to be used. Some studies use 9 to 15 electrodes (e.g Háden et al., 2013; Németh et al., 2015), although other studies have used set-ups with higher electrode density with up to 32 electrodes in newborns (Mahmoudzadeh et al., 2017), and even 128 electrodes in older infants (Kouider et al., 2013; Von Holzen et al., 2018). Another important specificity of infants is the presence of fontanels in the skull. Fontanels are membranes between the skull bones that ossify during the first year of life. These membranes have different electric properties than the skull, and as such introduce inhomogeneity in skull conductivity, which impact the EEG signal. Specifically, fontanels have higher conductivity than bone (Lew et al., 2013; Gargiulo et al., 2015). As a result, field potentials are estimated with higher amplitude above the fontanels than above bone (Lew et al., 2013). However, these results are tempered by the fact that the neonatal skull has higher electrical conductivity than the adult one, close to soft tissues (Odabaee et al., 2014). For this reason the difference in conductivity between bone and fontanels is small, and the effect of the fontanels on the EEG signal is negligible for classical electrophysiological analysis, both in terms of topography and amplitude (Gargiulo et al., 2015). Nevertheless, as compared to adults, EEG in infants is more spatially focal due to higher skull conductivity than in adults (Odabaee et al., 2014).

In newborns, due to the limited time participants can stay awake, cognitive experiments are typically performed during sleep. At term birth, sleep is composed of an alternation of two stages, called active and quiet sleep (e.g. Ellingson and Peters, 1980). Quiet sleep is characterized by activity mainly in the theta band. Active sleep is characterized by lower amplitude activity with higher frequency (Anders et al., 1971). Hence, at term birth, oscillatory activity occurs mainly in the theta and delta bands at rest. This activity changes almost week to week, with less and less active sleep over the course of the first months of life (Husain, 2005). Although no significant effect of sleep stages was found on the MMN in neonates (Martynova et al., 2003), another study with two-month-old infants provided ev-

idence that sleep stage potentially influences evoked responses (Friederici et al., 2002). Therefore, the effect of sleep on evoked responses in infants remains poorly defined, but background oscillatory activity can potentially interfere with induced stimulus-related activity.

A key question is whether the typical evoked responses observed in adults can be observed in infants. Regarding ERPs, newborn experiments have typically looked at mismatch response in response to deviant stimuli. MMN response is observed from birth with its mature amplitude, but increased latency that decreases with age (Cheour et al., 1998). Regarding oscillatory components, speech sounds evoke an increase in power in a 10-20 Hz frequency band in premature newborns (Mahmoudzadeh et al., 2017). Later in development, in full-term newborns, deviant tones induce an increase in power in the theta (4-8 Hz), beta (13-25 Hz) and gamma bands (18-45 Hz) (Isler et al., 2012; Stefanics et al., 2007). Similarly, induced power is observed in the typical gamma range (32-48 Hz) at 6 months in response to visual objects, but appears later and smeared out over long time intervals (240 to 320 ms post stimulus onset). In contrast, 8-month-old infants show earlier (from 100 ms post stimulus onset), temporally more precise burst-like activity, more similar to adults (Csibra et al., 2000). Taken together, these results suggest that oscillatory activity evoked by sensory stimuli becomes more and more spectrally and temporally precise during the first months of life. The central frequency values of the different oscillation bands also change across development. As compared to moments of sustained attention from the infant, periods of darkness in the testing room (used as functional analogues to eye closing when the participants cannot follow instructions), elicited functional analogues of the alpha rhythm centered around 6 Hz at 8 months, and around 7 Hz at 11 months (Stroganova et al., 1999). This suggests that the frequency ranges associated with specific sensory and cognitive states are already present from birth, although they might be slower, less spectrally and temporally precise than the ones observed in adults.

2.2.3 Limitations of the EEG method

EEG has several limitations, both due to its internal properties and those of the brain tissues.

First, EEG records only slow electrophysiological activity. The magnitude of the EEG signal is inversely related to the frequency: the higher the frequency of oscillation, the lower the power. This, combined with the fact that surrounding head tissues, especially the skull, have high electrical resistance, therefore acting as a low-pass filter, makes it extremely difficult to record high-frequency activity (superior to 80 Hz) at the scalp surface. Because spikes occur at a high rate ($>$ 300 Hz), it is not possible to record them with scalp EEG. Therefore, only a subset of neuronal activity is represented in the EEG signal.

Another limitation of the EEG technique is that only a subset of neuronal populations in the cortex can be recorded. The signal originates from the superficial

layers of the cortex, contributions of deeper brain structures are very difficult to record (Cohen, 2014). Only radial sources contribute to the EEG signal. These sources are cortical populations with axons oriented radially to the scalp, i.e. cortical gyri. On the opposite, sources oriented tangentially to the scalp, i.e. cortical sulci, tend to cancel out and therefore don't contribute to the EEG signal (Cohen, 2014; Luck, 2014).

A last limitation of EEG is its low spatial resolution. Even before reaching the head surface, the signal arriving at the cortical surface potentially results from a weighted sum of activity at multiple depths (cortical and sub-cortical). Then the skull diffuses electrical signals. Signals from different cortical areas can overlap when smeared by the skull. Therefore the electric potential measured at each electrode is a weighted sum of different underlying currents mixed at the skull. Three parameters influence the EEG signal: the sources where neural activity takes place, the transmission properties of the head tissues (e.g. cerebrospinal fluid, cerebral tissues, bone, scalp), and the EEG sensors themselves (e.g. impedance, sampling rate). The electric field produced at the scalp surface by a set of known sources (i.e. the direct problem) can be estimated using conductivity models. On the opposite, localizing the sources of a given surface electric field, known as the inverse problem, is more challenging. An infinite number of source configurations can produce any given voltage distribution on the scalp (Nunez et al., 2006; Plonsey and Heppner, 1967). Methods exist to estimate the sources of the EEG signal, but they require a large number of electrodes and rely on heavy assumptions about head anatomy, hence give large margin of errors (Diallo, 2017).

All these limitations point to the fact that EEG has complementary properties to other techniques, in particular those with higher spatial resolution, such as fMRI and fNIRS for developmental populations. This work uses both EEG and fNIRS, to characterize electrophysiological mechanisms supporting invariant perception of speech with EEG, and localize cortical regions involved with fNIRS.

Part II

Experimental contributions

Chapter 3

Adult-like perception of time-compressed speech at birth

This chapter is a verbatim of Issard & Gervain (2017)

3.1 Introduction

Understanding speech is remarkably constant: despite considerable differences in voice quality, accent or speech rates between speakers, we have the subjective impression of hearing the same speech sounds and words under a wide variety of circumstances. Indeed, our auditory and linguistic systems readily normalize the highly variable speech signal in order to extract linguistic units that are necessary for comprehension. One important dimension along which speech may vary considerably is time. Speech rate differs within and across speakers, but this typically doesn't impede communication.

Indeed, successful adaptation to time-compressed speech has been observed in adults and older children, in tasks such as word comprehension (Dupoux and Green, 1997; Orchik and Oelschlaeger, 1977), sentence comprehension (e.g. Ahissar et al., 2001; Peelle et al., 2004) or reporting syllables (Mehler et al., 1993; Pallier et al., 1998; Sebastián-Gallés et al., 2000). In these latter studies, participants had to report the words or syllables they perceived in sentences compressed to 38%–40% of their initial duration, after having been habituated with compressed speech or after having received no habituation. Specifically, using 40% compression, Mehler et al. (1993) presented French or English sentences to French or English monolinguals and to French-English bilinguals. Participants reported higher numbers of words when they were initially habituated to and then tested in their native language. A follow-up study showed that Spanish-Catalan bilinguals adapted to Spanish or Catalan sentences compressed at 38% after habituation in the other language (i.e. habituation in Spanish before test in Catalan, and habituation in Catalan before test in Spanish). In two subsequent studies, English monolingual adults tested with 40% compressed English sentences benefited from habituation to time-compressed speech in Dutch, which is rhythmically similar to English,

and Spanish monolinguals tested with 38% compressed Spanish sentences benefited from habituation with Catalan and Greek, languages that shares rhythmic properties with Spanish (Pallier et al., 1998; Sebastián-Gallés et al., 2000). These studies thus show that adults are able to adapt to moderately compressed speech in their native language, as well as in unfamiliar languages, if those belong to the same rhythmic class as their native language. This suggests that adaption to time-compressed speech is based on auditory/phonological mechanisms ('sound-based' adaptation), rather than on top-down linguistic knowledge regarding the lexicon, the syntax or semantics of the native language ('lexical/grammatical' adaptation).

Neuroimaging studies also provided evidence for a dissociation between lexical/grammatical processing and sound-based adaptation to time compressed speech, the two processes involving different neural pathways. In an fMRI study, Peelle et al. (2004) presented syntactically simple or complex sentences compressed to 80%, 65%, or 50% of their normal duration. Time-compressed sentences produced activation in the anterior cingulate, the striatum, the premotor cortex, and portions of temporal cortex, regardless of syntactic complexity. Other studies found that some brain regions, e.g. the Heschl's gyrus (Nourski et al., 2009) and the neighboring sectors of the superior temporal gyrus (Vagharchakian et al., 2012), showed a pattern of activity that followed the temporal envelop of compressed speech, even when linguistic comprehension broke down, e.g. at 20% compression rate. Other brain areas, such as the anterior part of the superior temporal sulcus, by contrast, showed a constant response, not locked to the compression rate of the speech signal for levels of compression that were intelligible (40%, 60%, 80% and 100% compression), but ceased to respond for compression levels that were no longer understandable, i.e. 20% (Vagharchakian et al., 2012).

These studies investigated adults and older children, i.e. participants who are proficient speakers of a language and have considerable experience with speech and language processing in general. Thus, the developmental origins and the existence of a critical period for this ability remain unexplored. It is unclear if adaptation to time-compressed speech can occur independently of any experience with speech (at least with broadcast speech, transmitted through the air, as experienced postnatally). Several hypotheses may be considered. First, this ability might rely on top-down linguistic knowledge of the lexicon and/or the grammar (morphology, syntax, semantics), which helps listeners discover linguistically relevant constants in the time-altered speech stream. In this case, newborns and young infants should fail to adapt to time-compressed speech. This hypothesis is relatively unlikely, since adults and older children can adapt to compressed speech in an unknown language, as long as that is rhythmically similar to their mother tongue. Second, one may assume that the adaptation ability depends on experience with spoken language in general (not specifically with the native language). Such a hypothesis predicts that adults and children can adapt to time-compressed speech, as has been observed, but that newborns, who only have experience with speech as heard in utero, which is very different from regular speech transmitted through the air, would fail. Third, it may be the case that little or no experience

is needed for the adaptation ability to occur, so the degraded, low-pass filtered speech signal experienced prenatally, which only preserves the prosodic properties of the native language (Gerhardt et al., 1992; Querleu et al., 1988), may be sufficient. In this case, newborns may adapt to time-compressed speech successfully. Here, we show that this is indeed the case, suggesting that adaptation occurs at the auditory/phonological and not at the lexical/grammatical levels. This finding brings the first developmental evidence to the hypothesis that adaptation to time-compressed speech is an auditory/phonological phenomenon.

Specifically, using near-infrared spectroscopy (NIRS) we tested the ability of the newborn brain to discriminate and to adapt to speech at three levels of compression: (i) normal, uncompressed speech, i.e. 100% of its original duration, (ii) speech compressed at a level comprehensible for adults, i.e. 60% of its initial duration, and (iii) speech compressed at a level that is no longer comprehensible for adults, i.e. 30% of its original duration. NIRS is a powerful and easy-to-use neuroimaging method, well suited to test young infants (Gervain et al., 2011; Rossi et al., 2012). It uses the absorbance properties of oxygenated and deoxygenated hemoglobin to assess the metabolic correlates of brain activity, i.e. the hemodynamic response function (HRF).

3.2 Materials and methods

3.2.1 Participants

Fifty-nine healthy full-term neonates participated in the experiment (mean age 2.34 days, range 1–4 days; 34 females). They were recruited during their stay at the hospital after birth. To participate, newborns had to be full-term (gestational weeks $\geq$37), weight more than 2700 g, and have an Apgar score superior or equal to 8 at 10 min after birth. Newborns' hearing was assessed by a measurement of their oto-acoustic emissions during their stay at the maternity and through screening by a local pediatrician: no hearing disabilities, neurological disorders, prenatal or perinatal complications were reported for any of the participants. A questionnaire filled out by the parents assessed for parental handedness, antecedents of language or hearing impairments in the family. One parent reported a non-hereditary auditory impairment (mother having a deaf ear following an ear infection). No other relevant condition was present.

Four participants were excluded from the analysis due to crying (n = 3) and technical problems (n = 1). Data from remaining fifty-five newborns were preprocessed, but thirty-four were not retained for the final analysis due to poor data quality (mostly because of movement artifacts and noise related to dark thick hair). To be retained for the final analysis, participants had to have at least 50% good data in at least two of the three conditions. Importantly, this uniform rejection criterion was applied in batch to all infants whose data was pre-processed, leading to the above reported rejection, prior to statistical analysis. Twenty-one participants were thus included in the final analysis.

All parents gave informed written consent. The study was approved by the local ethics committee at Paris Descartes University (CERES number 2011-013).

3.2.2 Material

The stimuli were 120 utterances of 11 or 12 syllables, randomly chosen from the CHILDES (MacWhinney, 2000) French corpora and assessed for grammaticality and naturalness by a native French speaker. A native female French speaker recorded the selected utterances in an infant-directed manner. To generate time-compressed speech, we compressed the original utterances to 60% and 30% of their initial duration. The two compression rates, 60% and 30%, were chosen because adults perceive them differently. Indeed, speech compressed at 60% is intelligible for adults, while speech compressed at 30% is not (Pallier et al., 1998). For compression, we used the Audacity software, which allows for the correction of the pitch shift due to compression, maintaining the pitch at the same level as in the original recordings. The compression algorithm (implemented by the "change tempo" function of Audacity) maintained the spectral content of the normal utterances but accelerated the amplitude modulations by a factor of 1.67 for the 60% compression rate and by 3.33 for the 30% compression rate (3.1). Intensity was normalized across stimuli. The utterances were finally concatenated into blocks (see below). Within blocks, utterances were separated by silences of 0.5–1.2 s.

The experiment used an alternating/non-alternating design, often used in behavioral studies to test fine-grained perceptual discrimination (Best and Jones, 1998; Gervain et al., 2014), and successfully implemented with NIRS, (e.g. Sato et al., 2010; Gervain et al., 2012). Two types of blocks were used (3.2): alternating blocks contained an alternation of normal and accelerated speech tokens (e.g. 100% and 60% or 100% and 30%), whereas non-alternating blocks contained only one type of speech sounds (only normal speech utterances, only 60% compressed utterances, or only 30% compressed speech). The 30% and 60% compression rates were used in two distinct halves of the experiment (3.2). The order of the 30% and 60% compression halves was counterbalanced across participants. The advantage of the alternating/non-alternating design, e.g. over presenting only non-alternating blocks, is that it can address two different questions. When analyzing non-alternating blocks only, we can explore how and where speech at different compression rates is processed in the newborn brain. Additionally, comparing alternating and non-alternating blocks tells us whether newborns can discriminate between the three different levels of compression.

The number of utterances per block was adjusted such that all five block types (non-alternating 100%, non-alternating 60%, non-alternating 30%, alternating 100% and 60%, and alternating 100% and 30%) had approximately the same length (mean duration 18,39 s, range 17–19 s). This choice was made due to the sensitivity of the hemodynamic response function to the length and absolute amount of stimulation. Non-alternating normal blocks thus contained six utterances, 60% compressed non-alternating blocks contained 8 utterances, and

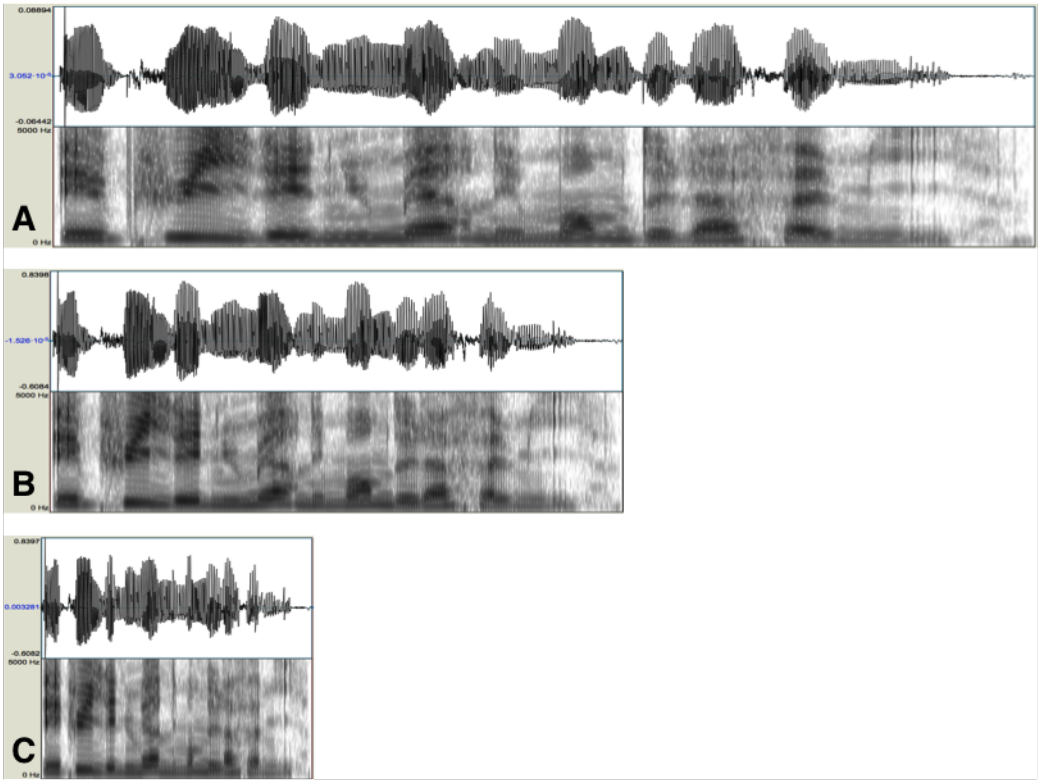


Figure 3.1: Waveforms and spectrograms of one sentence at its normal duration (A), compressed to 60% (B), and to 30% (C) of its original duration.

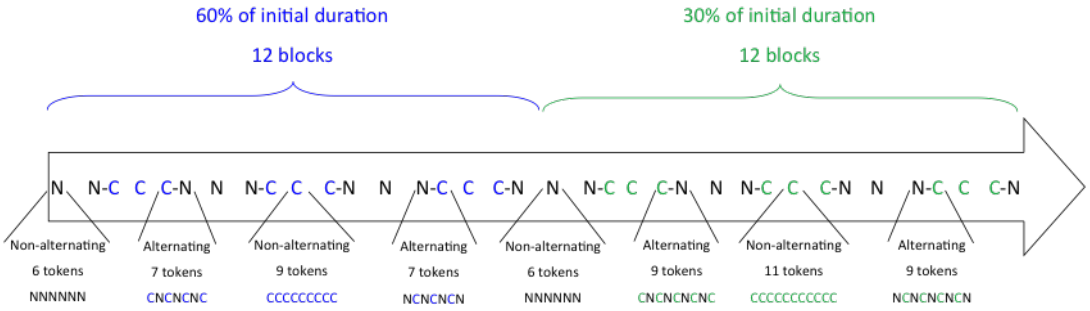


Figure 3.2: Experimental design. N: Normal sentences. C: Compressed sentences. All the blocks had the same duration.

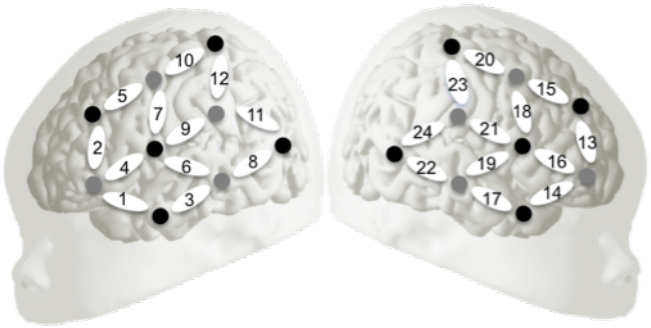


Figure 3.3: Optode placement overlaid on the schematic newborn head. Dark circles indicate detectors, light circles indicate sources and numbered ovals correspond to measurement channels.

30% compressed non-alternating blocks contained 11 utterances. Blocks alternating between normal and 60% compressed tokens contained 7 utterances and those alternating between normal and 30% compressed tokens contained 9 utterances. Each utterance was presented once per compression rate. The order of the stimuli was randomized within each condition. Blocks were separated by pauses varying in duration between 26 and 35 s.

3.2.3 Procedure

The hemodynamic response of the newborn brain to the auditory stimuli was recorded using a NIRx NIRScout 816 (NIRx Medizintechnik GmbH, Berlin, Germany) NIRS machine (two pulsated wavelengths of 760 nm and 850 nm). LED light sources and detectors were placed on a stretchy cap, each adjacent source-detector pair forming a channel (source–detector separation: 3 cm). The relative concentration of oxy- and deoxyhemoglobin was computed in each channel based on the difference between the intensities of the incident light projected onto the head and the light measured by the detectors, using a modified Beer-Lambert law (for further details, see Gervain et al., 2011). In the apparatus we used, the optodes were placed on the fronto-temporal, temporal and temporo-parietal regions (3.2.3). This localization was verified against an average newborn MRI head template (Shi et al., 2011) and was already used successfully in previous work from our laboratory (Abboub et al., 2016).

Infants were tested in a quiet room of the hospital, lying in their cribs throughout the 22–25-min testing session, assisted by an experimenter. Parents attended the session. Testing was done with the infants in a state of quiet rest or sleep. The stimuli were presented through two loudspeakers positioned at a distance of approximately 1 m and an angle of 30° from the infant's head respectively. The stimulus presentation was controlled by E-Prime 2.10. The experiment was dis-

continued if the infant started to cry or upon parental request.

The experiment was divided into two parts: in half of the experiment, we used 60% as the compression rate, in the other half 30%. Each of these two parts contained 12 blocks. The order of the two parts was counterbalanced between participants. The alternating and non-alternating blocks strictly alternated, with half of the infants hearing an alternating block first and the other half hearing a non-alternating block first. If the first non-alternating block was a normal one, the second one was a time-compressed one and vice-versa. Likewise if the first alternating block began with a normal utterance, the second one began with a time-compressed utterance (3.2).

3.2.4 Data processing and analysis

Data were processed using Matlab (Mathworks) custom scripts developed by the McDonnell consortium "Infant Methodology" (Gervain et al., 2011). First, attenuation changes were converted into changes of concentration in oxy- and deoxyhemoglobin. Data were then filtered to remove components of the signal due to heartbeat, as well as to remove noise, general trends and systemic blood flow variations, using a band pass filter between 0.01 Hz and 1 Hz.

Artifacts (mainly due to movements) were then removed according to the following criteria: each block-channel pair containing a concentration change over 0.1 mmol × mm on two consecutive samples (i.e. 200 ms) was removed from the analysis. Only participants who had at least 50% artifact-free blocks were kept (see 3.2.1 above).

We then computed the mean concentration change in a time window starting 7 s after stimulus onset (to allow for adaptation to take place), and lasting 25 s (comprising the stimulus block starting 7 s after its onset as well as a post-stimulus period of 14 s) in each channel for oxy- and deoxyhemoglobin in each experimental condition. All analyses were carried out for both hemoglobin species, but only oxyhemoglobin yielded significant results, as is common with infant studies (Gervain et al., 2011). We thus only report statistical tests for oxyhemoglobin.

We first compared each condition to a zero baseline with permutation tests (Maris and Oostenveld, 2007; Nichols and Holmes, 2002). Permutation tests have the advantage of controlling for the multiple comparisons problem without loss of statistical power, which typically occurs when Bonferroni or other corrections are applied to infant NIRS data (Nichols and Holmes, 2002). The experimental conditions were then compared directly by running permutation tests. Alternating and non-alternating blocks were compared for each compression rate (60% and 30%) separately.

We further analyzed our results using analyses of variance (ANOVAs) looking at specific regions of interest (ROIs). We defined two ROIs per hemisphere, following previous NIRS studies on newborns (Gervain et al., 2012, 2008), as well as the pattern of activity found in response to time-compressed speech in previous imaging studies (Adank and Devlin, 2010; Vagharchakian et al., 2012). We thus

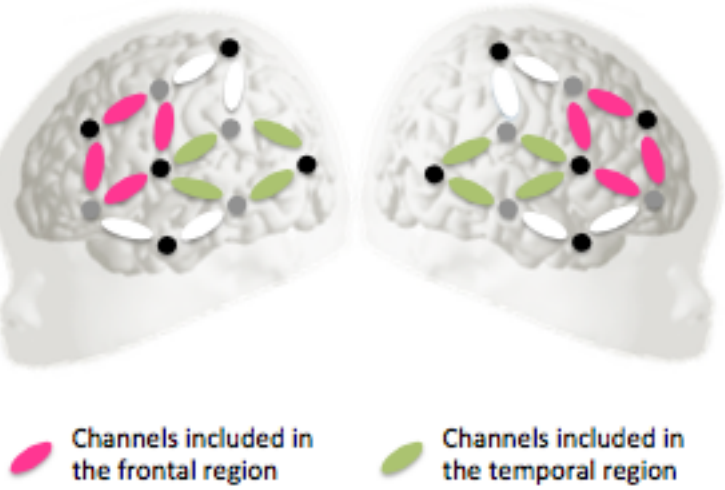


Figure 3.4: Channels included in the two regions of interest. Dark circles indicate detectors, light circles indicate sources.

defined a temporal region comprising channels 6, 8, 9 and 11 in the left hemisphere and channels 19, 21, 22 and 24 in the right hemisphere, and a frontal region comprising channels 2, 4, 5 and 7 in the left hemisphere and channels 13, 15, 16 and 18 in the right hemisphere. These regions are plotted in 3.4.

3.3 Results

3.3.1 Channel-by-channel comparisons

Non-alternating blocks

We first analyzed the non-alternating blocks alone. The obtained grand average responses are shown in 3.3.1. The 30% compression non-alternating blocks evoked a significant decrease in oxyhemoglobin as compared to baseline in channels 10, 12, 15, and 22 ($p10 < 0.001$, $p12 = 0.025$, $p15 < 0.001$, $p22 < 0.001$, Fig. 6A). No significant channel-by-channel results were obtained for the other two compression rates. The results are summarized in 3.1 and 3.3.1A (only the significant comparisons are shown).

Directly comparing the three compression rates in channel-by-channel permutation tests, we observed a significant effect of compression rate (normal/60%/30%) in channel 10 ($p = 0.002$). Pair-wise comparisons with permutations to explore this effect yielded a significant difference between 60% and 30% compressed speech in channel 10 ($p = 0.011$), and between normal and 30% compressed speech in channels 10 ($p = 0.001$), 12 ($p = 0.020$), 15 ($p = 0.002$), and 22 ($p = 0.007$), but no significant difference between normal and 60% compressed speech. These results are summarized in 3.2 and in 3.3.1B–C.

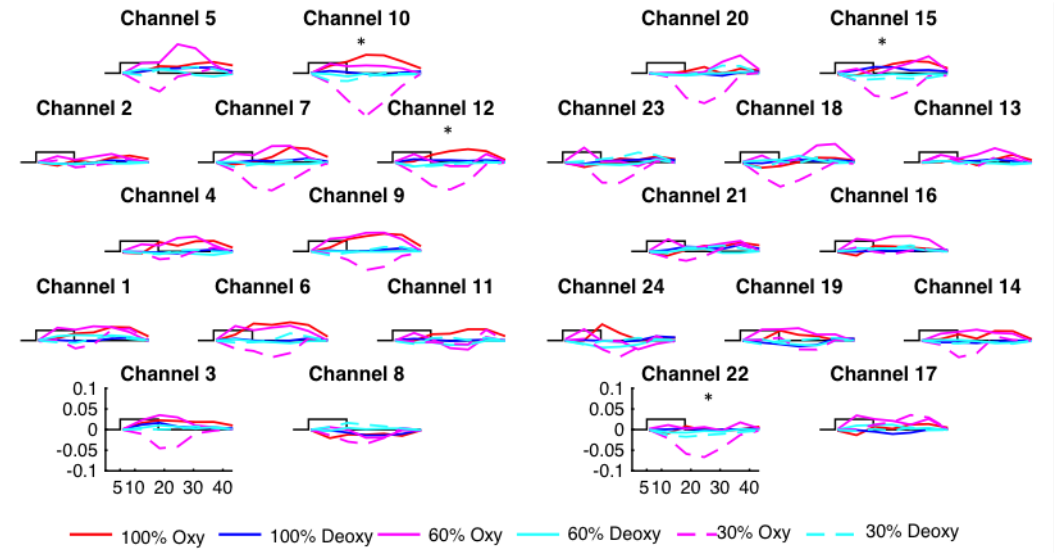


Figure 3.5: Grand averages of the hemodynamic response evoked by each condition in each channel in the non-alternating blocks (mmol x mm). The rectangle represents the time of stimulation. Asterisks indicate significant comparisons for the 30% compression rate with respect to baseline.

	results
normal	n.s.
60% compression rate	n.s.
30% compression rate	Ch. 10: p<0.001 Ch. 12: p=0.025 Ch. 15: p<0.001 Ch. 22: p<0.001

Table 3.1: Statistical comparisons of the three different types of non-alternating blocks to baseline

main effect of compression rate		Ch. 10: p=0.002
pairwise comparisons	normal vs. 60%	n.s.
	normal vs. 30%	Ch. 10: p=0.001 Ch. 12: p=0.020 Ch. 15: p=0.002 Ch. 22: p=0.007
	60% vs. 30%	Ch. 10: p<0.011

Table 3.2: Statistical comparisons of the three types of non-alternating blocks to each other

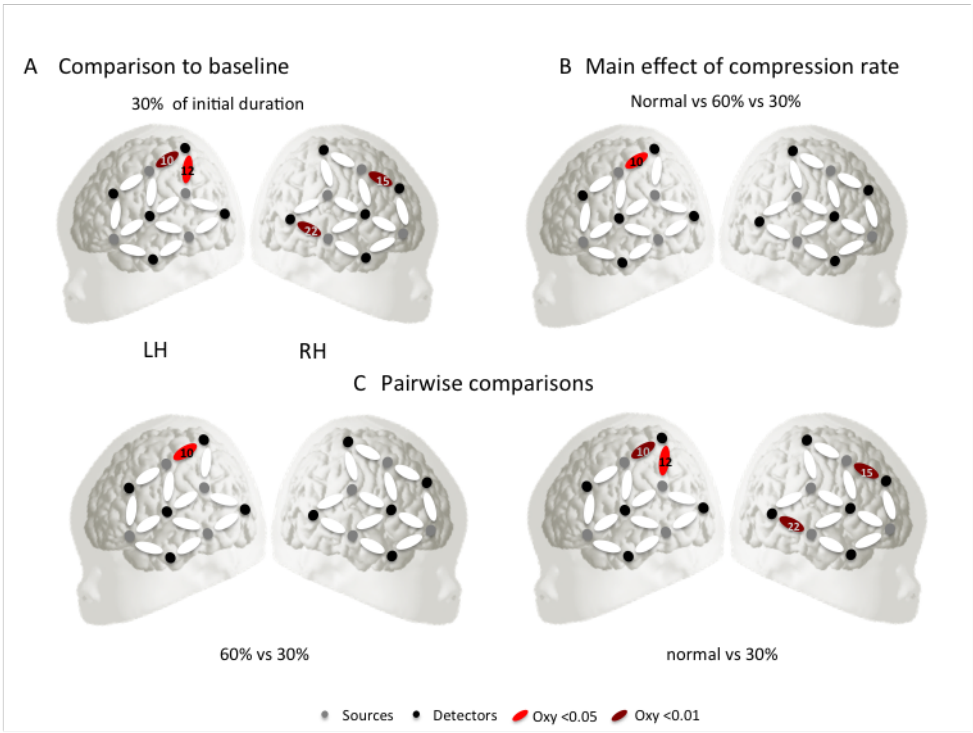


Figure 3.6: Statistical maps for the non-alternating blocks. Only comparisons with at least one significant channel are shown. (A) comparison of each condition to baseline, (B) direct comparison of the three conditions, (C) pairwise comparisons of the conditions.

alternating blocks vs. baseline	60%	Ch. 24: p=0.0396
	30%	n.s.
non-alternating vs. baseline	60%	n.s.
	30%	n.s.
alternating vs. non-alternating	60%	n.s.
	30%	n.s.

Table 3.3: Statistical comparisons for the alternating and non-alternating blocks to baseline and to each other

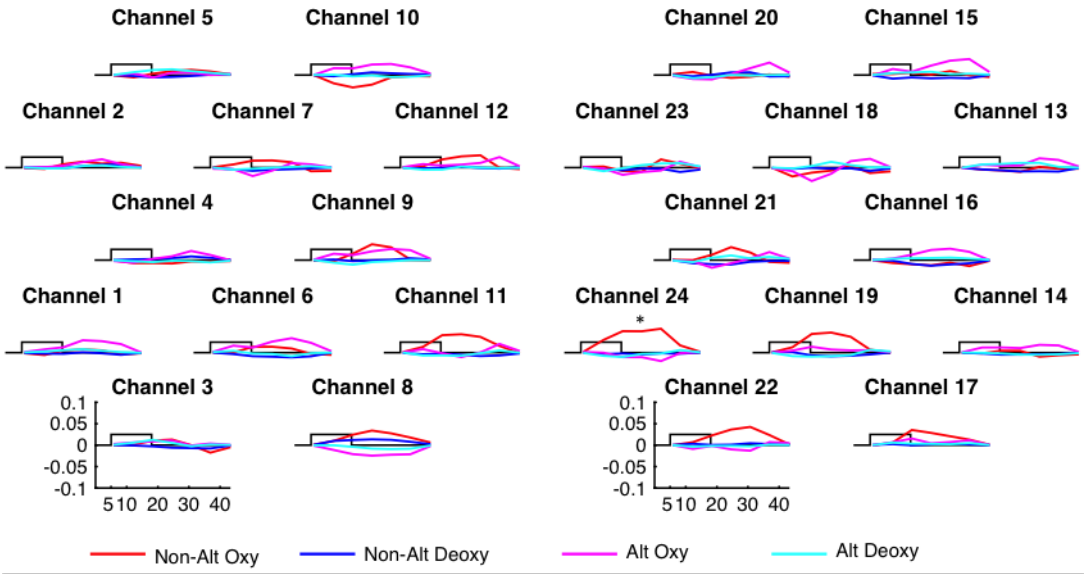


Figure 3.7: Grand averages of the hemodynamic responses to the alternating and non- alternating blocks for the 60% compression rate. The rectangle represents the time of stimulation. The asterisk indicates a significant difference between alternating blocks and baseline.

Comparison between alternating and non-alternating blocks

We also compared the hemodynamic response evoked by alternating and non-alternating blocks to the baseline and to each other separately in each channel for each of the compression rates.

For the 60% compression rate, there was a significant increase in oxyhemoglobin concentrations in channel 24 ($p = 0.0396$) for the alternating blocks as compared to baseline. For the 30% compression level, we did not obtain any significant difference between the two.

We did not obtain any significant difference between alternating and non-alternating blocks in channel-by-channel comparisons for either compression rate.

The above results are summarized in Table 3.3.

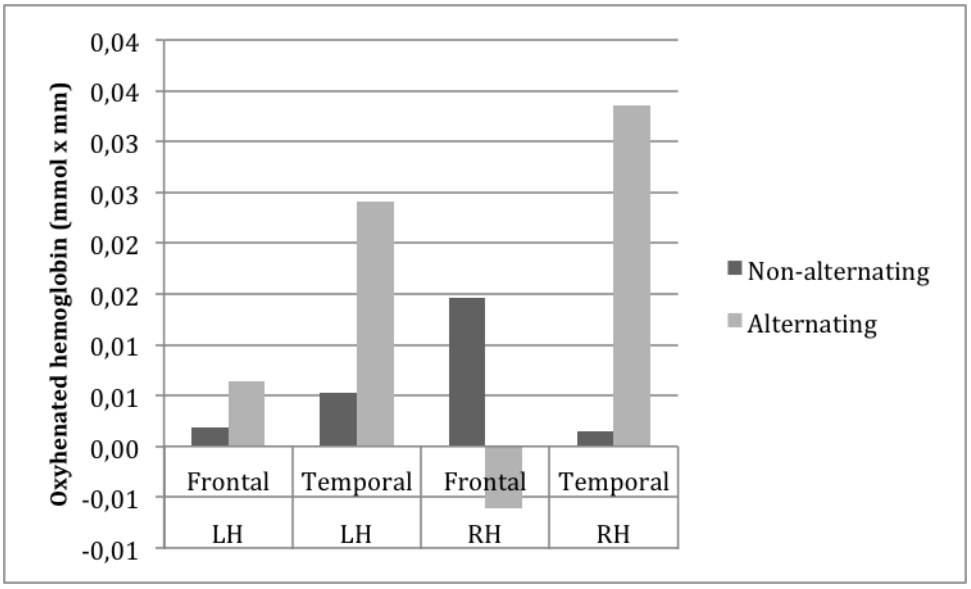


Figure 3.8: Analysis by ROI for 60%-compressed speech.

3.3.2 Analyses of variance

In addition to the above channel-by-channel comparisons, we also conducted analyses of variance (ANOVA) to reveal general patterns in two regions of interests, temporal and frontal, defined on the basis of previous results on the perception of time-compressed speech (Figure 3.4).

Non-alternating blocks

An ANOVA with factors Condition (normal/60%/30%), Hemisphere (LH/RH), and ROI (temporal/frontal) as within-subject factors over the three types of non-alternating blocks did not revealed any significant effects or interactions.

Alternating vs. non-alternating blocks

For the 60% compressed part of the experiment, an ANOVA with Block Type (alternating/non- alternating), ROI (frontal/temporal) and Hemisphere (LH/RH) as within-subject factors revealed a main effect of ROI ($F(1, 20) = 7.719$, $p = 0.014$), and an interaction between Block Type and ROI ($F(1, 20) = 8.069$, $p = 0.010$). As depicted in Figure 3.8, the main effect of ROI was due to a higher activity in the temporal regions ($d_{temporal-frontal} = 1.18 \times 10 - 2, p = 0.014$). The interaction between Block Type and ROI was due to the fact that non-alternating blocks evoked more activity in the frontal regions, whereas alternating blocks evoked more activity in the temporal region ($d_{temporal-frontal} = 2.86 \times 10 - 2, p = 0.007$)(Figure3.8). The three-way interaction between Block Type, ROI and Hemisphere was not significant.

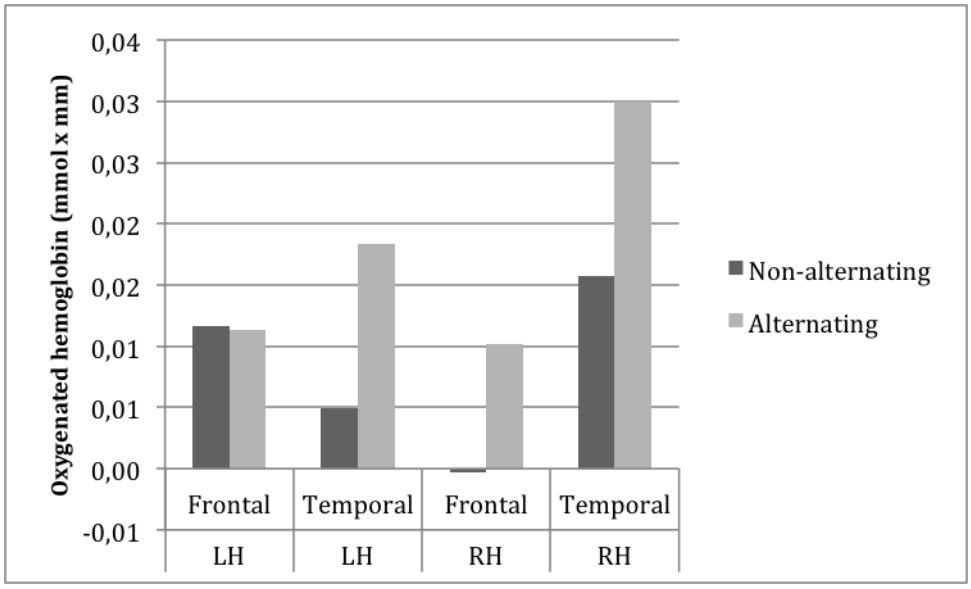


Figure 3.9: Analysis by ROI for 30%-compressed speech.

The same ANOVA on the 30% compressed part of the experiment showed a main effect of ROI ($F(2, 20) = 4.391$, $p = 0.049$) and an interaction between Hemisphere and ROI ($F(2, 20) = 7.807$, $p = 0.011$), but no effect of the type of block (Figure 3.3.2). The main effect of ROI was due to greater activation in the temporal region than in the frontal region ($d_{temporal-frontal} = 8.91 \times 10 - 3, p = 0.049$, Figure 3.3.2). This effect was modulated by Hemisphere: In the frontal region higher activity was observed in the left hemisphere, whereas the opposite pattern was observed for the temporal region (Figure 3.3.2). The three-way interaction between Block Type, ROI and Hemisphere was not significant.

3.4 Discussion

In the current study, we have tested how the newborn brain perceives time-compressed speech. We compared normal speech with speech compressed to 60% and to 30% of its original duration. The former, moderately compressed speech rate is intelligible for adults, the latter, higher compression is not. We presented the stimuli using an alternating/non-alternating design and measured newborns' brain responses using NIRS.

When comparing the neural responses to the three speech rates as presented separately in the non-alternating conditions, we have found that the fastest, unintelligible speech rate gave rise to a strong deactivation (or inverted hemodynamic response), while the response to the two intelligible rates, i.e. the normal and the moderately compressed speech rates, was weak, but canonical. This suggests that the newborn brain processes moderately compressed speech in a similar fashion than normal speech, but responds to highly compressed speech differently.

This pattern of results resembles the intelligibility performance of adult listeners, who can adapt to moderately compressed, but not to highly compressed speech. Furthermore, the comparison of alternating and non-alternating blocks suggests that while the newborn brain processes normal and moderately compressed speech similarly, it is nevertheless able to discriminate between the two.

Several aspects of our results warrant further discussion before the more general implications of the findings can be considered. First, a difference between alternating and non-alternating blocks was found for the 60% compression rate, but not for the 30% compression rate alone. This can be explained when considering the pattern of responses obtained in the non-alternating blocks. Normal speech and moderately compressed speech evoked a mild canonical response, whereas the fastest speech rate evoked a strong inverted response. In the direct comparison of the alternating/non-alternating blocks, data from normal and time-compressed non-alternating blocks were averaged together before being compared to the alternating blocks. The response evoked by the normal and 30% compressed blocks averaged each other out, masking the bimodal distribution of the responses in non-alternating blocks. This made it impossible to show a difference between the non-alternating and alternating blocks in the 30% compressed condition. By contrast, in the moderately compressed 60% speech rate, the normal and the 60% compressed non-alternating blocks evoked a similar, canonical hemodynamic response, so their mean yielded a non-null, canonical response. This pattern of results is in line with the observation that normal and moderately compressed speech rates are encoded the same way by the newborn brain, while the highest compression rate is treated differently.

We hypothesize that the relative weakness of the canonical response to the normal and the 60% compression rates is due to habituation to the stimuli across time. This habituation can be quite fast, and may happen already within the first block. While habituation across blocks is a commonly used measure in NIRS (e.g. Benavides-Varela et al., 2012; Bouchon et al., 2015), habituation within a stimulation block is hard to detect with NIRS. Existing fMRI data on adults show that adaptation to time-compressed speech occurs within the first 16 sentences (Adank and Devlin, 2010), and this number is already reached by the first alternating/non-alternating block pair. Newborns therefore might have been habituated to 60% compressed speech during the first pair of blocks, leading to reduced amplitudes between the hemodynamic responses evoked by normal and 60% compressed speech blocks. This might explain why these responses are not significantly different from baseline. Such repetition suppression effects are often taken to be neural signatures of the existence of a stable representation of the signal (see Nordt et al., 2016, for a review). Given newborn infants' prenatal experience with their native language as well as their broad-based, universal speech perception abilities, it is not implausible to assume that they may build a stable representation of normal and moderately time-compressed speech fast and efficiently. Our experimental design was not set up to specifically test the time course of adaptation itself. A more detailed description of this process is nevertheless theoretically relevant, and will

require further research.

Second, in our design, different block types contained different numbers of utterances. This choice was made to equate for the absolute amount of stimulation across block types, as this is known to influence the hemodynamic response, especially in young infants (Minagawa-Kawai et al., 2013). However, one might argue that the differences we observed across conditions might be due to this difference in the number of utterances, and not to compression rate itself. In particular, it may be the case that the inverted response observed in the fastest speech rate may be a deactivation or neural habituation effect due to the higher number of repetitions in that condition. Indeed, considerable redundancy in the stimuli has been shown to give rise to habituation effects in the NIRS response of newborns (Bouchon et al., 2015). While we did not explicitly test this alternative, there is indirect evidence in our data that this explanation is unlikely. While the non-alternating 30% compression blocks did indeed contain the greatest number of utterances (Figure 3.2), alternating blocks using this compression rate also had a much higher number of utterances than non-alternating normal blocks or alternating blocks with 60% compression rate. Yet, they did not evoke an inverted response. Furthermore, the number of utterances in the different block types ranged from 6 through 7, 8 and 9 to 11. This should have resulted in graded NIRS responses for the different block types if the number of utterances played a role, which was not the case.

More generally, and returning to our research question about the origin of the mechanisms underlying adaptation to compressed speech, we had evoked several potential hypotheses: adaptation might happen at the level of (i) knowledge about the lexicon and/or grammar of the native language, (ii) processing of broadcast speech or (iii) general auditory processing requiring no experience with broadcast speech. Indeed, in the adult literature, the drop of performance at high compression ratios has been explained by several models: (i) the saturation of the lexical buffer, which gets filled up at speech rates faster than the inherent speed of linguistic read-off and processing (Vagharchakian et al., 2012), (ii) an impossibility to map phonological representations to articulatory motor plans (Adank and Devlin, 2010; Peelle et al., 2004), or (iii) a failure to find the subphonemic and rhythmic landmarks in the compressed speech signal (Pallier et al., 1998; Sebastián-Gallés et al., 2000).

Our developmental results support the third hypothesis for at least two reasons. First, despite their sophisticated abilities to process speech sounds, newborns do not have sufficient lexical or grammatical knowledge of their native language and lack experience with the processing of broadcast speech. While lexical mechanisms are certainly involved in adaptation to time-compressed speech in adults, they cannot constitute the core mechanism in newborns. Even if newborns have access to proto-lexical information, e.g. via statistical learning (Teinonen et al., 2009) or sensitivity to certain phonological features (Shi et al., 2011), they still lack a sizeable lexicon, so an explanation based on the saturation of a lexical buffer, as proposed by Vagharchakian et al. (2012), is unlikely.

Second, in adults, highly compressed and normal speech elicit different acti-

vations in the temporal cortex. Although in our study, we cannot localize the source of activation with as much precision as in previous fMRI studies, it seems that we reproduce the posterior temporal activation found in adults (Adank and Devlin, 2010; Vagharchakian et al., 2012), as shown by the higher activation in the temporal/temporo-parietal regions for both compression rates (Figures 3.8 and 3.3.2). This suggests that the neural mechanisms underlying adaptation to time-compressed speech are similar across development and are related to auditory/phonological processing.

At exactly what level of auditory processing/speech perception (from low-level acoustic processing to abstract phonological representations) this adaptation takes place remains an open question. Our study doesn't address this question directly, but on the basis of the existing literature on newborns' auditory and speech perception abilities and certain aspects of our data, we formulate possible hypotheses. First, the temporal and temporo-parietal localization of our effects suggests that prosody may play a role. These brain areas have been shown to be involved in prosodic processing in adults and in infants (Homae et al., 2006, 2007). This involvement of prosodic processing in time-compressed speech in newborns is consistent with the large body of evidence regarding newborns' extensive use of sound patterns, in particular prosody, to encode speech and language. For instance, newborns are able to recognize their native language based on its linguistic rhythm (Bertoncini et al., 1995; Nazzi et al., 1998), discriminate unknown languages if those are rhythmically different (Ramus, 2002) and their brain discriminates the prosodic structure of their native language from that of other languages (Abboub et al., 2016).

Indeed, in the present study babies adapted to time-compressed speech in their native language. While newborns lack experience with broadcast speech, they do have intrauterine experience with their native language. The intrauterine speech signal is different from the broadcast one, as maternal tissues act as low-pass filters, suppressing most individual sounds, but the low frequency modulations carrying prosody are well transmitted in the womb (Querleu et al., 1988). Even thought the fetal auditory system is not fully functional yet, frequencies below 300 Hz are transmitted to the fetal inner ear (Gerhardt et al., 1992). The ability of newborns to recognize their native language (Moon et al., 1993) and its prosodic structure (Abboub et al., 2016) is evidence that they readily perceive and learn about the prosody of their native language during the end of gestation. Thus to process time-compressed speech, newborns may have relied on their knowledge of the rhythmic structure of their mother tongue to track phonological landmarks in the time-compressed signal, at least for the moderate compression rate. If newborns do indeed rely on their prenatal experience, then it is possible that they can also adapt to time-compressed speech in unfamiliar languages, provided that those belong to the same rhythmic class as their native language, just as adults can (Pallier et al., 1998). By contrast, they should not be able to adapt to time-compressed speech in a language from a different rhythmic class. Furthermore, testing adaptation to time-compression in non-linguistic sounds will make it possible to disentangle

general auditory vs. speech-/language-specific mechanisms. Future research is needed to investigate these predictions.

3.5 Conclusions

Using moderately and highly compressed speech, we have shown that newborns are able to discriminate normal from time-compressed speech, and process normal and moderately compressed speech in similar ways. Highly compressed speech, by contrast, is treated differently. These results mirror those found in adults, bringing the first developmental evidence for the hypothesis that the ability to adapt to time-compressed speech relies on auditory, most probably prosodic/rhythmic, mechanisms.

Chapter 4

Role of Linguistic Rhythm

This chapter is adapted from Issard & Gervain (submitted)

4.1 Introduction

Despite the stable percept that we experience in everyday conversations, speech is a highly variable signal. Speech sounds may vary as a function of the emotional content of the message, the speaker's accent, gender and identity (Magnuson and Nusbaum, 2007). Speech also varies considerably in time. Speech rate differs within and across speakers, but this typically does not impede communication. While the perceptual and neural mechanisms underlying the ability to adapt to speech at different rates is relatively well understood in the context of speech comprehension, the respective contributions of low-level auditory mechanisms and linguistic expertise remain poorly understood. The current study seeks to address this question.

Adaptation to time-compressed speech has been investigated both at the behavioral and neural levels. Successful adaptation to time-compressed speech in the native language has been observed in adults and older children, in tasks such as word comprehension (Dupoux and Green, 1997; Orchik and Oelschlaeger, 1977), sentence comprehension (Ahissar et al., 2001; Peelle et al., 2004), or reporting syllables (Mehler et al., 1993; Pallier et al., 1998; Sebastián-Gallés et al., 2000) for speech compressed up to about 30-20% of its original duration.

In an fMRI study with adults, Peelle et al. (2004) presented syntactically simple and complex sentences compressed to 80%, 65%, and 50% of their original duration. Time-compressed sentences produced activation in the anterior cingulate, the striatum, the premotor cortex, and portions of temporal cortex, regardless of syntactic complexity. This shows that adaptation to time-compressed speech is supported by a sensory-motor network, operating independently of regions involved in syntactic processing. Other studies found that some brain regions, e.g. the Heschl's gyrus (Nourski et al., 2009) and the neighboring sectors of the superior temporal gyrus (Vagharchakian et al., 2012) showed a pattern of activation that followed the temporal envelop of compressed speech, even when linguistic

comprehension broke down, e.g. at 20% compression rate. Other brain areas, such as the anterior part of the superior temporal sulcus, by contrast, showed a constant response, not locked to the compression rate of the speech signal for levels of compression that were intelligible (40%, 60%, 80% and 100% compression), but ceased to respond for compression levels that were no longer understandable, i.e. 20% (Vagharchakian et al., 2012). This suggests that the auditory and linguistic levels of processing may be dissociable, at least in adults.

Since adults and older children are experienced listeners, the question arises whether their ability to adapt to time-compressed speech is a result of experience with the native language, a more basic auditory ability or may be influenced by both. To address this question, some studies investigated whether listeners can adapt to time-compressed speech in unfamiliar languages, where participants cannot rely on top-down linguistic knowledge and familiarity. Specifically, using 40% compression, Mehler et al. (1993) presented French or English sentences to French and English monolinguals and to French-English bilinguals. After a few minutes of familiarization with time-compressed sentences, participants correctly reported higher numbers of words when they were initially familiarized with and then tested in their native language. A follow-up study showed that Spanish-Catalan bilinguals adapted to Spanish or Catalan sentences compressed at 38% after habituation in the other language (i.e. habituation in Spanish before test in Catalan, and habituation in Catalan before test in Spanish). In two subsequent studies, English monolingual adults tested with 40% compressed English sentences benefited from habituation to time-compressed speech in English, and in a lesser extent in Dutch, but showed no adaptation when familiarized with French. Spanish monolinguals tested with 38% compressed Spanish sentences benefited from habituation with Catalan and Greek, but not with English and Japanese (Pallier et al., 1998; Sebastián-Gallés et al., 2000). Importantly, these studies used unfamiliar languages that were either rhythmically similar to (Spanish & Catalan; Spanish & Greek; English & Dutch), or rhythmically different from (French & English; Spanish & Japanese, English) the participants' native language, following the operational definition of rhythm proposed by Ramus et al. (2000). The results show that adults are able to adapt to moderately compressed speech in unfamiliar languages, if those belong to the same rhythmic class as their native language. This implies that adaptation to time-compressed speech does not take place at the level of abstract linguistic representations, but rather at the phonological level, with linguistic rhythm playing a key role.

Investigating the developmental origins of the ability to process time-compressed speech is particularly relevant in this regard. Newborns do not have abstract linguistic knowledge about the grammar or the lexicon of their native language. Nor do they have extensive experience with listening to broadcast speech transmitted through the air. Investigating their ability to adapt to time-compressed speech is therefore highly informative about the respective roles of auditory processes and experience.

Newborns have the experience-independent ability to discriminate languages,

even those unfamiliar to them, on the basis of their rhythm (Ramus et al., 2000). It is important to note, however, that newborns do have some language experience. Specifically, they have intrauterine experience with the sound patterns of the language(s) their mothers spoke during pregnancy. This experience mostly consists of prosody, i.e. melody and rhythm, as the more fine-grained details of the speech signal are filtered out by the uterine environment (Querleu et al., 1988). Newborns have indeed been shown to recognize their mother's voice (DeCasper and Fifer, 1980), their native language (Mehler et al., 1988; Moon et al., 1993) and its most common prosodic patterns (Abboub et al., 2016) on the basis of their prenatal experience with speech. Some of newborns' sophisticated speech perception abilities are thus broadly based and universal, while others are already influenced by prenatal experience.

In a previous study (Issard and Gervain, 2017), we have shown that newborns are able to adapt to time-compressed speech in their native language within a range of compression similar to the adult adaptation range (compression up to about 30% of the original duration). In the language they heard prenatally, French, they responded similarly to uncompressed speech and speech compressed to 60% of its original duration, but differently to speech compressed to 30% of its original duration (Issard and Gervain, 2017). Specifically, normal and 60% compressed speech elicited canonical, positive responses in the left temporo-parietal, right frontal and right temporal cortices, whereas 30% compressed speech evoked an inverted, negative response (i.e. a decrease instead of an increase of blood oxygenation) in the same areas.

These results confirm that the ability to adapt to time-compressed speech does not rely on grammatical/syntactic processes or experience with broadcast speech. Whether experience with the language sound patterns is relevant remains an open question. To address this issue, here we ask whether newborns' capacity to adapt to time-compressed speech relies on prenatal experience with the native language, and if so by what cortical network it is supported. To this end, we presented newborns with uncompressed (100%) speech, as well as speech compressed to 60% and 30% of its initial duration in Spanish, an unfamiliar language that is similar from the participants' native language, French in several phonological features, e.g. rhythm, lexical stress and phoneme repertoire, and in English, an unfamiliar language that is different from the participants' native language along these dimensions. Similarly to our previous study, we measured evoked cortical activity with functional Near Infrared Spectroscopy (fNIRS). If adaptation to time-compressed speech is initially a broad-based, general auditory mechanism, newborns might show similar brain responses to normal and time-compressed speech in any language irrespective of its familiarity or its similarity to the native language, with activity mainly in sensory (i.e. temporal) cortex. If experience with sound patterns of the language heard is crucial, then newborns may show different brain responses to normal and time-compressed speech in English, a language that is rhythmically different from the participants' native language, with activity over a broader cortical network, but not in Spanish, a language that is rythmically similar

to the participants' native language.

4.2 Experiment 2.a.: Spanish

4.2.1 Materials and Methods

Participants

We recruited 66 newborns (31 girls, 35 boys) during their stay at the maternity of the Robert Debré Hospital, Paris, France. All were born full-term with a gestational age superior to 37 gestational weeks and a birth weight appropriate for their gestational age. All had Apgar score greater than 8 at 5 minutes, indicating that they had not suffered from hypoxia at birth. All were healthy, as assessed by a local pediatrician, and had normal audition, as assessed by oto-acoustic emissions. All participants were less than 4 days of age (mean = 1.96 days, min = 0 d, max = 4 d). All participants had been exposed to French from the last trimester of gestation, and none of them had been exposed to Spanish during the same period. Parents of all participants gave written informed consent prior to participation. The study was approved by the ethics committee of the Université Paris Descartes (nr. 2011-13). 36 participants were excluded because of poor data quality (n = 12), failure to finish the experiment due to crying (n = 1), or machine dysfunction (n = 13). 30 participants were included in the final sample.

Apparatus

We recorded hemodynamic responses in 24 cortical channels with a NIRx NIRScout 1616 machine (NIRx Medizintechnik GmbH, Berlin, Germany). Sources were LED lights emitting pulsated narrow-band red and near-infrared light at 760 and 850 nm with a maximal power of 5 mW. The optodes were mounted on a flexible cap (EasyCap®). The optode configuration comprised 8 sources and 10 detectors arranged in two chevron patterns, forming 24 channels, with 12 channels over each hemisphere (source-detector separation: 3 cm). Channels covered the frontal, temporal and parietal cortices bilaterally (Figure 4.1. The layout was the same as in (Issard and Gervain, 2017), and other studies from our laboratory (Abboub et al., 2016; Benavides-Varela and Gervain, 2017).

Channel localization

We determined the approximate location of each channel on the cortical surface using a neonatal brain atlas (Shi et al., 2011). We obtained fiducial points from photographs of the participants wearing the headgear. Each optode was located on the skull surface on the 3D newborn atlas image based on the fiducial points derived from the photographs. The coordinates of the optodes were then saved. These points were subsequently projected onto the cortical surface and measure-

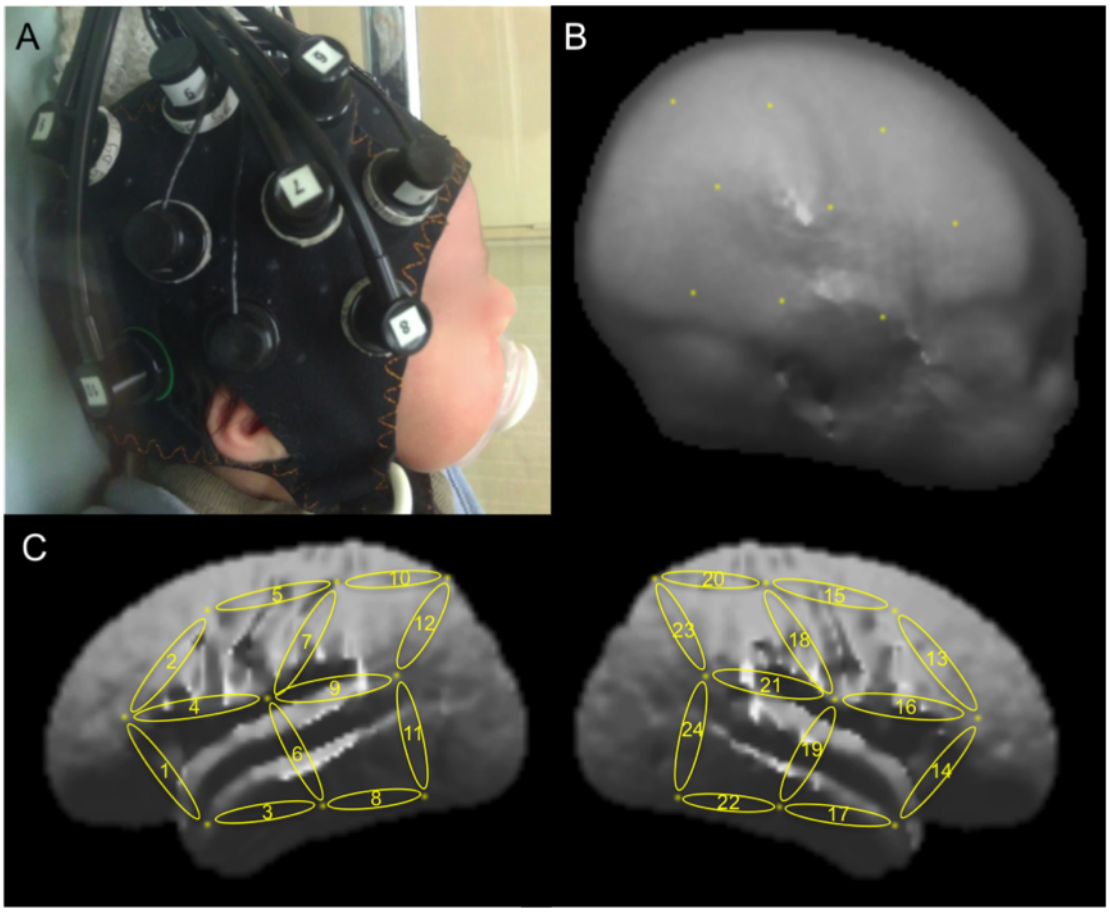


Figure 4.1: A: Picture of a participant wearing the cap. B: approximate optode localization on the skull. C: projection of the optodes and channels on the cortical surface.

ment channels were manually located. The resulting localizations on the skull and cortical surface are shown in Figure 4.1.

Stimuli

Stimuli were 88 infant-directed Spanish sentences retrieved from the CHILDES corpus (MacWhinney, 2000), e.g. "Y que es lo que quiere comer entonces ?". Sentences consisting of 11 or 12 syllables were selected. A female native English speaker recorded the stimuli in an infant-directed speech manner. Each sentence was then time-compressed to 60% and 30% of its initial duration using the "Change Tempo" function in the Audacity sound editing software. This function can modify duration while compensating for the pitch shift, so pitch remained unchanged. The spectrogram of a representative sentence and its time-compressed counterparts are presented in Figure 4.2.

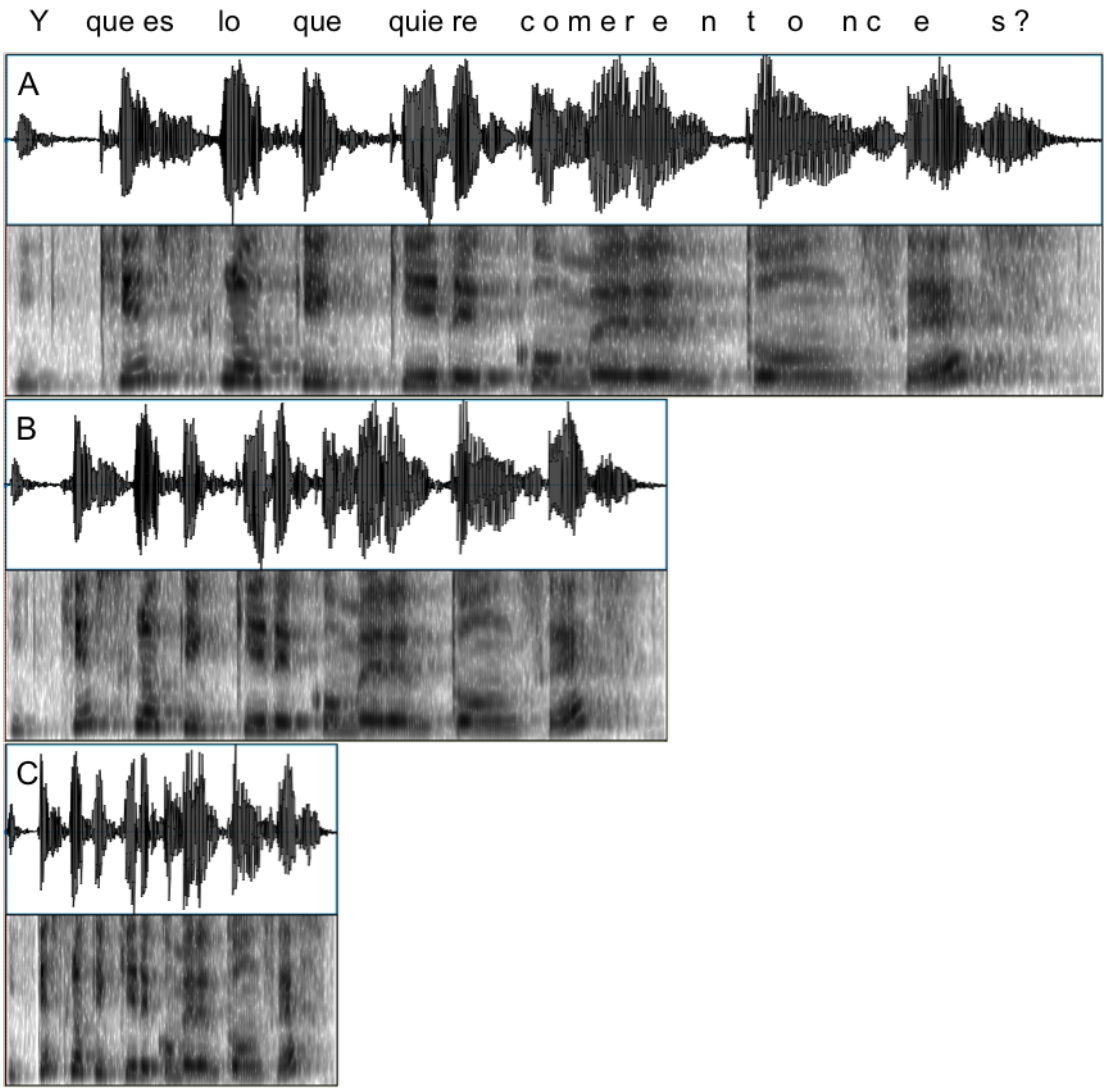


Figure 4.2: Waveform and spectrogram of a Spanish sentence for each compression rate. A: Normal duration, B: 60%-compressed, C: 30%-compressed.

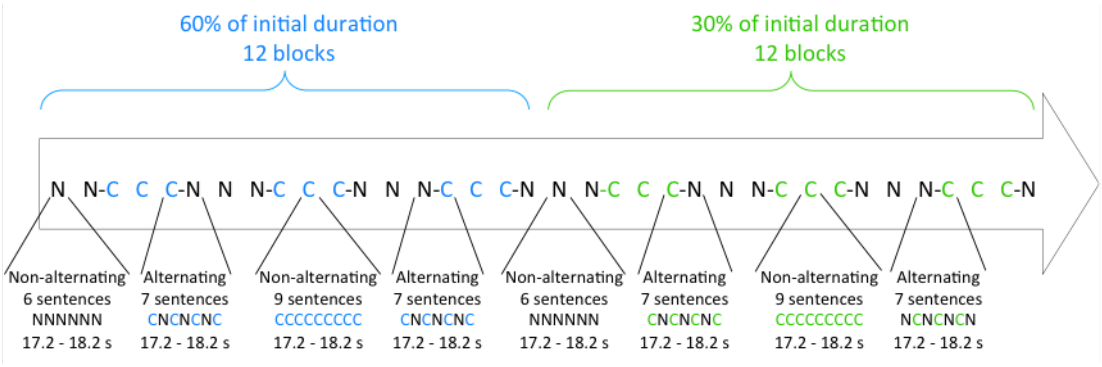


Figure 4.3: The alternating/non-alternating design used in experiment 2.a. and 2.b. N: non-compressed, C: compressed

Procedure

Newborns were tested in their bassinet while naturally asleep. All parents gave informed written consent prior to participation. Stimuli were presented through loudspeakers placed in front of the bassinet, approximately 1.5 meters away. Stimuli were presented in an alternating/non-alternating design, often used in behavioral and NIRS studies for fine discrimination (Best and Jones, 1998; Gervain et al., 2012; Sato et al., 2010) as well as in our previous NIRS study on time-compression (Issard and Gervain, 2017). In half of the blocks (alternating blocks), normal sentences alternated with time-compressed sentences. Half of the alternating blocks started with a normal sentence, the other half with a time-compressed sentence. These alternating blocks alternated with non-alternating blocks with only normal or only time-compressed sentences. Half of the non-alternating blocks contained normal sentences, the other half time-compressed sentences. The order of the blocks was counter-balanced between participants. In one part of the experiment, 60% compression was used, in the other part, 30% (Figure 4.2.1). The order of the two halves was counter-balanced across participants. We adjusted the number of sentences for each type of block to have approximately the same duration, from 17.2 to 19.2 s.

Data analysis

Signal processing The NIRS signal was processed with custom MATLAB scripts from the McDonnell consortium (Gervain et al., 2011). The raw signal was filtered using a 4th order Butterworth band-pass filter with zero phase-shift between 0.01 and 0.075 Hz to remove artifacts such as heartbeat and potential drifts. The time series was then segmented into epochs starting at – 5 s from block onset to + 20 s from block offset. Each epoch was detrended by fitting a linear trend between the first and the last five seconds of the epoch. Artefacts were defined as variations superior to 0.05 mmol within a 200 ms time window. Epochs containing artefacts

were removed from the data set. Light intensity was transformed to oxygenated hemoglobin (oxyHb) and deoxygenated hemoglobin (deoxyHb) concentrations with a modified Beer-Lambert law. Differential pathlength factor (DPF) values were chosen according to the age of participants (Scholkmann and Wolf, 2013). Participants were included in the data if they contributed at least 2 trials per condition and had satisfactory data quality upon visual inspection.

Statistical analyses The statistical significance of hemodynamic responses with respect to baseline was assessed for each channel using a permutation test (Holmes et al., 1996; Maris and Oostenveld, 2007; Nichols and Holmes, 2002) over oxyHb concentrations, known to produce more reliable and robust effects in infant data (Gervain et al., 2011; Lloyd-Fox et al., 2010). To assess the significance of the hemodynamic response in each experimental condition in each channel, we compared the mean amplitude of the response of each condition to a zero baseline. The mean amplitude of the responses was computed by averaging the amplitude in a pre-defined time-window from 12 to 36 s block onset. We then computed a one-sample t-statistic. The significance of this t-value was assessed by randomly permuting the labels of the experimental conditions within each participant's data a thousand times. For each permutation, the maximal t-value over the 24 channels was retrieved to draw the null distribution of maxima. Finally, the statistical significance of each channel was determined by checking its t-value in the original data against the null distribution. This procedure offers control of the Type I error and deals with the multiple comparisons problem (Holmes et al., 1996; Nichols and Holmes, 2002).

To determine if the activity was clustered within regions, we also used cluster-based permutation tests. We compared the mean response of each condition to a zero baseline computing a one-sample t-test for each channel-time sample. Clusters of activity were then identified within the data as samples with a significant t-value adjacent to one another in time or space. The labels of the experimental conditions were then randomly permuted within each participant's data, and clusters of significant t-statistics in time and space were calculated. This permutation was repeated 1000 times to draw the null distribution of clusters sizes. The statistical significance of each cluster was finally determined by checking its size against the null distribution. The scripts written to perform both types of permutation tests are available at github.com/cissard/NIRS.

For the main effect of compression rate, two comparisons were made. First, the non-alternating blocks with the three different compression rates were compared directly to evaluate their processing. We used the same permutation tests as above, except for replacing the t-test with a one-way ANOVA with three levels (hence using the F-statistic). If clusters emerged as significant in this latter comparison, follow-up pair-wise comparisons using the cluster-based permutation test were conducted to determine the source of the effects. Differences between conditions for each channel were assessed using an F-statistic and a two-tailed t-statistic. The time window over which the amplitude of the response was averaged

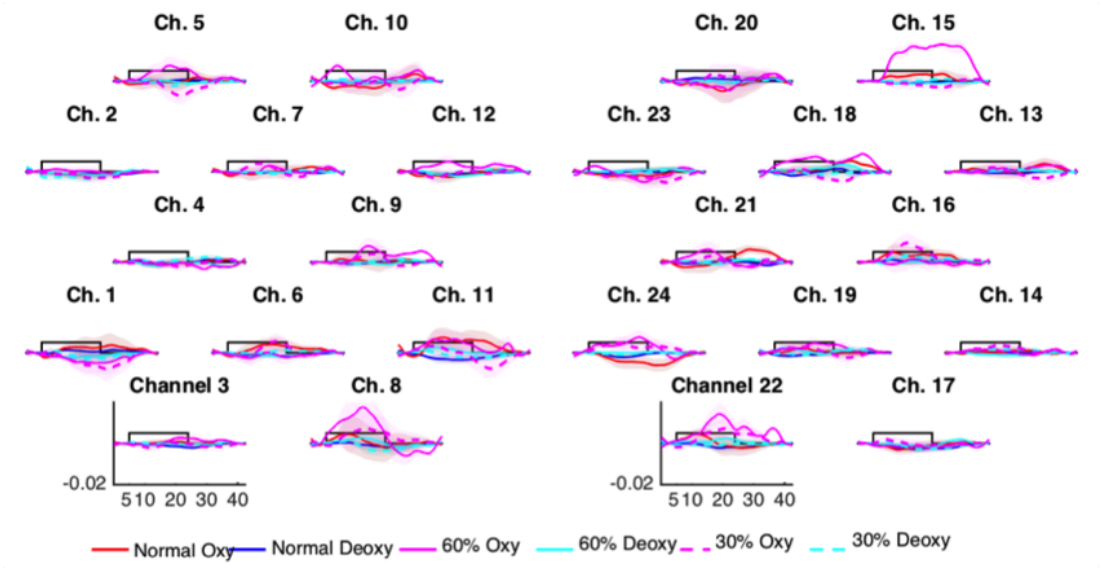


Figure 4.4: Hemodynamic responses evoked by each of the three compression rates in experiment 2.a. (Spanish). Shaded areas represent SEM. Rectangles represent the stimulation.

was the same as for baseline comparisons.

Finally, the alternating and non-alternating blocks were compared within each half of the experiment (30% and 60% compression) to test for discrimination between normal and time-compressed speech at each compression rate using the same permutation tests as above.

4.2.2 Results

Non-alternating blocks

We found no significant response to any of the three conditions (normal, 60% and 30% compressed speech) as compared to baseline in any channel, nor any significant cluster. There was no main effect of compression rate. The mean hemodynamic responses evoked by each compression rate in each channel are plotted on figure 4.2.2.

Alternating vs. non-alternating blocks

We compared the alternating and non-alternating blocks for the 60% compressed and the 30% compressed parts (i.e. each half of the experiment) separately.

We found no significant cluster in the 60% compressed part of the experiment. Nor for the alternating, nor for the non-alternating, nor for the difference between the alternating and the non-alternating blocks. The hemodynamic response elicited by each type of block in each channel are plotted in Figure 4.2.2.

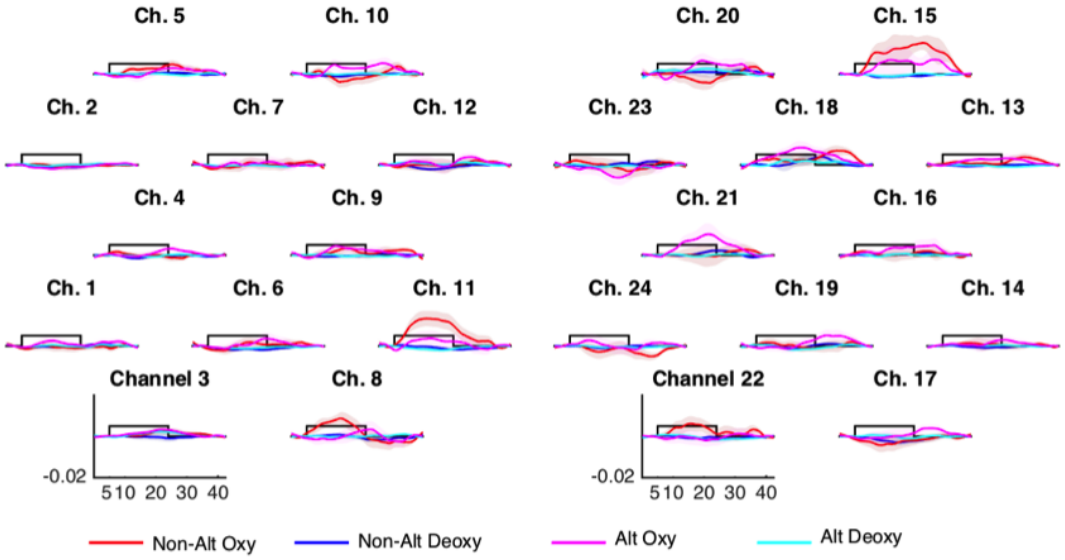


Figure 4.5: Hemodynamic responses evoked by the alternating and non-alternating blocks in the 60% compressed part of experiment 2.a. (Spanish). Shaded areas represent SEM. Rectangles represent the stimulation.

We found no significant cluster in the 30% compressed part of the experiment. Nor for the alternating, nor for the non-alternating, nor for the difference between the alternating and the non-alternating blocks. The hemodynamic response elicited by each type of block in each channel is plotted in Figure 4.2.2.

4.3 Experiment 2.b: English

4.3.1 Materials and methods

Participants

We recruited 43 newborns during their stay at the maternity of the Robert Debré Hospital, Paris, France. All were born full-term with a gestational age superior to 37 weeks. All had Apgar scores greater than 8 at 5 minutes after birth, indicating that they had not suffered from hypoxia at birth. All were healthy, as assessed by a local pediatrician, and had normal hearing, as assessed by the oto-acoustic emissions test. Participants were between 0 and 3 days of age (mean = 1.95 days, min = 0 d, max = 3 d). All participants had been exposed to French from the last trimester of gestation, and none of them had been exposed to English during the same period. Parents of all participants gave written informed consent prior to participation. The study was approved by the ethics committee of the Université Paris Descartes (nr. 2011-13). Fourteen participants were excluded from final data analysis because of poor data quality (n = 13) or machine dysfunction (n = 1). Exclusion was done in a batch, before statistical results were known.

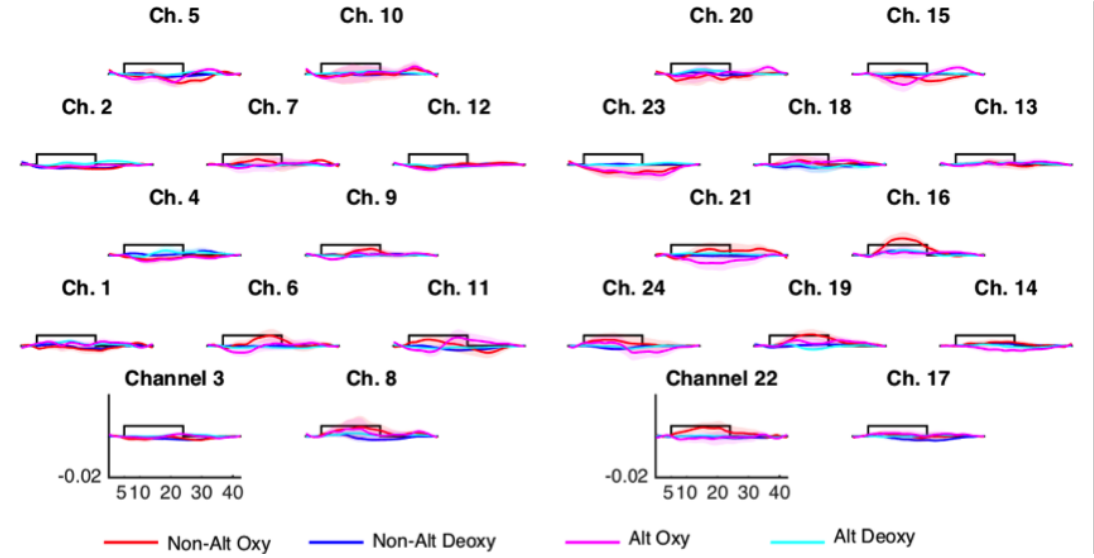


Figure 4.6: Hemodynamic responses evoked by the alternating and the non-alternating blocks in the 30% compressed part of experiment 2.a. (Spanish). Shaded areas represent SEM. Rectangles represent the stimulation.

Apparatus

The apparatus was the same as for experiment 2.a.

Channel localization

Channels localization was the same as in experiment 2.a.

Stimuli

Stimuli were 88 infant-directed English sentences retrieved from the CHILDES corpus (MacWhinney, 2000), e.g. "Here, the dog's gonna go up the ladder!". Sentences consisting of 11 or 12 syllables were selected. A female native English speaker recorded the stimuli in an infant-directed speech manner. Each sentence was then time-compressed to 60% and 30% of its initial duration using the "Change Tempo" function in the Audacity sound editing software. This function can modify duration while compensating for the pitch shift, so pitch remained unchanged. The spectrogram of a representative sentence and its time-compressed counterparts are presented in Figure 4.3.1.

Procedure

The procedure was the same as in experiment 2.a.

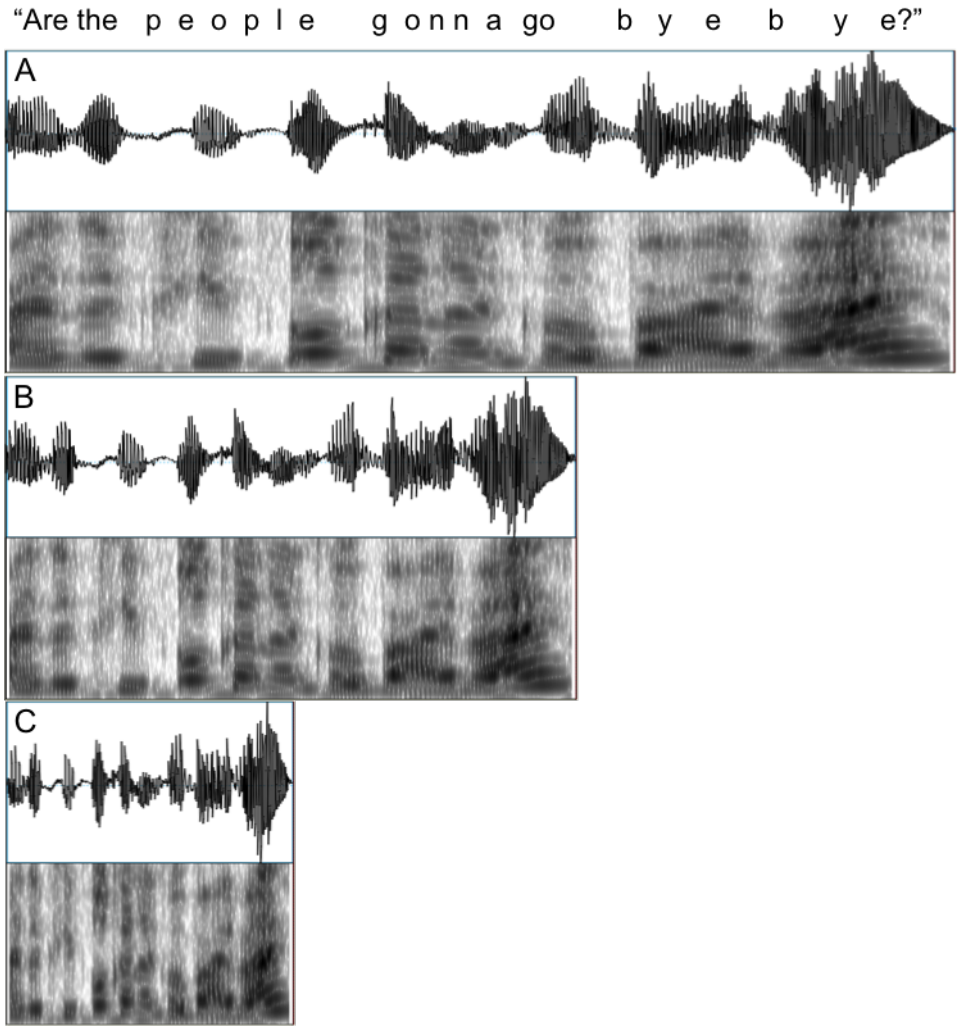


Figure 4.7: Waveform and spectrogram of a sentence for each compression rate. A: Normal duration, B: 60%-compressed, C: 30%-compressed.

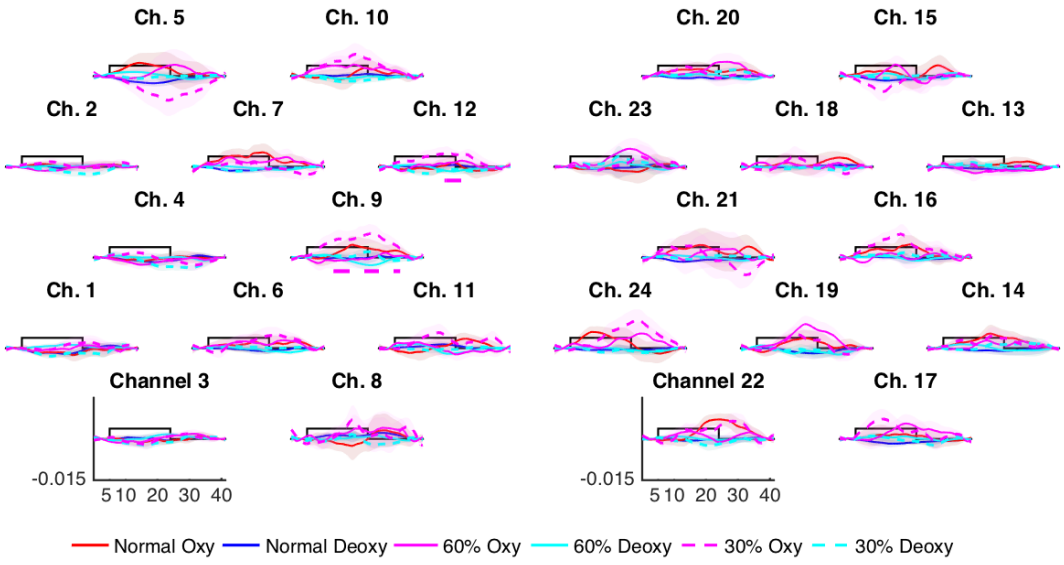


Figure 4.8: Hemodynamic response to each compression rate in each channel. Magenta dashed lines under the plots of channels 9 and 12 represent significant activation for 30%-compressed speech (as compared to baseline). Rectangles represent the time of stimulation. The x-axis represents time in seconds. The y-axis represents concentration change of oxygenated or deoxygenated hemoglobin in mmol.

Data analysis

The data was analyzed in the same way as in experiment 2.a.

4.3.2 Results

Non-alternating blocks

Comparing the three types of non-alternating blocks to baseline, we found that the 30% compressed blocks elicited a significant response in channel 9 and channel 12 (p=0.05 and p= 0.001, respectively). These two channels formed a significant cluster of activity (p=0.038) (Figures 4.8 and 4.9). The normal and 60% compressed blocks didn't elicit any significant response in any of the channels nor any significant cluster. The hemodynamic responses evoked by each compression rate in each channel are plotted on Figure 4.8.

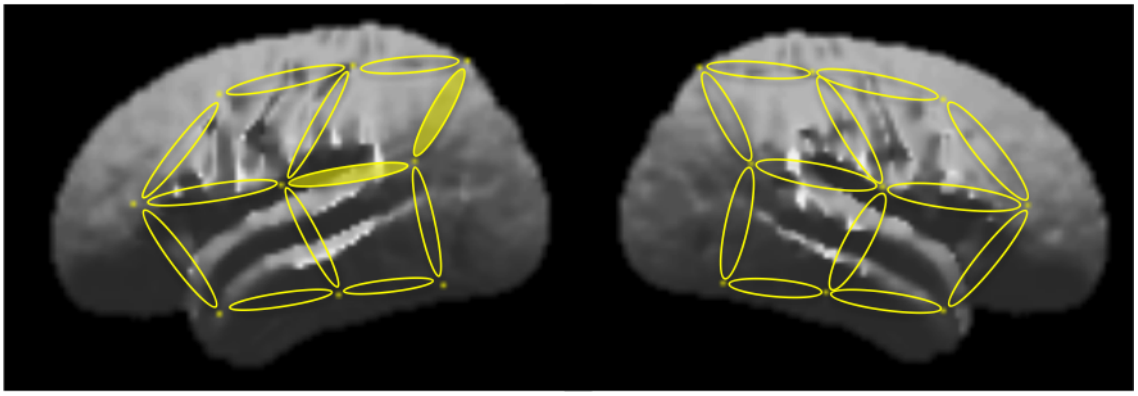


Figure 4.9: Cortical regions showing a larger hemodynamic response to 30%-compressed as compared to baseline.

Comparing the three compression rates directly, we found no main effect of compression rate at the channel or the cluster level.

Alternating vs. non-alternating blocks

We found no significant response with respect to baseline, nor any significance difference between the two types of blocks in the 60% compressed part of the experiment. The hemodynamic response elicited by each type of block in each channel is plotted on Figure 4.10.

In the 30% compressed part of the experiment, the non-alternating blocks (normal and 30% compression pooled together) elicited a significant response in channels 9 and 12 (p=0.034 and p=0.006, respectively). The cluster analysis yielded two significant clusters: one in the left hemisphere, covering channels 6, 9, 10 and 12; and one in the right hemisphere, covering channels 14, 16, 17, 19, and 21 (p= 0.001) (Figure 4.12). No activity in any channel or cluster for the alternating blocks, nor the difference between alternating and non-alternating blocks was significant (Figures 4.11).

4.4 Discussion

Presenting normal, moderately (60%) and highly (30%) compressed speech in two unfamiliar languages to newborns, we found that the neonatal brain processes highly-compressed speech differently than normal and moderately-compressed speech in an unfamiliar language that has a different rythmic structure from their native one. This difference manifests itself as a larger hemodynamic response to highly compressed speech in the left superior temporal and inferior parietal regions in English (experiment 2.b.). These results suggest that even in an unfamiliar language, highly compressed speech evokes a different response than speech that is within the potential adaptation range. These findings mesh well with the existing adult data

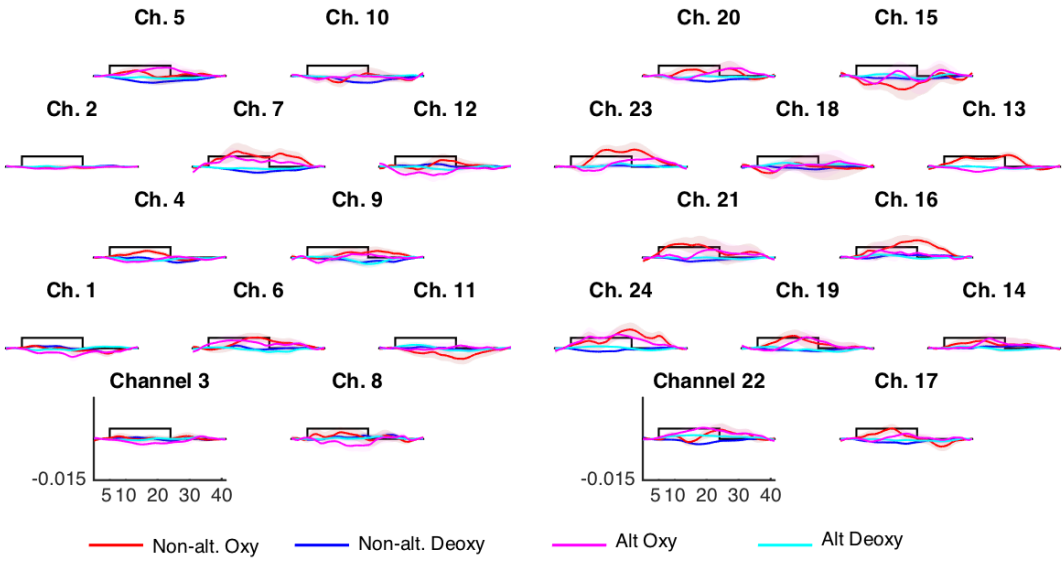


Figure 4.10: Hemodynamic responses evoked by the alternating and non-alternating blocks in the 60%-compressed part of experiment 2.b. (English). Rectangles represent the time of stimulation.

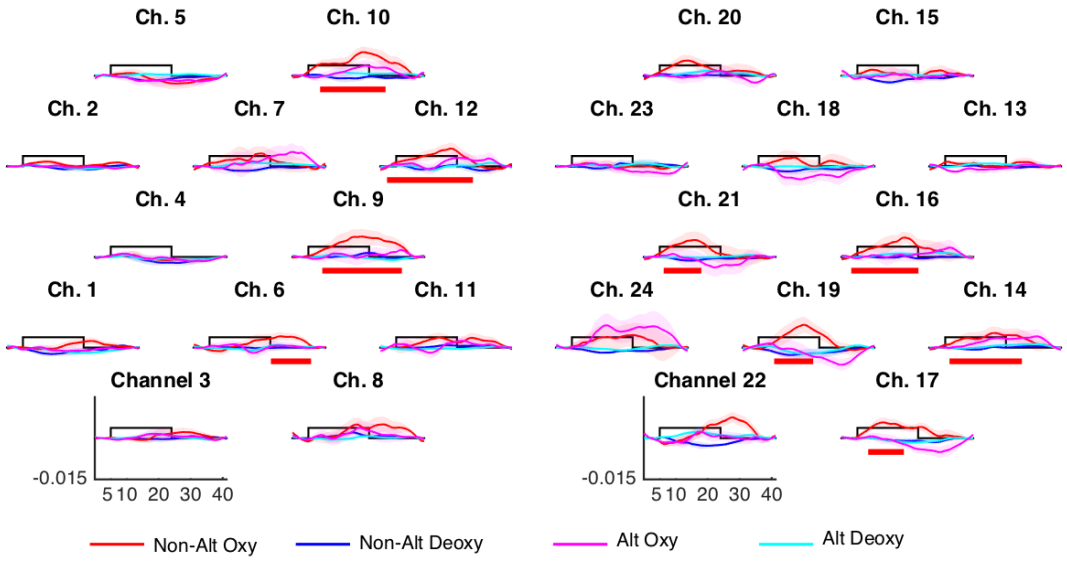


Figure 4.11: Hemodynamic responses evoked by the alternating and non-alternating blocks in the 30%-compressed part of experiment 2.b. (English). Rectangles represent the time of stimulation. Shaded areas represent SEM. Red bars under the channel plots represent the time windows of significant activation to the non-alternating blocks.

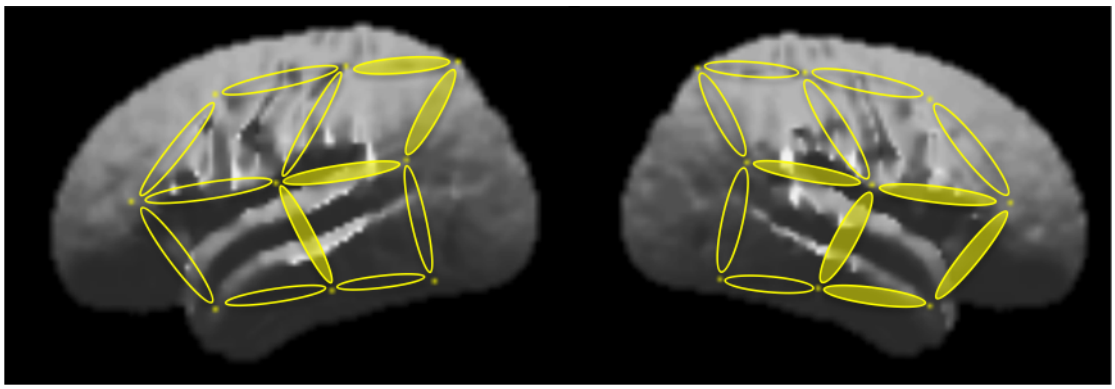


Figure 4.12: Cortical regions showing a larger activation to non-alternating blocks as compared to baseline for 30%-compressed speech.

showing a sudden drop of comprehension and neural responses at around 30% compression (Adank and Devlin, 2010; Dupoux and Green, 1997; Pallier et al., 1998; Sebastián-Gallés et al., 2000; Vagharchakian et al., 2012).

The global pattern of results is also similar to what we previously found with the participants' native language in that there was no difference between normal and moderately compressed speech, but a different response to highly compressed speech in the left temporo-parietal and right temporal region (Issard and Gervain, 2017). However, interesting differences also exist between the current and our previous results. Contrary to what we found for the native language, where the high compression rate evoked an inverted hemodynamic response, here it elicited a canonical, positive hemodynamic response. The exact interpretation of the inverted hemodynamic response in young infants remains poorly understood (Issard and Gervain, 2018). One possibility is that the inverted response we had obtained for the native language reflects an effortful and unsuccessful processing of a stimulus that should otherwise (in the uncompressed and moderately compressed conditions) be processed normally leading not only to auditory, but more abstract linguistic representations, which would be the case for the native language. By contrast, newborns most likely do not process an unfamiliar language beyond the basic auditory level, which explains the weak activation to the uncompressed and moderately compressed condition in the current experiment. The increased activation to the highly compressed condition in this case would reflect its auditory 'ill-formedness' or unnaturality.

This interpretation is all the more plausible for an unfamiliar language that is rhythmically different from the native language. As newborns are known to discriminate between rhythmically different, but not between rhythmically similar languages, their prenatal experience with speech rhythm may benefit the processing of an unfamiliar language that is rhythmically similar to the native one. In case of a rhythmically different language, no such advantage may be expected. It is thus not unlikely that French newborns may not process English stimuli beyond the basic auditory level. The absence of significant difference between normal and highly-compressed speech in Spanish is puzzling. However, we had to reject more than 50% of the participants in this experiment due to low data quality. This experiment also suffered from several machine dysfunction that hindered data quality. Due to the difficulty to test newborn participants, we chose to include as many participants as possible despite data of variable quality between channels. Another parameter that could have hindered data quality is the design. Alternating/non-alternating designs are typically used in behavioral experiments to show discrimination between two stimulus categories. In our case, this design was initially chosen to show subtle differences between encoding of normal and time-compressed speech if the response were similar for the three compression rates in the non-alternating blocks. We used the same design in the two experiments of the present study to make them comparable to the one presented in Chapter 2 and Issard and Gervain (2017). This design reduces the number of trial per condition as each condition is presented both during the non-alternating and

the alternating blocks. This might be a significant caveat in NIRS experiments, which necessitate long trials due to the slow hemodynamic response timecourse. The resulting number of trial that is possible to run is low, in our case three non-alternating blocks per condition. This made difficult to attain a good signal-to-noise ratio, and might have led to the non-significant differences obtained in the Spanish experiment (experiment 2.a.).

The fact that only highly-compressed speech elicited a significant activation as compared to baseline is consistent with previous adult results showing larger responses to time-compressed speech and a rapid decrease of response amplitude in various cortical regions, including auditory cortex (Adank and Devlin, 2010).

Regarding localization, highly compressed speech elicited larger responses than moderately compressed speech in the left superior temporal and inferior parietal regions. This pattern is consistent with brain regions activated during similar tasks in adults. In these studies, a superior activation for time-compressed speech was observed in the superior temporal gyrus (Adank and Devlin, 2010), with a positive correlation between compression rate and the amount of activity (Peelle et al., 2004). Previous adult studies also found activation in the inferior parietal and angular cortices. However, in adults, these activations were linked to intelligibility, as they collapsed when participants didn't understand the sentences any more (Vagharchakian et al., 2012). It is possible that in the present study, activation in this region was high because it wasn't constrained by comprehension, as newborns don't understand language yet. Taken together, the similarity between regions activated in the present study at birth and regions involved in adults highlights the fact that these regions rely on low-level acoustical processing to adapt to time-compressed speech, and not linguistic expertise that develops later in development. This idea is further supported by the fact that a similar region was activated in newborns when processing prosodic cues of their native language (Abboub et al., 2016), and in 3 month-old infants listening to natural as compared to speech with flat prosody (Homae et al., 2006, 2011). Hence, this temporo-parietal region may process the prosodic cues of speech in a certain time-range, allowing adaptation to speech rate independently of comprehension.

Interestingly, we didn't find any significant activation in the frontal regions. The frontal regions have been shown to be involved when adapting to time-compressed speech both in adults (Adank and Devlin, 2010; Vagharchakian et al., 2012) and in our previous newborn study (Issard and Gervain, 2017). However, these studies used the participants' native language. In newborn infants, frontal regions have been involved in tasks such as detecting structural regularities in speech (Bouchon et al., 2015; Gervain et al., 2008), or recognizing familiar speech sequences (Benavides-Varela et al., 2012). All of these studies investigated abilities that go beyond auditory processing. As suggested above, it is possible that stimuli from an unfamiliar language do not trigger processing beyond the auditory level, at least at birth, and do not therefore activate the frontal cortices.

In other words, although newborns are not linguistic experts yet, they have significant pre-natal experience with their native language, and this prenatal ex-

perience has been shown to shaping cortical response to speech at birth (Abboub et al., 2016; May et al., 2011). Participants' knowledge of the temporal structure of their native language may lead to a more in-depth processing, leading to more robust differences between normal and moderately compressed speech as opposed to highly compressed speech in a broad cortical network involving sensory and higher-level regions (Issard and Gervain, 2017). By contrast, in the present study, the unknown sound patterns and rhythmic structure of English presented may have led to more limited processing, with less cortical regions recruited and weaker responses, and no discrimination between normal and moderately compressed speech in this unfamiliar language, as suggested by the absence of difference between alternating and non-alternating blocks in the 60%-compression part of the experiment.

4.5 Conclusions

Our study shows differential processing for normal and moderately compressed speech as compared to highly-compressed speech in the newborn temporo-parietal cortex in an unfamiliar language, rhythmically different from the participants' native one. Although results are overall similar to the native language, differences in the exact neural response and the cortical regions recruited may be explained by different levels of processing activated by the two languages: a (pre-)linguistic level for the native language involving both temporo-parietal and frontal cortices, while a lower-level acoustic one for the unfamiliar language limited to the temporo-parietal region. The similarity of localization with adult studies This suggests that prenatal experience with the native prosodic structure modulates neural responses to time-compressed speech.

Chapter 5

Electrophysiological mechanisms for the stable encoding of speech

5.1 Introduction

In the previous chapters, we have shown that infants can adapt to variability in speech rate already at birth, both in the native language and in an unfamiliar language. We will now explore the electrophysiological mechanisms supporting the stable encoding of speech in this population.

The basic mechanisms are now starting to be understood in adults. A series of studies investigated the role of amplitude envelope tracking to encode speech, specifically whether, and if yes, how envelope tracking contributes to the processing of time-altered speech. A first behavioral study provided evidence for the role of the theta rhythm (4-8 Hz): When silent intervals were introduced within sentences compressed to 33% of their initial duration, comprehension (used as an index of successful speech encoding) was restored if silences occurred at a periodical rate that restored the theta-syllable rhythm – 80ms silences following 40 ms of time-compressed speech, aligned with the typical 120 ms duration of syllables (Ghitza and Greenberg, 2009). Accordingly, when adults (Pefkou et al., 2017) and typically developing children (Guiraud et al., 2018) listened to speech at various rates, their oscillatory activity was phase-locked to the envelope of the sound in the theta band. Importantly, the oscillations followed the syllable rate of speech even beyond the theta band (~12Hz). Power also increased with syllabic rate in the theta-band in adults (Pefkou et al., 2017). Other studies also found phase-locking to the time-compressed speech envelope in other frequency bands, such as a broad 0-20 Hz band (Ahissar et al., 2001), and the high-gamma band (70 Hz and above) (Nourski et al., 2009). Phase-locking was observed whether or not the participant understood the sentences across various cortical areas, regions as early as the Heschl's gyrus (Davidesco et al., 2018; Nourski et al., 2009). To sum up, the brain encodes the amplitude envelope of speech by phase-locking its oscillatory activity in the theta and high-gamma bands to the input, corresponding to two important rhythmic scales of speech sounds (i.e. syllabic and phonemic scales), as well as the

entrainment of the neural oscillation to the amplitude modulation of the speech envelope (Giraud and Poeppel, 2012).

Other authors pointed to the fact that the amplitude envelope is easily degraded in noisy backgrounds and reverberant environments, and that spectral contents may play a central role in speech encoding (Obleser et al., 2012). Indeed, using the phase pattern of the theta-band to discriminate sentences in various acoustical conditions, sentences with degraded spectral content were less accurately discriminated than sentences with degraded envelope (Luo and Poeppel, 2007). This underlines the idea that spectral content also plays a crucial role in speech encoding, even in quiet. Interestingly, the perception of amplitude (i.e. envelope) and frequency modulations might rely on similar encoding strategies. Several studies specifically investigated the encoding of frequency modulations. In a first study, human participants listened to a tone modulated in amplitude at 37 Hz and in frequency between 0.3 and 8 Hz. An auditory steady-state response (aSSR, Picton et al., 2003) was observed in the auditory cortex at the AM frequency (i.e. 37 Hz) with spectral sidebands (i.e. secondary peaks) at AM+/- FM frequencies (e.g. 37.8 Hz if the tone FM rate was 0.8 Hz), providing evidence for integrated encoding of amplitude and frequency modulations in the auditory cortex. Moreover, for slower FM (< 5 Hz), the phase of the aSSR at the AM frequency tracked the stimulus carrier frequency change (Luo et al., 2006). This encoding of frequency modulations through phase-locking was confirmed by subsequent studies. In particular, using complex tones frequency-modulated at 3 Hz, an aSSR was observed at the same frequency, and these delta oscillations were phase-locked to the tone FM (Henry and Obleser, 2012). Finally, using a larger range of FM, a last study replicated the aSSR at the FM frequency up to 8 Hz, but not above, providing further evidence for two distinct encoding mechanisms for FM (Millman et al., 2010). These studies show that frequency modulations are encoded by entrainment of neural activity at the FM rate, as shown by an aSSR at the corresponding frequency. Therefore, these results show that, like for the amplitude envelope, frequency modulations are encoded by a synchronization of cortical populations at the modulation rate, tracking the phase of the modulation.

However, animal studies showed that different mechanisms can be used. Contrary to human studies that found different responses to the different acoustical parameters (speech or FM rate), several animal studies presenting con-specific vocalizations have found that the response of neuronal populations in the inferior colliculus and the auditory cortex remained stable despite acoustical distorsions, both in the spectral and the temporal dimensions (Holmstrom et al., 2010; Carruthers et al., 2015). This stable response was obtained by neuronal gain control, i.e. neurons modulating their firing rate according to the range of acoustical parameters present in the stimulus (Rabinowitz et al., 2011; Natan et al., 2017). Therefore, these studies did not show a change in neuronal excitability, as would be expected for changes in oscillatory power and phase (Lakatos et al., 2005), as observed in humans. At a larger scale, hence closer to human studies, presenting natural sounds to awake ferrets reduced the inter-trial variance of low-frequency

LFPs (<16 Hz) in the primary auditory cortex as compared to spontaneous activity during silence, but this reduction was not correlated to mean LFP power. Surprisingly, mean LFP power was even reduced during the presentation of the sounds (Ding et al., 2016). Together, these studies point to an adaptation mechanism, where the brain adjusts its activity to the statistics of incoming sensory stimuli to encode them more efficiently, which is reflected in reduction in variance.

On which of these mechanisms does the newborns brain rely to encode temporal and spectral content in speech? Among the human studies cited above, those investigating the encoding of frequency modulations all used synthetic stimuli (i.e. modulated tones). To our knowledge no study have generalized their findings to natural sounds, and in particular speech. This chapter aims to fill this gap. We also wanted to investigate whether spectral and temporal variability are managed with similar strategies. Indeed, if brain activity adapts to the global statistical properties of the stimuli, there should be no difference between the spectral and the temporal domains. If speech is encoded through phase-locking of ongoing oscillations to the frequency or amplitude modulations, we should observe the highest phase-locking values in higher frequency-bands for time-compressed speech than for normal speech, corresponding to the faster amplitude modulations than in normal speech. No change with the range of frequency modulations should be observed due to the range of F0 variations in speech. If speech is encoded through increased neural synchrony, time-compressed speech should lead to aSSR at a higher rate than for normal speech. The aSSR should also appear at a frequency corresponding to the amplitude modulation rate of speech. Regarding frequency modulations, increase in speech frequency modulations should lead to an increase in power in a wider frequency band than for normal speech, whereas reduced frequency modulations should lead to an increase in power in a tighter frequency band than for normal speech. However, the discrepancy between the frequencies visible in EEG and the range of F0 variations in speech might make it difficult to observe such an effect. Finally, if speech is encoded through a adaptation to frequency and amplitude modulation range, then we should observe a decrease in inter-trial variance in all stimulus conditions.

To test these predictions, we investigated the electrophysiological responses of the newborn brain to variable speech sounds in two experiments. In a first experiment, we presented normal, moderately, or highly-compressed speech to the participants while the electrophysiological activity of their brain was recorded through EEG. In a second experiment, we presented speech with normal, enhanced or reduced frequency modulations to another group of participants, also recording the electrophysiological activity with EEG.

5.2 Experiment 3.a.: temporal variability

5.2.1 Materials and methods

Participants

Eighteen healthy full-term neonates participated in the experiment (mean age 1.63 days, range 1–4 days; 6 females). They were recruited during their stay at the hospital after birth. To participate, newborns had to be born at at least 40 gestational weeks, weight more than 2700 g, and have an Apgar score superior or equal to 8 at 10 min after birth. Newborns' hearing was assessed by a measurement of their oto-acoustic emissions during their stay at the maternity and through screening by a local pediatrician: no hearing disabilities, neurological disorders, prenatal or perinatal complications were reported for any of the participants. A questionnaire filled out by the parents assessed for parental handedness, antecedents of language or hearing impairments in the family. No relevant condition was present. Data from eleven newborns were not retained for the final analysis due to poor data quality (mostly because of movement artifacts and noise related to heart beat). To be retained for the final analysis, participants had to have at 40 good trials per condition. Importantly, this uniform rejection criterion was applied in batch to all infants whose data was pre-processed,leading to the above reported rejection, prior to statistical analysis. Seven participants were thus included in the final sample.

Apparatus

EEG activity was recorded from 12 Ag/AgCl electrodes mounted on a elastic cap and located over the frontal, central and parietal regions bilaterally at positions defined in the 10-20 system (F3, Fz, F4, C3, Cz, C4, T7, T8, F7, F8, TP9, TP10). We chose this restricted set of electrode to keep preparation time short and leave the occipital region free to allow participants to rest on their back without any discomfort, as long preparation or discomfort during the experiment may have led to movements or crying. The data of each electrode was referenced at Cz, forming 11 recording channels. The signal was amplified and digitized at 500 Hz, with an online low-pass filter at the Nyquist frequency (i.e. 250 Hz), using a Brain Vision ActiChamp (Brain Products GmbH, Germany) amplifier.

Stimuli

The stimuli were 6 utterances of 11 or 12 syllables, randomly chosen from the CHILDES (MacWhinney, 2000) French corpora and assessed for grammaticality and naturalness by a native French speaker, similarly to experiment 1. This reduced set of utturances was chosen to make acoustical analysis easier. We kept only the 6 utterances lasting more than 2 seconds to allow for longer epochs during data analysis. These 6 utterances lasted between 2.03 and 2.77 seconds at their normal duration. To generate time-compressed speech, we compressed the original utterances to 60% and 30% of their initial duration.

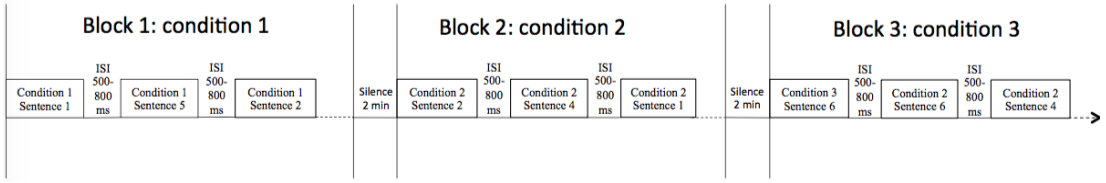


Figure 5.1: Experimental design used in experiment 3.a. and 3.b.

Procedure

The participants were tested directly in their bassinet while naturally asleep. The three experimental conditions (normal, 60% and 30% compressed speech) were presented in a block design: all trials from one condition were presented together on a row. Within blocks, the sentences were separated by silent ISI of random duration from 500 to 800 milliseconds. The sentences were presented in a random order within blocks. The blocks were separated by 2 minutes of silence with marks sent at random moments every 3 to 4 seconds. These silent periods were inserted to have a resting state measure as a negative control. This experimental design is schematized on Figure 5.1. The experiment was ended if participants showed signs of discomfort or upon parental request.

Data-analysis

Data were analyzed using EEGlab (Delorme and Makeig, 2004) and custom MATLAB scripts.

Event-related potentials Data from each participant was band-pass filtered between 1 and 30 Hz using a zero phase shift Butterworth filter, with a slope of -6 dB and a transition band width of 0.5 Hz (cutoff frequencies [0.5 30.5] Hz). We chose these values after visual inspection of the raw data to remove slow fluctuations in the data and high-frequency noise. These values were chosen in accordance to the newborn EEG literature (e.g. Stefanics et al., 2007; Mahmoudzadeh et al., 2017).

The time-series was then segmented into epochs around each stimulus, starting 0.2 second before stimulus onset and ending 600 ms seconds after stimulus onset. This post-stimulus length corresponded to the duration of the shortest sentence compressed at 30%. Each epoch was normalized by subtracting the mean of the pre-stimulus period ([-0.2 0] s) to each post-stimulus onset time point.

The data was sorted using a semi-automated procedure: epochs containing voltage deviations exceeding +/-100 μV relative to baseline at any of the electrodes were automatically removed from the dataset. All the epochs were then visually inspected to remove epochs containing non-stereotyped artefacts that were not detected automatically, such as sharp local deviations.

The normalized trial epochs were finally averaged within channels and conditions to create the ERPs.

Time-frequency analysis The trials retained in the analysis were the same as for event-related potentials.

Original trials (i.e., not subjected to any offline filtering) from each electrode were decomposed in the time–frequency domain using Complex Morlet Wavelets. The wavelets used to convolve the data spanned between 2 and 30 Hz, with a step of 1 Hz and a varying number of cycles of 3 to 10 to deal with the loss of precision when frequency increases. Time-frequency power was estimated in epochs spanning from -300 to 2000 ms post-stimulus onset to have a sufficient number of cycles at the lowest frequency. Each trial was normalized computing the percentage change in power relative to a prestimulus baseline period from -300 ms to stimulus onset. For each participant, changes in spectral power were averaged for each time-point and frequency band across trials.

5.2.2 Results

Due to the low number of participants included, we here only provide descriptive results.

Event-related potentials The event-related potentials obtained for each condition in each electrode are plotted on Figure 5.2. For normal speech, a negative deflection in the voltage starts around 200 ms and reaches its maxima around 500 ms after stimulus onset. For 60%-compressed speech, a negative deflection appears at the frontal electrodes (F7, F8, F3, F4, Fz) around 300 ms. For 30%-compressed speech, a peak appears at the stimulus onset at electrodes F7, T7, T8, TP9 and TP10. Large deflections in the voltage also appears during the silent periods at electrodes F3, F4, F7, T7, T8, TP9 and TP10. These peaks were larger at T7, T8 and TP10, and occured at similar latencies than the ones observed for the three experimental conditions.

Time-frequency analysis The time-frequency maps obtained for each speech rate and the silent trials in each electrode are plotted on Figure 5.3. During the silent period, a transient increase in power as compared to the pre-stimulus period appears to occur around 8 Hz, mostly visible at T7 and C3. By contrast, activity during the speech sounds at the three speech rates appears to be less pronounced.

5.3 Experiment 3.b.: spectral variability

5.3.1 Materials and methods

Participants

Fifteen healthy full-term neonates participated in the experiment (mean age 2.38 days, range 1–4 days; 10 females). They were recruited during their stay at the hospital after birth. To participate, newborns had to be born at at least 40 gestational

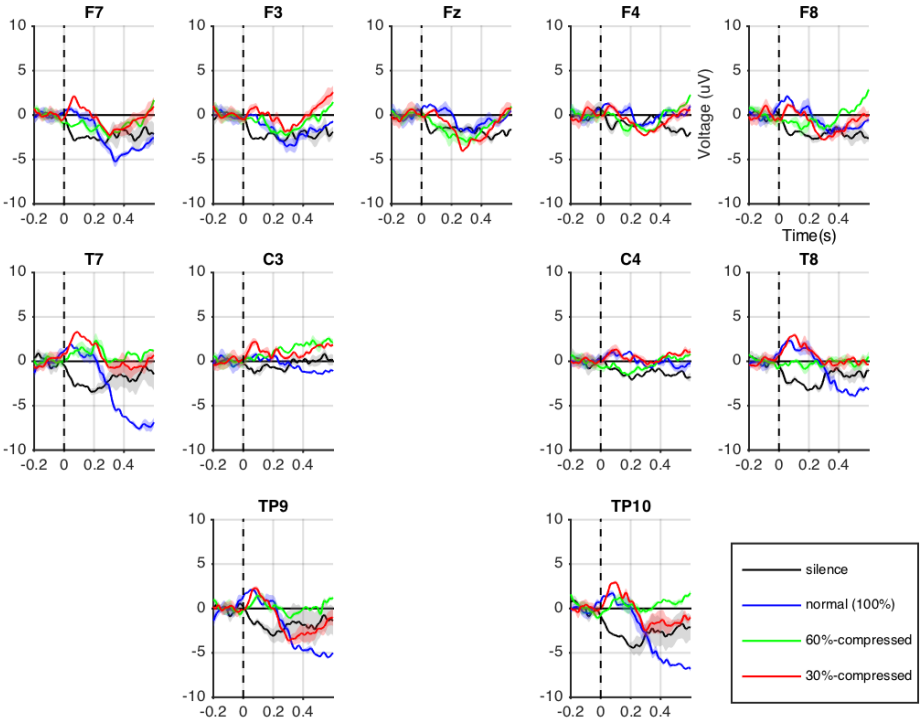


Figure 5.2: Event-related potentials obtained in experiment 3.a. for each speech rate. Shaded areas represent SEM.

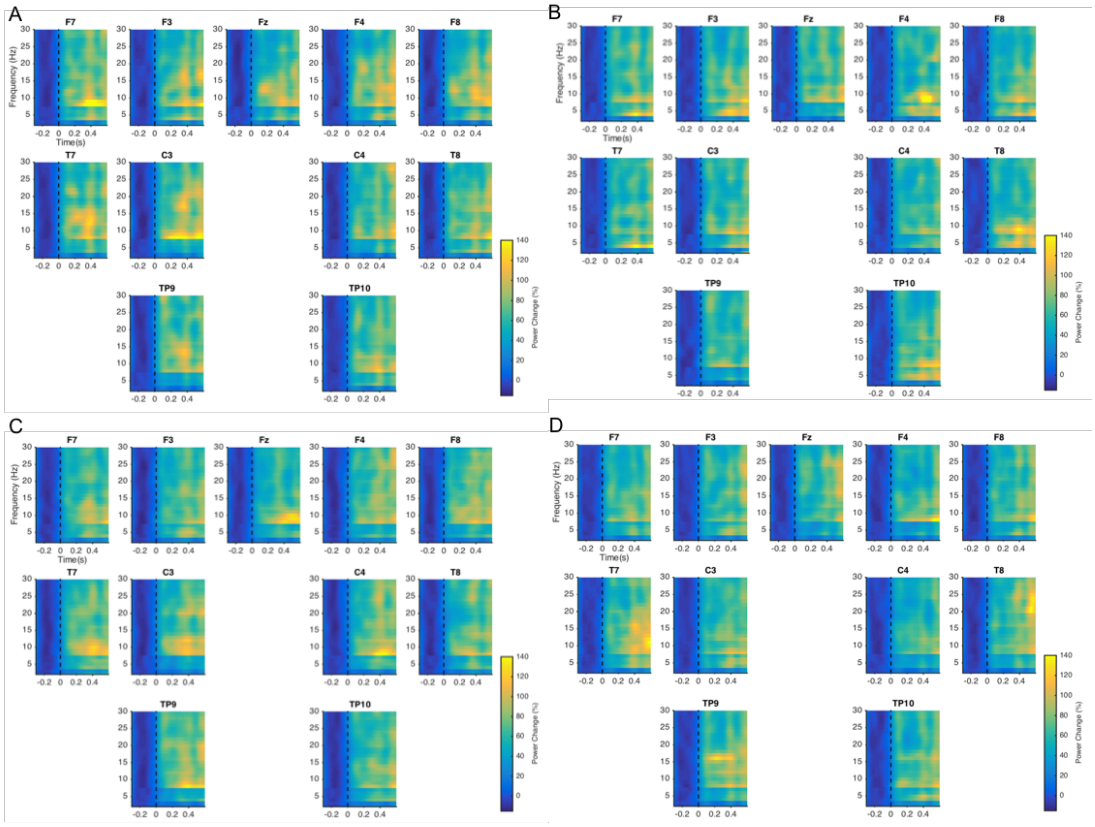


Figure 5.3: Mean power change for each time-frequency bin in experiment 3.a. for each speech rate. A: silence, B: 60%-compressed, C: 30%-compressed, D: normal (100%).

weeks, weight more than 2700 g, and have an Apgar score superior or equal to 8 at 10 min after birth. Newborns' hearing was assessed by a measurement of their oto-acoustic emissions during their stay at the maternity and through screening by a local pediatrician: no hearing disabilities, neurological disorders, prenatal or perinatal complications were reported for any of the participants. A questionnaire filled out by the parents assessed for parental handedness, antecedents of language or hearing impairments in the family. No relevant condition was present. To be retained for the final analysis, participants had to have at least 40 good trials per condition. Importantly, this uniform rejection criterion was applied in batch to all infants whose data was pre-processed, leading to the above reported rejection, prior to statistical analysis. Data from eight newborns were pre-processed, but were not retained for the final analysis due to poor data quality (mostly because of movement artifacts). Seven participants were thus included in the final analysis. All parents gave informed written consent. The study was approved by the local ethics committee at Paris Descartes University (CERES number 2011-013).

Stimuli

The stimuli were 6 utterances of 11 or 12 syllables, chosen from the CHILDES (MacWhinney, 2000) French corpora and assessed for grammaticality and naturalness by a native French speaker, similarly to experiment 1. The final 6 sentences were chosen within the initial set of sentences recorded in experiment 1 if they lasted at least 2 seconds, to allow segmentation of epochs of 2 seconds in the EEG data without offset effects. A native female French speaker recorded the selected utterances in an infant-directed manner. We then generated spectrally distorded speech using STRAIGHT (Kawahara et al., 2008). We compressed or dilated the spectrum of each sentence. First, the F0 value at each time point was extracted, and the F0 trajectory was log-transformed. A 4th degree polynomial was fitted to the F0 trajectory. The residuals at each time-point were either multiplied (spectral dilation) or divided by 4 (spectral compression). This created a new F0 trajectory with enhanced (*4) or reduced (/4) frequency modulations. The full speech sounds were finally resynthetized by recreating the original formants. This resulted in larger or smaller variations of frequency over time, hence in enhanced or reduced frequency modulations [1]. As a result, the F0 range was modified but not the spectral shape. An example of sentence, along with its two spectrally distorded analogs, is plotted on figure 5.4.

Apparatus

The apparatus was the same as in experiment 3.b.

[1] These stimuli were conceived and created by Maria N. Geffen and Chris Angeloni, University of Pennsylvania

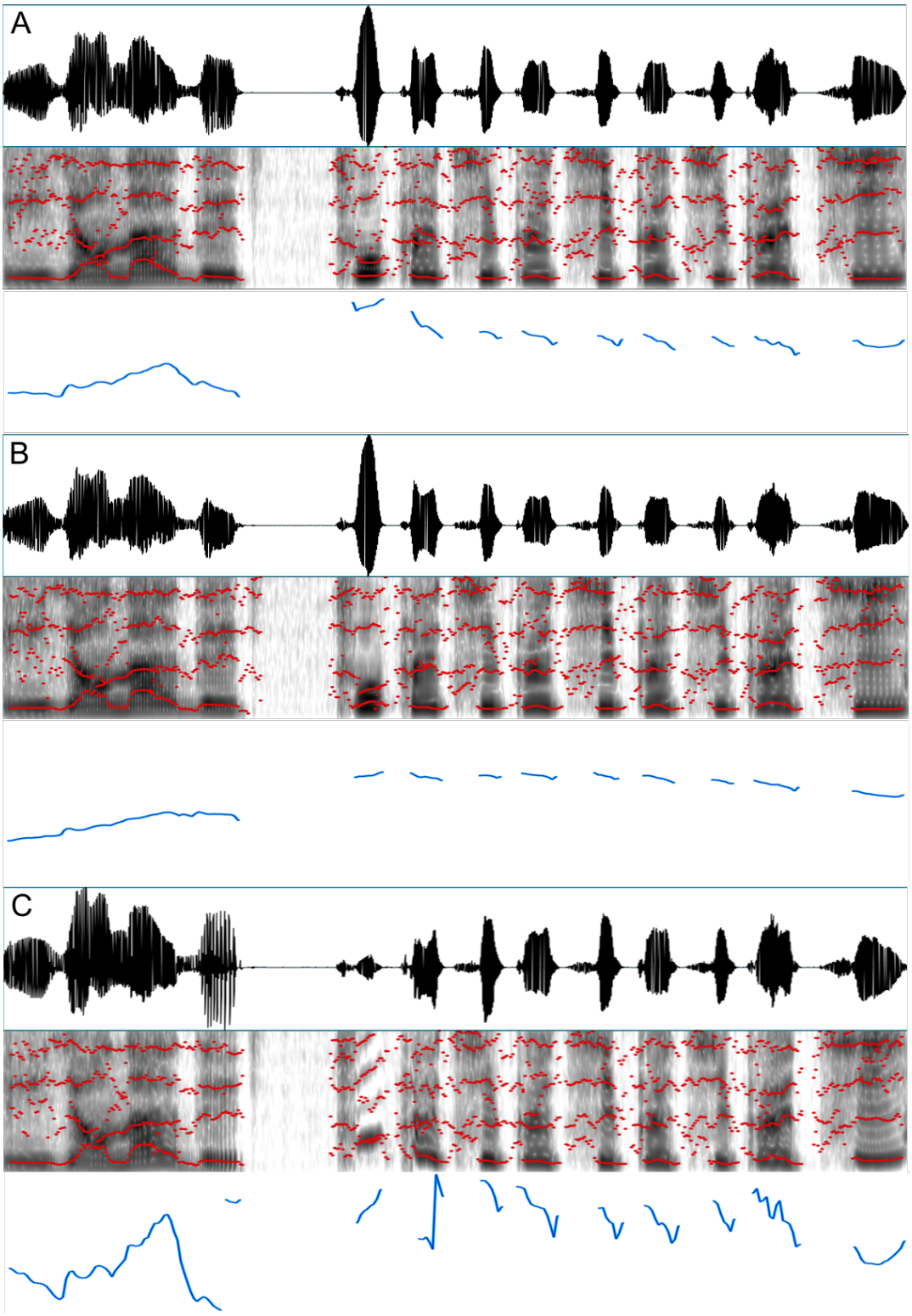


Figure 5.4: Example of a sentence used in experiment 4.b., along with its spectrally distorded analogs. A: Normal. B: FM/4. C: FM*4. Top: sound waveform, middle: spectrogram with formants drawn in red on top, bottom: pitch (F0) contour.

Procedure

The participants were tested directly in their bassinet while naturally asleep. Then the three experimental conditions (normal, FM/4 and FM*4) were presented in a block design: all trials from one condition were presented together on a row. Each condition contained 78 trials, separated by silent ISI of random duration from 500 to 800 milliseconds. The sentences were presented in a random order within blocks. The blocks were separated by 2 minutes of silence with marks sent every 3 to 4 seconds. This experimental design is schematized on Figure 5.1. The experiment was ended if participants showed signs of discomfort or upon parental request.

Data analysis

Event-related potentials The data was pre-processed the same way as in experiment 3.a. As all the stimuli lasted more than 2 seconds in every condition, we computed event-related potentials across longer epochs of -300 to 2000 ms post stimulus onset.

Time-frequency analysis The data was pre-processed the same way as in experiment 3.a. As all the stimuli lasted at least 2 seconds in every condition, we computed time-frequency decompositions across longer epochs of -300 to 1700 ms post stimulus onset.

We then explored the activity evoked by each of the three experimental conditions (i.e. normal speech, 60%-compressed speech and 30%-compressed speech) to the activity measured during the silent periods. To this end, we computed t-values for each time-frequency point in each electrode between silence and one of the three types of sound. The uncorrected significance of each t-value was plotted as "significance maps" for each experimental condition. Although uncorrected, these maps were computed to obtain a visual estimate of the effect of each type of sound as compared to resting-state activity.

5.3.2 Results

Due to the low number of participants included, we here only provide descriptive results.

Event-related potentials

The event-related potentials obtained for each condition in each electrode are plotted on Figure 5.5. Similarly to experiment 3.a, a voltage deflection seems to occur around 300 ms post-stimulus onset in every condition. This deflection is particularly visible for normal speech at T7, for speech with enhanced F0 variations at F3 and Fz, and for speech with reduced F0 range at C3. As for experiment 3.a,

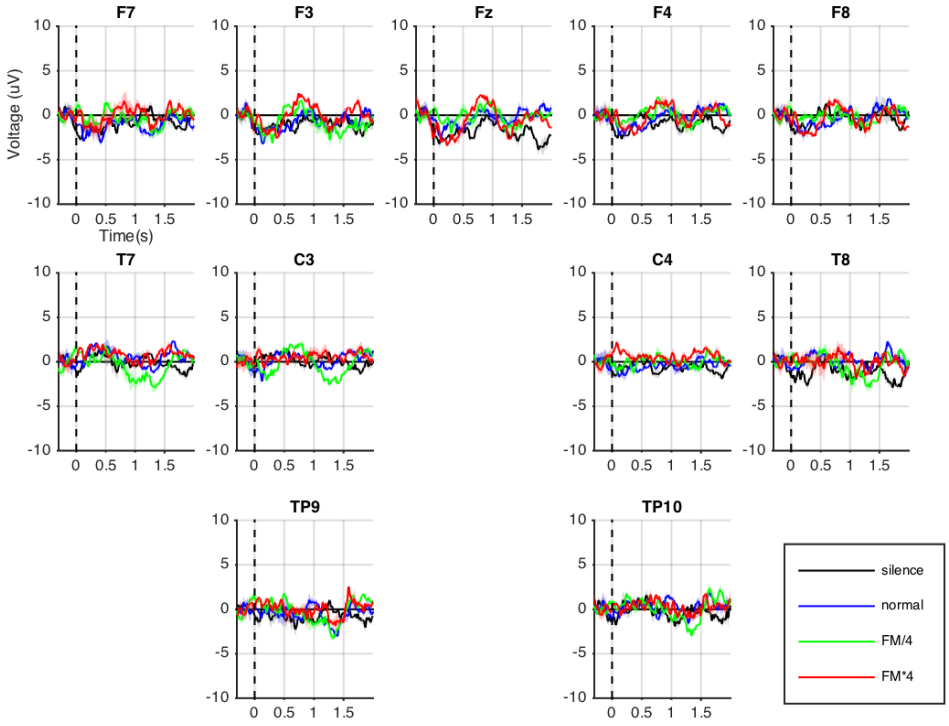


Figure 5.5: Event-related potentials obtained in experiment 3.b. for each F0 range. Shaded areas represent SEM.

voltage deflection are also visible for the silent trials, with similar latencies than for the three speech conditions, in particular at Fz, F3, F4, F7, and F8.

Time-frequency analysis

The time-frequency maps obtained for each condition in each electrode are plotted on Figure 5.6. During the silent period, a transient increase in power as compared to the pre-stimulus period appears to occur around 10 Hz in all the electrodes. A visually similar activity appeared during the presentation of speech with smaller F0 variations, although it may span across a larger frequency band, potentially up to 20 Hz at C3. By contrast, this increase of power was less visible during the presentation of normal speech and speech with increased F0 variations.

5.4 Discussion

In these two experiments, we presented the same sentences with their temporal (exp. 3.a.) or spectral structure (exp 3.b.) compressed or dilated. Although only descriptive, the preliminary results presented here warn several comments.

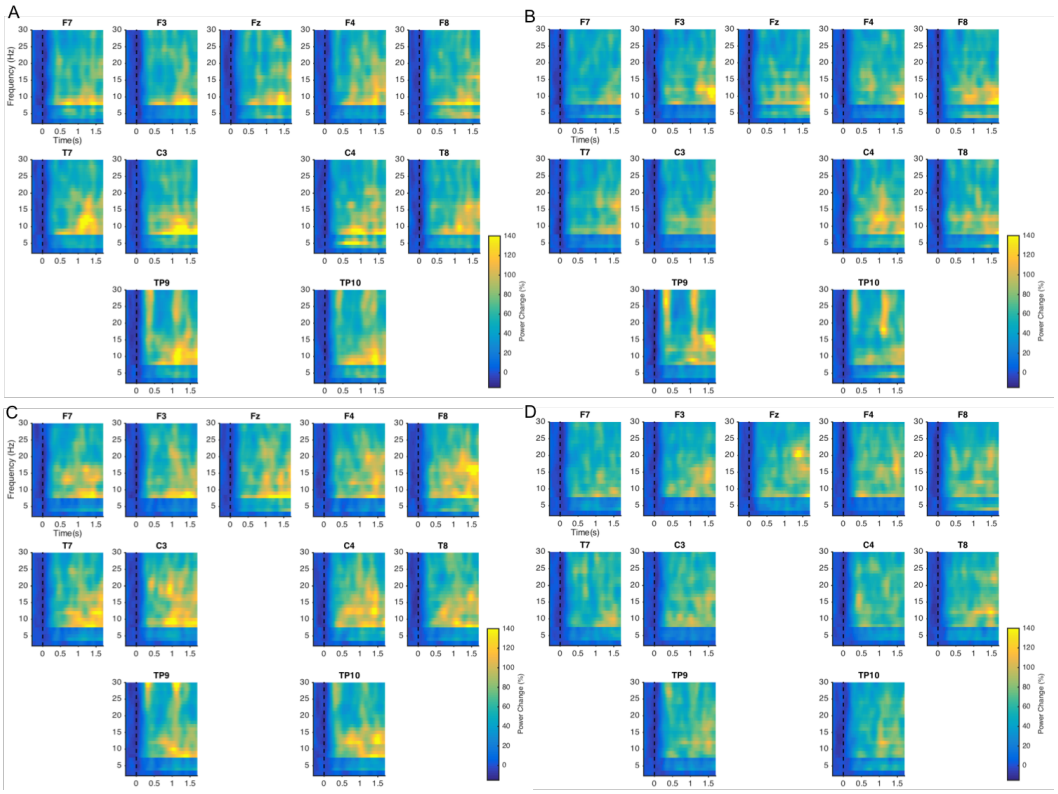


Figure 5.6: Mean power for each time-frequency bin in experiment 3.b. A: silence, B: normal, C: FM/4, D: FM*4.

First, an apparently surprising pattern is the absence of any visible component in the three speech conditions in the two experiments. ERPs have repeatedly observed in newborns (e.g. Cheour et al., 2002; Mahmoudzadeh et al., 2017; Winkler et al., 2009). However, all these experiments used oddball designs to look at mismatch responses. Their stimuli were also short (about a hundred of milliseconds). By contrast, our study used longer stimuli, of more than 600 milliseconds, which didn't contain any embedded event of interest that could have elicited time-locked transient components to be seen in ERPs. We still computed ERPs to check the quality of the signal and as exploratory analysis, but it is not surprising that we didn't observe ERP components.

Then, some activity appears during the silent periods. The participants of the two experiments were sleeping during the whole experiment, which leads to strong activity even in the absence of cognitive activity (e.g. Eisermann et al., 2013). However, given the fact that these portions of the signal were epoched with marks placed at random moments of the silent periods, any cyclic activity should have been cancelled by phase differences when averaging. Moreover, the fact that in experiment 3.a. the largest peaks seem to appear at the most distant electrodes from the reference suggest that these shapes might come from a poorly chosen reference. Although our pre-processing steps were chosen based on the newborn EEG literature, they might not have been optimal for our specific set-up and experiments. As this literature remains sparse, future analysis should compare several references and pre-processing methods to determine the optimal one in this particular population.

Regarding time-frequency analysis, we used complex morlet wavelets as they are the most commonly used method for time-frequency decomposition, and that several studies have used them with newborn data (Isler et al., 2012; Stefanics et al., 2007). However, wavelets require a high signal to noise ratio in the data, and these experiments had at least 100 trials per condition and participants. Our data contained much less trials, as we included participants with a minimum of 40 trials per condition. Hence, other methods might be more appropriate, such as multitapers that tolerates a lower signal-to-noise ratio than wavelets.

For all of these reasons, and in particular the presence of electrical activity during the silent periods, the data presented here prevent us to draw any conclusion regarding electrophysiological activity supporting encoding of speech sounds at birth.

Part III

General Discussion

Chapter 6

Theoretical discussion

This thesis explored how the newborn brain encodes speech with variable acoustic parameters, both in the temporal and the spectral dimension. To this end, we recorded hemodynamic localized responses with NIRS, as well as electrophysiological, temporally precise activation with EEG to temporally or spectrally distorted utterances. In the first experiment, we presented normal speech, speech compressed to 60% of its initial duration, and speech compressed to 30% of its initial duration. We recorded the hemodynamic response to these stimuli over the frontal, temporal and parietal cortices using NIRS. The results show no difference between normal and 60%-compressed speech, but differential responses between normal and 30%-compressed speech as well as between 60%- and 30%-compressed speech in the right frontal, right temporal, and left temporo-parietal cortices were observed. In a second set of experiments, we asked whether this ability relies on prenatal experience with the prosodic pattern of the native language, already experienced in utero. We ran experiments with designs identical to the first one using stimuli in two unfamiliar languages. One group of newborns heard utterances in a language, Spanish, that is rhythmically similar to the native language. Another group of newborns heard a language, English, that is rhythmically different from French. No difference between the three compression rates was observed in Spanish. In English, only 30%-compressed speech evoked significant responses in a left temporo-parietal region also activated for French, but the exact pattern of activations was different from those for French. In the last two experiments, we explored newborns' brain responses to temporally or spectrally distorted speech sounds using EEG. In one experiment, we presented the same time-compressed speech as in the first NIRS study. In the other experiment, we tested speech in French with normal, enhanced (FM*4) or reduced (FM/4) frequency modulations. We now discuss the general implications of these results in a broader theoretical framework.

6.1 Limits in auditory processing

The results from Experiment 1-2 show that highly-compressed speech evoked a different response that normal speech in the temporo-parietal cortex of newborns, while normal and moderately compressed speech did not. This was the case when participants heard their native language (French), as well as when they heard an unfamiliar language (English). These results show that even in the absence of substantial linguistic expertise and experience with broadcast speech, the newborn brain processes speech at a high temporal scale differently than normal speech. By contrast, the lack of differences between responses to normal and moderately compressed speech shows that the newborn brain processes speech in a stable, invariant way within a certain time range. Interestingly, this time range appears to be the same as in adults, and doesn't depend on the language heard. (We did observe differences between the responses to French and English, which will be discussed later, but crucially the two patterns of results did not differ in the compression rate that was processed differently.) This suggests that there is a temporal limit on speech processing. "Bottleneck" models of speech-processing have been proposed in the literature, with studies showing that sensory cortical areas respond to oral or written language independently of time-compression, while other modality-general linguistic areas cease to respond below a certain time-scale (i.e. above a certain level of time-compression) (Vagharchakian et al., 2012). Consistently, other studies showed that θ oscillations follow the slow AM of the speech signal independently of comprehension (Pefkou et al., 2017; Howard and Poeppel, 2010; Nourski et al., 2009). Instead, the drop in comprehension was associated with a drop in β-power approximately 1.5 seconds after the sentence onset, suggesting that this limit comes from a disruption of top-down predictive mechanisms (Pefkou et al., 2017). When speech rate is too fast, the time between successive information packets is too short for the speech-perception system to read the buffer and dynamically deploy predictions. Our results provide further support for this model, but also complement it. Indeed, the results we obtained in the three experiments presented in chapter 3 and chapter 4 show that this limit does not operate at a linguistic level (e.g. semantic or syntactic, as claimed in Vagharchakian et al., 2012), since newborns do not have these linguistic representations, yet they show the same processing limit. Rather, this limit would come from general properties of auditory encoding already present prior to comprehension, i.e. in newborns. Thus, the data presented in chapter 3 and chapter 4 suggest that the drop in comprehension observed in adults would be a consequence of more general limits in auditory processing.

This limit on the spectro-temporal parameters that the auditory system can process is consistent with the acoustical properties of vocal and communication sounds. Typical human screams, speech, or bird songs occupy a certain part of the modulation spectrum (Arnal et al., 2015; Singh and Theunissen, 2003), meaning that they are contained within a certain range of temporal and frequency modulations. Accordingly, the avian auditory cortex preferentially responds to sounds contained in the corresponding region of the spectrum. The spectro-temporal tun-

ing of the neurons in the avian auditory cortex match the statistical structure (i.e. the spectro-temporal modulations) of various natural sounds, including conspecific songs. Interestingly, this tuning overlaps with the regions of the spectrum that differ the most between sound classes (e.g. songs and other types of natural sounds), providing a coding strategy to identify the various songs as exemplars of a unified category (Woolley et al., 2005). Similarly, it is possible that the human brain is tuned to the spectro-temporal modulations of speech, consistently to the idea of an acoustic subspace for human vocal sounds, such as screams (Arnal et al., 2015), and speech. Normal and moderately compressed speech would be in the range of spectro-temporal modulations typical of speech, but the higher compression rate would be out of it, hence wouldn't elicit the same type of processing. Importantly, no causal relationship is implied: sensory systems might have evolved to adapt to the statistical structure of natural stimuli, but communication systems might also have evolved within the biological constraints of the organisms that use them.

The auditory system is thought to respond preferentially to natural sounds (Barlow, 1961; Mizrahi et al., 2014), among them speech. As speech sounds are variable, humans need to be sensitive to the range of acoustic parameters that speech can take. The auditory code needs to be flexible towards the broad range of parameters than speech sounds can take, consistent with the idea that the higher levels of processing in the auditory system are sensitive to the presence of increasingly more abstract auditory entities, rather than absolute spectro-temporal parameters (Mizrahi et al., 2014). The fact that we couldn't find any difference between the responses to normal and moderately compressed speech in the newborn cortex meshes well with this proposal. Our data also go in the same direction as some animal studies. In the ferret primary auditory cortex, neurons respond to a range of spectro-temporal parameters corresponding to the ones present in the natural soundscape. These enhanced responses are visible up to certain limits: as for our data, when the stimulus parameters are too far away from the natural values, the responses become more variable and smaller (Garcia-Lazaro et al., 2006). The consistency between our data on human newborns listening to speech, and data on other species with other kinds of stimuli support the hypothesis that a core property of the auditory cortex is that it is tuned to a specific range of statistical parameters corresponding to the ones found in natural sounds, among them speech. This range is limited: when stimulus parameters stray too far from the typical ones, the brain produces weaker or different responses, as for highly compressed speech in Experiments 1 & 2.b.

The fact that we find invariant responses in the cortex suggests that these are higher, more abstract level responses. This pattern of results is consistent with hierarchical models of auditory coding that state that auditory processing occurs in three stages, first with extraction of spectro-temporal features in the lower structures, then grouping of features to form an auditory object in the thalamus and auditory cortex, and finally cognitive processes in higher-level cortices (e.g. Nelken, 2008, 2004). Importantly, the auditory objects formed in the thalamus and the cortex are abstract: their coding tolerates variations in their spectro-temporal

structure (Blackwell et al., 2016; Carruthers et al., 2015). Similar arguments have been made in the speech processing domain. These models argue that abstract speech units are formed at a pre-lexical level (Obleser and Eisner, 2009). The similarity between our data and those obtained in adults (e.g. Peelle et al., 2004; Vagharchakian et al., 2012) provides further evidence for these models. In our case, speech appears to be coded as an abstract auditory object in the newborn brain, with a representation that is independent of the temporal scale as long as it remains within a certain range. Although we don't know precisely the nature of these representations, and in particular if they correspond to discrete speech units or a global auditory object, it is still the case that the lack of differences between normal and 60%-compressed speech indicates that the newborn brain encodes speech beyond its temporal parameters, hence as an abstract representation.

6.2 Influence of prenatal experience

Although the responses obtained for the native and the unfamiliar language follow a similar pattern, interesting differences between the two languages also exist. In French, the highly compressed stimuli evoked a non-canonical, inverted response, while in English, this condition triggered a more increased response than the other two compression rates. One possibility is that the inverted response we obtained for the native language reflects an effortful and unsuccessful processing of a stimulus that should otherwise (in the uncompressed and moderately compressed conditions) be processed normally leading not only to auditory, but also to linguistic representations in the native language (e.g. well-formed pitch contours or other prosodic units). By contrast, newborns most likely do not process an unfamiliar language beyond the basic auditory level, which explains the weak activation to the uncompressed and moderately compressed condition in the unfamiliar language. The increased activation to the highly compressed condition in this case would reflect its auditory 'ill-formedness' or unnaturality. This interpretation is all the more plausible for an unfamiliar language that is rhythmically different from the native language. As newborns are known to discriminate between rhythmically different, but not between rhythmically similar languages (e.g. Mehler et al., 1988; Nazzi et al., 1998), their prenatal experience with speech rhythm may benefit the processing of an unfamiliar language that is rhythmically similar to the native one. Had the results with Spanish been conclusive, we could have directly confirmed this interpretation. In case of a rhythmically different language, no such advantage may be expected. It is thus not unlikely that French newborns may not process English stimuli beyond the basic auditory level.

Several lines of evidence support the idea that prenatal experience with speech shapes auditory perception at birth. For instance, monolingual and bilingual newborns were presented with tone pairs in which the members differed along an acoustic dimension (pitch, duration or intensity) that either matched or did not match the acoustic cues carrying prosodic prominence in the language(s) the infants heard in utero. The newborns' left temporo-parietal and right temporal

cortices responded more to sound patterns that were inconsistent with the patterns found in the infants' native language, i.e. were unfamiliar, as compared to patterns that were consistent with it, i.e. were familiar. Importantly, monolinguals responded only for the acoustic dimension used in their native language (French), whereas bilinguals responded to the other acoustic dimensions (Abboub et al., 2016). Similarly, newborns responded to a pitch change in trisyllabic sequences only when they were exposed to these stimuli during their intra-uterine life, inducing sensitivity to pitch in speech streams although this cue not used in their native language (Finnish) (Partanen et al., 2013). Finally, another experiment compared responses to forward (FW) and backward (BW) English and FW and BW Spanish in English-exposed newborns. Forward speech evoked a larger response than backward speech in the left fronto-temporal cortex in the native language (English), but not in a foreign language (Spanish) (May et al., 2018). Comparing the native language, English, with a non-native language, Tagalog, in a similar paradigm, but using low-pass filtered speech, a direction-insensitive response to English was observed. By contrast, Tagalog, evoked different responses when played forward and backwards (May et al., 2011). These results converge with the ones obtained in this thesis: speech processing is influenced by prenatal experience with the native language. In our experiments, newborns showed more robust processing (more typical hemodynamic responses) to native than to non-native stimuli. Consequently, their ability to process speech similarly across several speech rates interacted with the speech they experienced in-utero.

The fact that the native language activated a wider cortical network than the unfamiliar language is consistent with the idea that early sensory stimulation, including intra-uterine life, impacts brain development and function. On a structural level, preterm newborns exposed to recordings of maternal physiological noise in the neonatal intensive care unit presented a thicker and larger auditory cortex than infants only exposed to the electronic sounds of hospital devices (Webb et al., 2015). Animal studies also support the idea that the spectro-temporal structure of sounds experienced early influences the functional development of the auditory cortex. Rats reared in an acoustic environment made of pulsed tones had deteriorated tonotopicity in their auditory cortex (Zhang et al., 2001). In zebra finches, noise-rearing and social isolation led to a decrease in response strength and in selectivity of the neural responses to song in the primary auditory cortex as compared to birds reared in a normal social environment (Amin et al., 2013). Although these studies contrasted acoustic environments with coarse differences, they suggest that cortical activity may be influenced by prenatal auditory experience. In our case, the temporal structure of speech heard prenatally may have tuned the newborn brain to these patterns.

6.3 Coherence with models of neural coding

The results presented in this thesis show that speech sounds elicit similar responses in the newborn brain when they are presented at their normal duration and when

they are moderately time-compressed, whether it is in the participants' native language or in an unfamiliar language. This suggests that the newborn brain encodes speech sounds in a stable manner across a range of temporal scales. Similar neural responses across different values on a physical dimension suggest a scale-invariant neural encoding. This, in turn, is consistent with the efficient neural coding hypothesis. The efficient coding hypothesis states that the perceptual systems have evolved to encode environmental signals in the most efficient, i.e. information theoretically optimal, least redundant way possible (Atick, 1991; Barlow, 1961). According to this theory, a stimulus is encoded in a concise way to remove redundancies present in the signal. Natural sounds, among them speech, obey statistical rules that make pieces of the signal predictable from the others. This makes the signal redundant. Considering the importance of natural stimulus in our everyday environment, and in particular the importance of speech for humans, the auditory system must provide a representation that is sufficiently rich to allow complex responses with limited brain resources. To maximize the rate of information in the representation, the auditory system should tune to their statistical structure to focus on the most informative regions of the spectro-temporal space. This way, responding preferentially to the acoustical properties of natural sounds allows to encode the spectro-temporal modulations that differ most across sounds and allow the better discrimination, object identification etc... (Elliott and Theunissen, 2009; Woolley et al., 2005).

In the visual domain, the efficient neural coding hypothesis is well established (Simoncelli and Olshausen, 2001). In the auditory domain, increasing evidence suggests that the auditory neural code is consistent with an optimal representation of the sound under sparse coding assumptions (Lewicki, 2002; Smith and Lewicki, 2006). Mathematically, this theoretical framework models neural coding by deriving maximally non-redundant representations such as independent-component analysis (Simoncelli and Olshausen, 2001), which maximizes sparseness between dimensions, in our case coding units. In the mouse inferior colliculus, single neurons respond to different acoustic parameters, and their firing rate and timing are affected by variations of these parameters, but at the population level activity is stable over variations. Crucially, the information conveyed by neurons responding to vocalizations was higher than for simulated data with more redundancy (Holmstrom et al., 2010). Evidence for an efficient auditory code has also been provided in the human planum temporale. In this study, pitch sequences with higher levels of entropy (hence lower levels of predictability) evoked higher hemodynamic responses in the planum temporale of human participants, showing that they engaged more demanding neuronal activity. This suggests that they used an efficient neural code that exploited the predictability of the sequences, probably through a sparse representation of the sounds (Overath et al., 2007). When they listen to music, listeners are able to predict up-coming musical events based on the music spectro-temporal structure (Rankin et al., 2014). Similarly, in our experiments, the newborn brain might have adapted to the statistical regularities in speech over different time-scales, and derived an efficient neural code for speech.

The fact that there is a limit (too high compression) is suggestive of an efficient code of speech sounds: if the capacity of auditory perception is limited, then the signal (i.e. speech) should be encoded in a compact way to fit with the observed bottleneck.

Chapter 7

Perspectives

The present thesis reports several experiments conceived to explore the ability of the neonatal brain to deal with acoustic variability in speech. As for every experimental project, this implied making choices to keep the experiments feasible. In this chapter we suggest further studies that would nicely complement and expand the ones presented in the experimental chapters of this thesis.

7.1 Replications with simpler experimental designs

A first follow-up study could replicate the Spanish and English experiment with a classical block design to have more trials and a higher signal-to-noise ratio. These experiments would not allow to replicate the difference between alternating and non-alternating blocks, hence would not allow to show discrimination between normal and time-compressed sentences. However, discrimination was not the main purpose of our studies. We had initially chosen this design to be able to show subtle differences between the processing of normal speech and speech compressed at different rates, with the reasoning that, based on the adult literature, newborn brain may produce similar responses to the different compression rates. Given the difficulty to formally show similarity between two data arrays, a null results in a block design experiment would have been difficult to interpret. We chose this design to be able to test both processing and discrimination, similarly to previous studies in infants (Gervain et al., 2012, 2016; Edwards et al., 2016). This is indeed what happened for the experiments in unfamiliar languages (see chapter 4). By contrast, the experiment in the participants' native language did show different responses between normal and highly-compressed speech on the one hand, and between moderately and highly-compressed speech. It would therefore be interesting to replicate this with a more classic block design, and test whether similar results between the three compression rates could be obtained in the unfamiliar languages with a larger number of trials.

7.2 Generalization to other acoustical parameters

Regarding experimental conditions, we restricted to a limited set of acoustical manipulations. In the time domain, our investigations were restricted to time-compression, following the rich literature on time-compressed speech (e.g. Peelle et al., 2004; Ahissar et al., 2001; Adank and Janse, 2009; Howard and Poeppel, 2010). However, variability in speech rate is also manifested as slowdown. A complete investigation of adaptation to temporal variability in speech would require to explore responses to dilated speech. Two studies have consistently demonstrated that brain activity rescales to speech sounds for both compression and dilation (Lerner et al., 2014; Davidesco et al., 2018). The consistency of our results to the one obtained in adults suggest that similar mechanisms take place in infants. Still, this remains to be demonstrated. By contrast, for spectral variability, we used both dilation and compression of the F0 range. Our experiments therefore lacked the possibility to investigate the limits to the ability to deal with spectral variability in speech. It is possible that, as for time-compression, the brain is able to extract an invariant representation of speech, but only up to a certain level of compression or dilation. Follow-up experiments should use several level of spectral dilation and compression, to test the extent this variability can be managed, and whether the ranges are the same as for time-compression (i.e. about 1/3). This would indicate a general mechanism, consistently with models of auditory coding that postulate partly common codes for frequency and amplitude modulations (e.g. through phase-locking) (Luo et al., 2006). We also restricted our stimuli to a subset of spectral parameters. We varied only the range of frequency modulations, but speech varies in many other spectral parameters (see section 1.1). In particular, it would be interesting to vary the mean formant frequencies of the speech tokens. This would be particularly relevant in newborns, as they experienced only low-passed filtered sounds during their intra-uterine life (Abrams and Gerhardt, 2000). They hear high-frequency sounds for the first time, and might therefore have more difficulty to adapt to a shift to higher mean formant frequencies.

7.3 Brain networks supporting adaptation to spectral variability

Our experiment manipulating the range of FM used EEG. The aim of this experiment was to study the electrophysiological mechanisms supporting adaptation to spectral variability, but couldn't provide any information about the cortical regions where these mechanisms take place. Therefore, to complement this experiment, it would be judicious to replicate the experiment using NIRS.

Moreover, this experiment could tell whether adaptation to spectral and temporal variability are supported by the same cortical networks. Adult studies have provided evidence that spectral and temporal processing may occur in homologue

regions but with different lateralization (Zatorre and Belin, 2001). This organization may be present readily from birth, but this may also interact with developmental processes. The potential lateralization of speech processing networks remains controversial in infants, with some studies showing lateralization to the left (Homae et al., 2014), whereas others show a right lateralization. Moreover, the recruited network may vary depending on whether the stimuli were tones or speech sounds. Further studies are necessary to fully investigate these questions.

7.4 Relationship between limits in auditory processing and the spectrum of natural speech

In this thesis, we have shown that the neonatal brain responds differently to normal and moderately compressed speech than to highly-compressed speech (experiments 1 and 2.b.). Is this limit observed in auditory perception of speech correlated with the acoustical structure of speech? It might be that the limit observed (30% of initial duration in the time-domain) corresponds to the border of the modulations observed in natural speech sounds. If the auditory system is adapted or responds preferentially to the statistical structure of natural sounds, then it should not respond to the acoustical values that are outside the typical ones for natural sounds, among them speech. In our case, it is possible that 30%-compressed speech lies outside the typical spectrum of speech, leading to a reduced response of the auditory system. A similar hypothesis has been proposed, namely that the limit in the ability to understand speech above a certain compression rate is correlated with the upper limit of theta oscillations (Ghitza and Greenberg, 2009; Ghitza, 2014). This hypothesis is based on the parallel between the typical syllabic rate in speech and the frequency of theta oscillations: the latter can phase-lock to the syllabic rhythm up to their maximal frequency, determining an auditory-channel capacity for speech (Ghitza, 2014). Although correlated to acoustical modulations, syllables are already linguistic units. We suggest that this relationship could be even more fundamental with auditory processing limited to the range of modulations itself that natural speech sounds can take, without considering any perceptual or linguistic entities. Acoustical analysis of the stimuli used in this thesis are necessary to support this hypothesis.

Bibliography

Abboub, N., Nazzi, T., and Gervain, J. (2016). Prosodic grouping at birth. *Brain and Language*, 162:46–59.

Abdala, C. and Keefe, D. H. (2012). Morphological and Functional Ear Development. In Werner, L., Fay, R. R., and Popper, A. N., editors, *Human Auditory Development*, volume 42, pages 19–59. Springer New York, New York, NY.

Adank, P. and Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *Neuroimage*, 49(1):1124–1132.

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., and Merzenich, M. M. (2001). 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):13367–13372.

Alho, K., Sainio, K., Sajaniemi, N., Reinikainen, K., and Näätänen, R. (1990). Event-related brain potential of human newborns to pitch change of an acoustic stimulus. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 77(2):151–155.

Amin, N., Gastpar, M., and Theunissen, F. E. (2013). Selective and Efficient Neural Coding of Communication Signals Depends on Early Acoustic and Social Environment. *PLOS ONE*, 8(4):e61417.

Andermann, M., Patterson, R. D., Vogt, C., Winterstetter, L., and Rupp, A. (2017). Neuromagnetic correlates of voice pitch, vowel type, and speaker size in auditory cortex. *NeuroImage*, 158:79–89.

Anders, T., Emde, R., and Parmelee, A. A. (1971). A standardized terminology, techniques and criteria for scoring states of sleep and wakefulness. *UCLA Brain Information Service*.

Arichi, T., Fagiolo, G., Varela, M., Melendez-Calderon, A., Allievi, A., Merchant, N., Tusor, N., Counsell, S. J., Burdet, E., Beckmann, C. F., and Edwards, A. D. (2012). Development of BOLD signal hemodynamic responses in the human brain. *Neuroimage*, 63(2):663–673.

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A.-L., and Poeppel, D. (2015). Human Screams Occupy a Privileged Niche in the Communication Soundscape. *Current Biology*, 25(15):2051–2056.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, 40(3):351–373.

Atick, J. J. (1991). Could information theory provide an ecological theory of sensory processing? *Network (Bristol, England)*, 22(1-4):4–44.

Atkinson, J. E. (1978). Correlation analysis of the physiological factors controlling fundamental voice frequency. *The Journal of the Acoustical Society of America*, 63(1):211–222.

Banse, R. and Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3):614–636.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory Communication*, pages 217–234.

Bartocci, M., Winberg, J., Ruggiero, C., Bergqvist, L. L., Serra, G., and Lagercrantz, H. (2000). Activation of Olfactory Cortex in Newborn Infants After Odor Stimulation: A Functional Near-Infrared Spectroscopy Study. *Pediatric Research*, 48(1):18–23.

Benavides-Varela, S. and Gervain, J. (2017). Learning word order at birth: A NIRS study. *Developmental Cognitive Neuroscience*, 25:198–208.

Benavides-Varela, S., Hochmann, J.-R., Macagno, F., Nespor, M., and Mehler, J. (2012). Newborn's brain activity signals the origin of word memories. *Proceedings of the National Academy of Sciences of the United States of America*, 109(44):17908–17913.

Benavides-Varela, S., Siugzdaite, R., Gómez, D. M., Macagno, F., Cattarossi, L., and Mehler, J. (2017). Brain regions and functional interactions supporting early word recognition in the face of input variability. *Proceedings of the National Academy of Sciences of the United States of America*, 114(29):7588–7593.

Bertoncini, J., Floccia, C., Nazzi, T., and Mehler, J. (1995). Morae and Syllables: Rhythmical Basis of Speech Representations in Neonates. *Language and Speech*, 38(4):311–329.

Best, C. and Jones, C. (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behavior and Development*, 21:295.

Birnholz, J. C. and Benacerraf, B. R. (1983). The development of human fetal hearing. *Science*, 222(4623):516–518.

Blackwell, J. M., Taillefumier, T. O., Natan, R. G., Carruthers, I. M., Magnasco, M. O., and Geffen, M. N. (2016). Stable encoding of sounds over a broad range of statistical parameters in the auditory cortex. *The European Journal of Neuroscience*, 43(6):751–764.

Bouchon, C., Nazzi, T., and Judit Gervain (2015). Hemispheric Asymmetries in Repetition Enhancement and Suppression Effects in the Newborn Brain. *PLOS ONE*, 10(10):e0140160.

Buckner, R. L. and Koutstaal, W. (1998). Functional neuroimaging studies of encoding, priming, and explicit memory retrieval. *Proceedings of the National Academy of Sciences*, 95(3):891–898.

Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience*, 13(6):407–420.

Buzsaki, G., Bickford, R. G., Ponomareff, G., Thal, L. J., Mandel, R., and Gage, F. H. (1988). Nucleus basalis and thalamic control of neocortical activity in the freely moving rat. *Journal of Neuroscience*, 8(11):4007–4026.

Carruthers, I. M., Laplagne, D. A., Jaegle, A., Briguglio, J., Mwilambwe-Tshilobo, L., Natan, R. G., and Geffen, M. N. (2015). Emergence of invariant representation of vocalizations in the auditory cortex. *Journal of Neurophysiology*, page jn.00095.2015.

Chandrasekaran, B., Chan, A. H. D., and Wong, P. C. M. (2011). Neural Processing of What and Who Information in Speech. *Journal of Cognitive Neuroscience*, 23(10):2690–2700.

Cheour, M., Alho, K., Čeponiené, R., Reinikainen, K., Sainio, K., Pohjavuori, M., Aaltonen, O., and Näätänen, R. (1998). Maturation of mismatch negativity in infants. *International Journal of Psychophysiology*, 29(2):217–226.

Cheour, M., Čėponiené, R., Leppänen, P., Alho, K., Kujala, T., Renlund, M., Fellman, V., and Näätänen, R. (2002). The auditory sensory memory trace decays rapidlyin newborns. *Scandinavian Journal of Psychology*, 43(1):33–39.

Chiba, T. and Kajiyama, M. (1941). *The vowel: its nature and structure*. Tokyo-Kaiseikan. Google-Books-ID: tpkKAAAAMAAJ.

Cohen, M. X. (2014). *Analyzing Neural Time Series Data: Theory and Practice*. Issues in Clinical and Cognitive Neuropsychology. MIT Press, Cambridge, Mass.

Cristia, A., Dupoux, E., Hakuno, Y., Lloyd-Fox, S., Schuetze, M., Kivits, J., Bergvelt, T., Gelder, M. v., Filippin, L., Charron, S., and Minagawa-Kawai, Y. (2013). An Online Database of Infant Functional Near InfraRed Spectroscopy Studies: A Community-Augmented Systematic Review. *PLOS ONE*, 8(3):e58906.

Cristia, A., Minagawa-Kawai, Y., Egorova, N., Gervain, J., Filippin, L., Cabrol, D., and Dupoux, E. (2014). Neural correlates of infant accent discrimination: an fNIRS study. *Developmental Science*, 17(4):628–635.

Csibra, G., Davis, G., Spratling, M. W., and Johnson, M. H. (2000). Gamma Oscillations and Object Processing in the Infant Brain. *Science*, 290(5496):1582–1585.

Csibra, G., Henty, J., Volein, À., Elwell, C., Tucker, L., Meek, J., and Johnson, M. H. (2004). Near infrared spectroscopy reveals neural activation during face perception in infants and adults. *Journal of Pediatric Neurology*, 02(02):085–089.

Davidesco, I., Thesen, T., Honey, C. J., Melloni, L., Doyle, W., Devinsky, O., Ghitza, O., Schroeder, C., Poeppel, D., and Hasson, U. (2018). Electrocorticographic responses to time-compressed speech vary across the cortical auditory hierarchy. *bioRxiv*, page 354464.

Davis, P. A. (1939). Effects of acoustic stimuli on the waking human brain. *Journal of Neurophysiology*, 2(6):494–499.

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

Dehaene-Lambertz, G., Dehaene, S., and Hertz-Pannier, L. (2002). Functional Neuroimaging of Speech Perception in Infants. *Science*, 298(5600):2013–2015.

Dehaene-Lambertz, G. and Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *NeuroReport*, 12(14):3155.

Delorme, A. and 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.

Di Liberto, G. M., O'Sullivan, J. A., and Lalor, E. C. (2015). Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. *Current Biology*, 25(19):2457–2465.

Diallo, M. M. (2017). *Problème inverse de sources en Electro-Encéphalo-Graphie chez le nouveau-né*. PhD Thesis, Université de Picardie-Jules Verne.

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., and Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews*, 81:181–187.

Ding, N., Simon, J. Z., Shamma, S. A., and David, S. V. (2016). Encoding of natural sounds by variance of the cortical local field potential. *Journal of Neurophysiology*, 115(5):2389–2398.

Dudley, H. (1940). The Carrier Nature of Speech. *Bell System Technical Journal*, 19(4):495–515.

Dupoux, E. and Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance*, 23(3):914–927.

Ecklund-Flores, L. and Turkewitz, G. (1996). Asymmetric headturning to speech and nonspeech in human newborns. *Developmental Psychobiology*, 29(3):205–217.

Edwards, L. A., Wagner, J. B., Simon, C. E., and Hyde, D. C. (2016). Functional brain organization for number processing in pre-verbal infants. *Developmental Science*, 19(5):757–769.

Eggermont, J. J. and Moore, J. K. (2012). Morphological and Functional Development of the Auditory Nervous System. In Werner, L., Fay, R. R., and Popper, A. N., editors, *Human Auditory Development*, volume 42, pages 61–105. Springer New York, New York, NY.

Eisermann, M., Kaminska, A., Moutard, M. L., Soufflet, C., and Plouin, P. (2013). Normal EEG in childhood: From neonates to adolescents. *Neurophysiologie Clinique/Clinical Neurophysiology*, 43(1):35–65.

Ellingson, R. J. and Peters, J. F. (1980). Development of EEG and daytime sleep patterns in normal full-term infants during the first 3 months of life: Longitudinal observations. *Electroencephalography and Clinical Neurophysiology*, 49(1):112–124.

Elliott, T. M. and Theunissen, F. E. (2009). The Modulation Transfer Function for Speech Intelligibility. *PLoS Comput Biol*, 5(3):e1000302.

Emberson, L. L., Boldin, A. M., Riccio, J. E., Guillet, R., and Aslin, R. N. (2017a). Deficits in Top-Down Sensory Prediction in Infants At Risk due to Premature Birth. *Current Biology*, 27(3):431–436.

Emberson, L. L., Cannon, G., Palmeri, H., Richards, J. E., and Aslin, R. N. (2017b). Using fNIRS to examine occipital and temporal responses to stimulus repetition in young infants: Evidence of selective frontal cortex involvement. *Developmental Cognitive Neuroscience*, 23:26–38.

Emberson, L. L., Crosswhite, S. L., Goodwin, J. R., Berger, A. J., and Aslin, R. N. (2016). Isolating the effects of surface vasculature in infant neuroimaging using short-distance optical channels: a combination of local and global effects. *Neurophotonics*, 3(3):031406–031406.

Emberson, L. L., Richards, J. E., and 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.

Fant, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings. *Speech Transmission Laboratory - Quarterly Progress and Status Report*, 7(4):13.

Fant, G. (1971). *Acoustic Theory of Speech Production: With Calculations based on X-Ray Studies of Russian Articulations*. Walter de Gruyter. Google-Books-ID: UY0iAAAAQBAJ.

Fant, G. (1975). Non-uniform vowel normalization. *Speech Transmission Laboratory - Quarterly Progress and Status Report*, 16(2-3):1–19.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B. d., and Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants*. *Journal of Child Language*, 16(3):477–501.

Ferry, A. L., Fló, A., Brusini, P., Cattarossi, L., Macagno, F., Nespor, M., and 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.

Formisano, E., Martino, F. D., Bonte, M., and Goebel, R. (2008). "Who" Is Saying "What"? Brain-Based Decoding of Human Voice and Speech. *Science*, 322(5903):970–973.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2):78–84.

Friederici, A. D., Friedrich, M., and Weber, C. (2002). Neural manifestation of cognitive and precognitive mismatch detection in early infancy. *NeuroReport*, 13(10):1251.

Friedrich, M. and Friederici, A. D. (2004). N400-like Semantic Incongruity Effect in 19-Month-Olds: Processing Known Words in Picture Contexts. *Journal of Cognitive Neuroscience*, 16(8):1465–1477.

Fujimura, O. (1962). Analysis of Nasal Consonants. *The Journal of the Acoustical Society of America*, 34(12):1865–1875.

Fukui, Y., Ajichi, Y., and Okada, E. (2003). Monte Carlo prediction of near-infrared light propagation in realistic adult and neonatal head models. *Applied Optics*, 42(16):2881–2887.

Garcia-Lazaro, J. A., Ahmed, B., and Schnupp, J. W. H. (2006). Tuning to Natural Stimulus Dynamics in Primary Auditory Cortex. *Current Biology*, 16(3):264–271.

Gargiulo, P., Belfiore, P., Friðgeirsson, E., Vanhatalo, S., and Ramon, C. (2015). The effect of fontanel on scalp EEG potentials in the neonate. *Clinical Neurophysiology*, 126(9):1703–1710.

Gerhardt, K. J., Otto, R., Abrams, R. M., Colle, J. J., Burchfield, D. J., and Peters, A. J. M. (1992). Cochlear microphonics recorded from fetal and newborn sheep. *American Journal of Otolaryngology*, 13(4):226–233.

Gervain, J., Berent, I., and 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.

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

Gervain, J. and Mehler, J. (2010). Speech Perception and Language Acquisition in the First Year of Life. *Annual Review of Psychology*, 61(1):191–218.

Gervain, J., Mehler, J., Werker, J. F., Nelson, C. A., Csibra, G., Lloyd-Fox, S., Shukla, M., and Aslin, R. N. (2011). Near-infrared spectroscopy: A report from the McDonnell infant methodology consortium. *Developmental Cognitive Neuroscience*, 1(1):22–46.

Gervain, J., Werker, J. F., Black, A., and Geffen, M. N. (2016). The neural correlates of processing scale-invariant environmental sounds at birth. *NeuroImage*.

Gervain, J., Werker, J. F., and Geffen, M. N. (2014). Category-Specific Processing of Scale-Invariant Sounds in Infancy. *PLoS ONE*, 9(5):e96278.

Ghitza, O. (2014). Behavioral evidence for the role of cortical theta oscillations in determining auditory channel capacity for speech. *Frontiers in Psychology*, 5.

Ghitza, O. and Greenberg, S. (2009). On the Possible Role of Brain Rhythms in Speech Perception: Intelligibility of Time-Compressed Speech with Periodic and Aperiodic Insertions of Silence. *Phonetica*, 66(1-2):113–126.

Giordano, B. L., Ince, R. A. A., Gross, J., Schyns, P. G., Panzeri, S., and Kayser, C. (2017). Contributions of local speech encoding and functional connectivity to audio-visual speech perception. *eLife*, 6.

Giraud, A.-L. and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, 15(4):511–517.

Gliga, T. and Dehaene-Lambertz, G. (2007). Development of a view-invariant representation of the human head. *Cognition*, 102(2):261–288.

Goldowsky, B. N. and Newport, E. L. (1993). Modeling the Effects of Processing Limitations on the Acquisition of Morphology: The Less Is More Hypothesis. In *The Proceedings of the 24th Annual Child Language Research Forum*. Center for the Study of Language (CSLI). Google-Books-ID: x4FCUNpgPRQC.

Gómez, D. M., Berent, I., Benavides-Varela, S., Bion, R. A. H., Cattarossi, L., Nespor, M., and Mehler, J. (2014). Language universals at birth. *Proceedings of the National Academy of Sciences*, 111(16):5837–5841.

Goswami, U. and Leong, V. (2013). Speech rhythm and temporal structure: Converging perspectives? *Laboratory Phonology*, 4(1).

Grabe, E. and Low, E. L. (2002). Durational variability in speech and the Rhythm Class Hypothesis. In *Laboratory Phonology 7*. De Gruyter, Berlin, Boston, reprint 2013 edition.

Granier-Deferre, C., Ribeiro, A., Jacquet, A.-Y., and Bassereau, S. (2011). Near-term fetuses process temporal features of speech. *Developmental Science*, 14(2):336–352.

Grill-Spector, K. and Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychologica*, 107(1–3):293–321.

Grossmann, T., Johnson, M. H., Lloyd-Fox, S., Blasi, A., Deligianni, F., Elwell, C., and Csibra, G. (2008). Early cortical specialization for face-to-face communication in human infants. *Proceedings of the Royal Society of London B: Biological Sciences*, 275(1653):2803–2811.

Grossmann, T., Oberecker, R., Koch, S. P., and Friederici, A. D. (2010a). The Developmental Origins of Voice Processing in the Human Brain. *Neuron*, 65(6):852–858.

Grossmann, T., Parise, E., and Friederici, A. D. (2010b). The Detection of Communicative Signals Directed at the Self in Infant Prefrontal Cortex. *Frontiers in Human Neuroscience*, 4.

Guiraud, H., Ferragne, E., Bedoin, N., and Boulenger, V. (2013). Adaptation to natural fast speech and time-compressed speech in children. In *INTERSPEECH*, pages 1370–1374. Citeseer.

Guiraud, H., Hincapié, A.-S., Jerbi, K., and Boulenger, V. (2018). Perception de la parole et oscillations cérébrales chez les enfants neurotypiques et dysphasiques. In *Proc. XXXIIe Journées d'Études sur la Parole*, pages 222–230, Aix-en-Provence, France. ISCA.

Háden, G. P., Németh, R., Török, M., Drávucz, S., and Winkler, I. (2013). Context effects on processing widely deviant sounds in newborn infants. *Frontiers in Psychology*, 4.

Háden, G. P., Stefanics, G., Vestergaard, M. D., Denham, S. L., Sziller, I., and Winkler, I. (2009). Timbre-independent extraction of pitch in newborn infants. *Psychophysiology*, 46(1):69–74.

Hagiwara, R. (1995). Acoustic Realizations of American /r/ as Produced by Women and Men. *UCLA Working Papers in Phonetics*, 90:1–187.

Heinz, J. M. and Stevens, K. N. (1961). On the Properties of Voiceless Fricative Consonants. *The Journal of the Acoustical Society of America*, 33(5):589–596.

Henry, M. J. and Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences*, 109(49):20095–20100.

Henson, R., Shallice, T., and Dolan, R. (2000). Neuroimaging Evidence for Dissociable Forms of Repetition Priming. *Science*, 287(5456):1269–1272.

Henson, R. N. A. and Rugg, M. D. (2003). Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia*, 41(3):263–270.

Henson, R. N. A., Shallice, T., Gorno-Tempini, M. L., and Dolan, R. J. (2002). Face Repetition Effects in Implicit and Explicit Memory Tests as Measured by fMRI. *Cerebral Cortex*, 12(2):178–186.

Holmes, A. P., Blair, R. C., Watson, J. D. G., and Ford, I. (1996). Nonparametric Analysis of Statistic Images from Functional Mapping Experiments. *Journal of Cerebral Blood Flow & Metabolism*, 16(1):7–22.

Holmstrom, L. A., Eeuwes, L. B. M., Roberts, P. D., and Portfors, C. V. (2010). Efficient Encoding of Vocalizations in the Auditory Midbrain. *The Journal of Neuroscience*, 30(3):802–819.

Homae, F., Watanabe, H., Nakano, T., Asakawa, K., and Taga, G. (2006). The right hemisphere of sleeping infant perceives sentential prosody. *Neuroscience Research*, 54(4):276–280.

Homae, F., Watanabe, H., Nakano, T., and Taga, G. (2007). Prosodic processing in the developing brain. *Neuroscience Research*, 59(1):29–39.

Homae, F., Watanabe, H., Nakano, T., and Taga, G. (2011). Large-Scale Brain Networks Underlying Language Acquisition in Early Infancy. *Frontiers in Psychology*, 2.

Homae, F., Watanabe, H., and Taga, G. (2014). The Neural Substrates of Infant Speech Perception. *Language Learning*, 64(s2):6–26.

Houtgast, T. and Steeneken, H. J. M. (1985). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *The Journal of the Acoustical Society of America*, 77(3):1069–1077.

Howard, M. F. and Poeppel, D. (2010). Discrimination of Speech Stimuli Based on Neuronal Response Phase Patterns Depends on Acoustics But Not Comprehension. *Journal of Neurophysiology*, 104(5):2500–2511.

Hunter, M. A. and Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, 5:69–95.

Husain, A. M. (2005). Review of neonatal EEG. *American Journal of Electroneurodiagnostic Technology*, 45(1):12–35.

Isler, J. R., Tarullo, A. R., Grieve, P. G., Housman, E., Kaku, M., Stark, R. I., and Fifer, W. P. (2012). Toward an electrocortical biomarker of cognition for newborn infants. *Developmental Science*, 15(2):260–271.

Issard, C. and Gervain, J. (2017). Adult-like processing of time-compressed speech by newborns: A NIRS study. *Developmental Cognitive Neuroscience*, 25:176–184.

Issard, C. and Gervain, J. (2018). Variability of the hemodynamic response in infants: Influence of experimental design and stimulus complexity. *Developmental Cognitive Neuroscience*.

Jeschonek, S., Marinovic, V., Hoehl, S., Elsner, B., and Pauen, S. (2010). Do animals and furniture items elicit different brain responses in human infants? *Brain and Development*, 32(10):863–871.

Johnson, K., Ladefoged, P., and Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acoustical Society of America*, 94(2):701–714.

Joos, M. A. (1948). Acoustic phonetics. *Language*, 24(2):1–136.

Kaganovich, N., Francis, A. L., and Melara, R. D. (2006). Electrophysiological evidence for early interaction between talker and linguistic information during speech perception. *Brain Research*, 1114(1):161–172.

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., and Banno, H. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3933–3936.

Keitel, A., Gross, J., and Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLOS Biology*, 16(3):e2004473.

Kello, C. T., Bella, S. D., Médé, B., and Balasubramaniam, R. (2017). Hierarchical temporal structure in music, speech and animal vocalizations: jazz is like a conversation, humpbacks sing like hermit thrushes. *Journal of The Royal Society Interface*, 14(135):20170231.

Kidd, C., Piantadosi, S. T., and 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.

Kobayashi, M., Otsuka, Y., Nakato, E., Kanazawa, S., Yamaguchi, M. K., and Kakigi, R. (2011). Do infants represent the face in a viewpoint-invariant manner? Neural adaptation study as measured by near-infrared spectroscopy. *Frontiers in Human Neuroscience*, 5:153.

Kouider, S., Stahlhut, C., Gelskov, S. V., Barbosa, L. S., Dutat, M., Gardelle, V. d., Christophe, A., Dehaene, S., and Dehaene-Lambertz, G. (2013). A Neural Marker of Perceptual Consciousness in Infants. *Science*, 340(6130):376–380.

Kozberg, M. and Hillman, E. (2016). Chapter 10 - Neurovascular coupling and energy metabolism in the developing brain. In Kazuto Masamoto, H. H. a. K. Y., editor, *Progress in Brain Research*, volume 225 of *New Horizons in Neurovascular Coupling: A Bridge Between Brain Circulation and Neural Plasticity*, pages 213–242. Elsevier.

Kreitewolf, J., Gaudrain, E., and von Kriegstein, K. (2014). A neural mechanism for recognizing speech spoken by different speakers. *NeuroImage*, 91:375–385.

Kutas, M. and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207(4427):203–205.

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., and Schroeder, C. E. (2005). An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. *Journal of Neurophysiology*, 94(3):1904–1911.

Leong, V., Kalashnikova, M., Burnham, D., and Goswami, U. (2014). Infant-directed speech enhances temporal rhythmic structure in the envelope. *Int Speech Commun Assoc*, 2014:2563–7.

Lerner, Y., Honey, C. J., Katkov, M., and Hasson, U. (2014). Temporal scaling of neural responses to compressed and dilated natural speech. *Journal of Neurophysiology*, 111(12):2433–2444.

Lew, S., Sliva, D. D., Choe, M.-s., Grant, P. E., Okada, Y., Wolters, C. H., and Hämäläinen, M. S. (2013). Effects of sutures and fontanels on MEG and EEG source analysis in a realistic infant head model. *NeuroImage*, 76:282–293.

Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nature Neuroscience*, 5(4):356–363.

Liao, S. M., Gregg, N. M., White, B. R., Zeff, B. W., Bjerkaas, K. A., Inder, T. E., and Culver, J. P. (2010). Neonatal hemodynamic response to visual cortex activity: high-density near-infrared spectroscopy study. *Journal of Biomedical Optics*, 15(2):026010–026010–9.

Lloyd-Fox, S., Begus, K., Halliday, D., Pirazzoli, L., Blasi, A., Papademetriou, M., Darboe, M. K., Prentice, A. M., Johnson, M. H., Moore, S. E., and Elwell, C. E. (2017). Cortical specialisation to social stimuli from the first days to the second year of life: A rural Gambian cohort. *Developmental Cognitive Neuroscience*, 25:92–104.

Lloyd-Fox, S., Blasi, A., and Elwell, C. E. (2010). Illuminating the developing brain: The past, present and future of functional near infrared spectroscopy. *Neuroscience & Biobehavioral Reviews*, 34(3):269–284.

Lloyd-Fox, S., Moore, S., Darboe, M., Prentice, A., Papademetriou, M., Blasi, A., Kumar, S., Westerlund, A., Perdue, K. L., Johnson, M. H., Nelson, C. A., and Elwell, C. E. (2016). fNIRS in Africa & Asia: an Objective Measure of Cognitive Development for Global Health Settings. *The FASEB Journal*, 30(1 Supplement):1149.18–1149.18.

Lloyd-Fox, S., Richards, J. E., Blasi, A., Murphy, D. G. M., Elwell, C. E., and Johnson, M. H. (2014). Coregistering functional near-infrared spectroscopy with underlying cortical areas in infants. *Neurophotonics*, 1(2):025006–025006.

Lloyd James, A. (1940). *Speech signals in telephony*. Sir I. Pitman & sons, ltd., London. Open Library ID: OL6442817M.

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique, Second Edition*. MIT press, Cambridge, Mass., 2nd edition edition.

Luo, H. and Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron*, 54(6):1001–1010.

Luo, H., Wang, Y., Poeppel, D., and Simon, J. Z. (2006). Concurrent Encoding of Frequency and Amplitude Modulation in Human Auditory Cortex: MEG Evidence. *Journal of Neurophysiology*, 96(5):2712–2723.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Third Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Magnuson, J. S. and Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology. Human Perception and Performance*, 33(2):391–409.

Mahmoudzadeh, M., Dehaene-Lambertz, G., Fournier, M., Kongolo, G., Goudjil, S., Dubois, J., Grebe, R., and Wallois, F. (2013). Syllabic discrimination in premature human infants prior to complete formation of cortical layers. *Proceedings of the National Academy of Sciences*, page 201212220.

Mahmoudzadeh, M., Wallois, F., Kongolo, G., Goudjil, S., and Dehaene-Lambertz, G. (2017). Functional Maps at the Onset of Auditory Inputs in Very Early Preterm Human Neonates. *Cerebral Cortex*, 27(4):2500–2512.

Makeig, S., Westerfield, M., Jung, T.-P., Enghoff, S., Townsend, J., Courchesne, E., and Sejnowski, T. J. (2002). Dynamic Brain Sources of Visual Evoked Responses. *Science*, 295(5555):690–694.

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

Martynova, O., Kirjavainen, J., and Cheour, M. (2003). Mismatch negativity and late discriminative negativity in sleeping human newborns. *Neuroscience Letters*, 340(2):75–78.

Matsui, M., Homae, F., Tsuzuki, D., Watanabe, H., Katagiri, M., Uda, S., Nakashima, M., Dan, I., and Taga, G. (2014). Referential framework for transcranial anatomical correspondence for fNIRS based on manually traced sulci and gyri of an infant brain. *Neuroscience Research*, 80:55–68.

May, L., Byers-Heinlein, K., Gervain, J., and Werker, J. F. (2011). Language and the newborn brain: does prenatal language experience shape the neonate neural response to speech? *Frontiers in psychology*, 2:222.

May, L., Gervain, J., Carreiras, M., and Werker, J. F. (2018). The specificity of the neural response to speech at birth. *Developmental Science*, 21(3):n/a–n/a.

Meek, J. H., Firbank, M., Elwell, C. E., Atkinson, J., Braddick, O., and Wyatt, J. S. (1998). Regional hemodynamic responses to visual stimulation in awake infants. *Pediatric Research*, 43(6):840–843.

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

Mehler, J., Sebastian, N., Altmann, G., Dupoux, E., Christophe, A., and Pallier, C. (1993). Understanding Compressed Sentences: The Role of Rhythm and Meaning a. *Annals of the New York Academy of Sciences*, 682(1):272–282.

Millman, R. E., Prendergast, G., Kitterick, P. T., Woods, W. P., and Green, G. G. R. (2010). Spatiotemporal reconstruction of the auditory steady-state response to frequency modulation using magnetoencephalography. *NeuroImage*, 49(1):745–758.

Minagawa-Kawai, Y., Cristia, A., Long, B., Vendelin, I., Hakuno, Y., Dutat, M., Filippin, L., Cabrol, D., and Dupoux, E. (2013). Insights on NIRS sensitivity from a cross-linguistic study on the emergence of phonological grammar. *Language Sciences*, 4:170.

Minagawa-Kawai, Y., Mori, K., Naoi, N., and Kojima, S. (2007). Neural Attunement Processes in Infants during the Acquisition of a Language-Specific Phonemic Contrast. *Journal of Neuroscience*, 27(2):315–321.

Mizrahi, A., Shalev, A., and Nelken, I. (2014). Single neuron and population coding of natural sounds in auditory cortex. *Current Opinion in Neurobiology*, 24:103–110.

Molavi, B., May, L., Gervain, J., Carreiras, M., Werker, J. F., and Dumont, G. A. (2014). Analyzing the resting state functional connectivity in the human language system using near infrared spectroscopy. *Frontiers in Human Neuroscience*, 7.

Moon, C., Cooper, R. P., and Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, 16(4):495–500.

Moon, I. J. and Hong, S. H. (2014). What Is Temporal Fine Structure and Why Is It Important? *Korean Journal of Audiology*, 18(1):1–7.

Moore, B. C. J. (2008). The Role of Temporal Fine Structure Processing in Pitch Perception, Masking, and Speech Perception for Normal-Hearing and Hearing-Impaired People. *JARO: Journal of the Association for Research in Otolaryngology*, 9(4):399–406.

Moore, B. C. J. (2012). *An Introduction to the Psychology of Hearing*. BRILL. Google-Books-ID: LM9U8e28pLMC.

Moore, J. K. and Linthicum, F. H. J. (2007). The human auditory system: A timeline of development. *International Journal of Audiology*, 46(9):460–478.

Mullennix, J. W. and Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, 47(4):379–390.

Mullinger, K. J., Mayhew, S. D., Bagshaw, A. P., Bowtell, R., and Francis, S. T. (2014). Evidence that the negative BOLD response is neuronal in origin: A simultaneous EEG–BOLD–CBF study in humans. *NeuroImage*, 94(Supplement C):263–274.

Munck, J. C. d. and Bijma, F. (2010). How are evoked responses generated? The need for a unified mathematical framework. *Clinical Neurophysiology*, 121(2):127–129.

Murakami, S. and Okada, Y. (2006). Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals. *The Journal of Physiology*, 575(Pt 3):925–936.

Naccache, L. and Dehaene, S. (2001). The Priming Method: Imaging Unconscious Repetition Priming Reveals an Abstract Representation of Number in the Parietal Lobes. *Cerebral Cortex*, 11(10):966–974.

Nakano, T., Watanabe, H., Homae, F., and Taga, G. (2009). Prefrontal Cortical Involvement in Young Infants' Analysis of Novelty. *Cerebral Cortex*, 19(2):455–463.

Natan, R. G., Carruthers, I. M., Mwilambwe-Tshilobo, L., and Geffen, M. N. (2017). Gain Control in the Auditory Cortex Evoked by Changing Temporal Correlation of Sounds. *Cerebral Cortex*, 27(3):2385–2402.

Nazzi, T., Bertoncini, J., and Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3):756–766.

Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Current Opinion in Neurobiology*, 14(4):474–480.

Nelken, I. (2008). Processing of complex sounds in the auditory system. *Current Opinion in Neurobiology*, 18(4):413–417.

Németh, R., Háden, G. P., Török, M., and Winkler, I. (2015). Processing of Horizontal Sound Localization Cues in Newborn Infants. *Ear and Hearing*, 36(5):550.

Nespor, M. and Vogel, I. (2007). *Prosodic Phonology: With a New Foreword*. Walter de Gruyter. Google-Books-ID: VQC9jY2qTCkC.

Nichols, T. E. and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, 15(1):1–25.

Nordt, M., Hoehl, S., and Weigelt, S. (2016). The use of repetition suppression paradigms in developmental cognitive neuroscience. *Cortex*.

Norman, M. G. and O'Kusky, J. R. (1986). The Growth and Development of Microvasculature in Human Cerebral Cortex. *Journal of Neuropathology & Experimental Neurology*, 45(3):222–232.

Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., Howard, M. A., and Brugge, J. F. (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(49):15564–15574.

Nunez, P. L., Nunez, E. P. o. B. E. P. L., Srinivasan, R., and Srinivasan, A. P. o. C. S. R. (2006). *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press. Google-Books-ID: fUv54as56_8C.

Obleser, J. and Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences*, 13(1):14–19.

Obleser, J., Herrmann, B., and Henry, M. J. (2012). Neural Oscillations in Speech: Don't be Enslaved by the Envelope. *Frontiers in Human Neuroscience*, 6.

Odabaee, M., Tokariev, A., Layeghy, S., Mesbah, M., Colditz, P. B., Ramon, C., and Vanhatalo, S. (2014). Neonatal EEG at scalp is focal and implies high skull conductivity in realistic neonatal head models. *NeuroImage*, 96:73–80.

Okada, E. and Delpy, D. T. (2003a). Near-infrared light propagation in an adult head model. I. Modeling of low-level scattering in the cerebrospinal fluid layer. *Applied Optics*, 42(16):2906–2914.

Okada, E. and Delpy, D. T. (2003b). Near-infrared light propagation in an adult head model. II. Effect of superficial tissue thickness on the sensitivity of the near-infrared spectroscopy signal. *Applied Optics*, 42(16):2915–2922.

Orchik, D. J. and Oelschlaeger, M. L. (1977). Time-compressed speech discrimination in children and its relationship to articulation. *Journal of the American Audiology Society*, 3(1):37–41.

Overath, T., Cusack, R., Kumar, S., Kriegstein, K. v., Warren, J. D., Grube, M., Carlyon, R. P., and Griffiths, T. D. (2007). An Information Theoretic Characterisation of Auditory Encoding. *PLOS Biology*, 5(11):e288.

Pallier, C., Sebastian-Gallés, N., Dupoux, E., Christophe, A., and Mehler, J. (1998). Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory & Cognition*, 26(4):844–851.

Partanen, E., Kujala, T., Näätänen, R., Liitola, A., Sambeth, A., and Huotilainen, M. (2013). Learning-induced neural plasticity of speech processing before birth. *Proceedings of the National Academy of Sciences*, 110(37):15145–15150.

Peelle, J. E., McMillan, C., Moore, P., Grossman, M., and Wingfield, A. (2004). Dissociable patterns of brain activity during comprehension of rapid and syntactically complex speech: Evidence from fMRI. *Brain and Language*, 91(3):315–325.

Pefkou, M., Arnal, L. H., Fontolan, L., and Giraud, A.-L. (2017). Theta- and beta-band neural activity reflect independent syllable tracking and comprehension of time-compressed speech. *Journal of Neuroscience*, pages 2882–16.

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

Peterson, G. E. and Barney, H. L. (1952). Control Methods Used in a Study of the Vowels. *The Journal of the Acoustical Society of America*, 24(2):175–184.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (1986). Speaking Clearly for the Hard of Hearing II: Acoustic Characteristics of Clear and Conversational Speech. *Journal of Speech, Language, and Hearing Research*, 29(4):434–446.

Picton, T. W., John, M. S., Dimitrijevic, A., and Purcell, D. (2003). Human auditory steady-state responses: Respuestas auditivas de estado estable en humanos. *International Journal of Audiology*, 42(4):177–219.

Pike, K. L. (1945). *The Intonation of American English*. University of Michigan Press. Google-Books-ID: Jz3iAAAAMAAJ.

Plonsey, R. and Heppner, D. B. (1967). Considerations of quasi-stationarity in electrophysiological systems. *The bulletin of mathematical biophysics*, 29(4):657–664.

Potter, R. K. and Steinberg, J. C. (1950). Toward the Specification of Speech. *The Journal of the Acoustical Society of America*, 22(6):807–820.

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., and Crèpin, G. (1988). Fetal hearing. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 28(3):191–212.

Rabinowitz, N. C., Willmore, B. D. B., Schnupp, J. W. H., and King, A. J. (2011). Contrast Gain Control in Auditory Cortex. *Neuron*, 70(6):1178–1191.

Ramus, F. (2002). Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition*, 2(1):85–115.

Ramus, F., Nespor, M., and Mehler, J. (2000). Correlates of linguistic rhythm in the speech signal. *Cognition*, 75(1):AD3–AD30.

Rankin, S. K., Fink, P. W., and Large, E. W. (2014). Fractal structure enables temporal prediction in music. *The Journal of the Acoustical Society of America*, 136(4):EL256–EL262.

Roche-Labarbe, N., Fenoglio, A., Radakrishnan, H., Kocienski-Filip, M., Carp, S. A., Dubb, J., Boas, D. A., Grant, P. E., and Franceschini, M. A. (2014). Somatosensory evoked changes in cerebral oxygen consumption measured non-invasively in premature neonates. *NeuroImage*, 85(0 1).

Rosen, S. (1992). Temporal Information in Speech: Acoustic, Auditory and Linguistic Aspects. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 336(1278):367–373.

Rossi, S., Telkemeyer, S., Wartenburger, I., and Obrig, H. (2012). Shedding light on words and sentences: near-infrared spectroscopy in language research. *Brain and Language*, 121(2):152–163.

Sakatani, K., Chen, S., Lichty, W., Zuo, H., and Wang, Y. P. (1999). Cerebral blood oxygenation changes induced by auditory stimulation in newborn infants measured by near infrared spectroscopy. *Early Human Development*, 55(3):229–236.

Sato, H., Hirabayashi, Y., Tsubokura, H., Kanai, M., Ashida, T., Konishi, I., Uchida-Ota, M., Konishi, Y., and Maki, A. (2012). Cerebral hemodynamics in newborn infants exposed to speech sounds: A whole-head optical topography study. *Human Brain Mapping*, 33(9):2092–2103.

Sato, Y., Sogabe, Y., and Mazuka, R. (2010). Development of Hemispheric Specialization for Lexical Pitch-Accent in Japanese Infants. *Journal of Cognitive Neuroscience*, 22(11):2503–2513.

Scherer, K. R., Banse, R., Wallbott, H. G., and Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion*, 15(2):123–148.

Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Mata Pavia, J., Wolf, U., and Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *NeuroImage*, 85:6–27.

Scholkmann, F. and Wolf, M. (2013). General equation for the differential pathlength factor of the frontal human head depending on wavelength and age. *Journal of Biomedical Optics*, 18(10):105004–105004.

Schwartz, M. F. (1968). Identification of Speaker Sex from Isolated, Voiceless Fricatives. *The Journal of the Acoustical Society of America*, 43(5):1178–1179.

Sebastián-Gallés, N., Dupoux, E., Costa, A., and Mehler, J. (2000). Adaptation to time-compressed speech: Phonological determinants. *Perception & Psychophysics*, 62(4):834–842.

Segaert, K., Weber, K., de Lange, F. P., Petersson, K. M., and Hagoort, P. (2013). The suppression of repetition enhancement: A review of fMRI studies. *Neuropsychologia*, 51(1):59–66.

Selkirk, E. O. (1986). *Phonology and Syntax: The Relationship Between Sound and Structure*. MIT Press, Cambridge, MA, USA.

Sevy, A. B. G., Bortfeld, H., Huppert, T. J., Beauchamp, M. S., Tonini, R. E., and Oghalai, J. S. (2010). Neuroimaging with near-infrared spectroscopy demonstrates speech-evoked activity in the auditory cortex of deaf children following cochlear implantation. *Hearing Research*, 270(1–2):39–47.

Sheft, S., Shafiro, V., Lorenzi, C., McMullen, R., and Farrell, C. (2012). Effects of Age and Hearing Loss on the Relationship between Discrimination of Stochastic Frequency Modulation and Speech Perception. *Ear and hearing*, 33(6):709–720.

Shi, F., Yap, P.-T., Wu, G., Jia, H., Gilmore, J. H., Lin, W., and Shen, D. (2011). Infant Brain Atlases from Neonates to 1- and 2-Year-Olds. *PLOS ONE*, 6(4):e18746.

Shmuel, A., Yacoub, E., Pfeuffer, J., Van de Moortele, P.-F., Adriany, G., Hu, X., and Ugurbil, K. (2002). Sustained Negative BOLD, Blood Flow and Oxygen Consumption Response and Its Coupling to the Positive Response in the Human Brain. *Neuron*, 36(6):1195–1210.

Simoncelli, E. P. and Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual review of neuroscience*, 24:1193–1216.

Singh, N. C. and Theunissen, F. E. (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. *The Journal of the Acoustical Society of America*, 114(6):3394.

Smith, E. C. and Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, 439(7079):978–982.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, 27(4):501–532.

Stefanics, G., Háden, G., Huotilainen, M., Balázs, L., Sziller, I., Beke, A., Fellman, V., and Winkler, I. (2007). Auditory temporal grouping in newborn infants. *Psychophysiology*, 44(5):697–702.

Steinbrink, J., Villringer, A., Kempf, F., Haux, D., Boden, S., and Obrig, H. (2006). Illuminating the BOLD signal: combined fMRI–fNIRS studies. *Magnetic Resonance Imaging*, 24(4):495–505.

Strangman, G., Boas, D. A., and Sutton, J. P. (2002). Non-invasive neuroimaging using near-infrared light. *Biological Psychiatry*, 52(7):679–693.

Stroganova, T. A., Orekhova, E. V., and Posikera, I. N. (1999). EEG alpha rhythm in infants. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 110(6):997–1012.

Taga, G. and Asakawa, K. (2007). Selectivity and localization of cortical response to auditory and visual stimulation in awake infants aged 2 to 4 months. *NeuroImage*, 36(4):1246–1252.

Taga, G., Asakawa, K., Hirasawa, K., and Konishi, Y. (2003a). Hemodynamic responses to visual stimulation in occipital and frontal cortex of newborn infants: a near-infrared optical topography study. *Early Human Development*, 75:203–210.

Taga, G., Asakawa, K., Maki, A., Konishi, Y., and Koizumi, H. (2003b). Brain imaging in awake infants by near-infrared optical topography. *Proceedings of the National Academy of Sciences*, 100(19):10722–10727.

Taga, G., Watanabe, H., and Homae, F. (2011). Spatiotemporal properties of cortical haemodynamic response to auditory stimuli in sleeping infants revealed by multi-channel near-infrared spectroscopy. *Phil. Trans. R. Soc. A*, 369(1955):4495–4511.

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

Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., Obrig, H., and Wartenburger, I. (2009). Sensitivity of Newborn Auditory Cortex to the Temporal Structure of Sounds. *The Journal of Neuroscience*, 29(47):14726–14733.

Theunissen, F. E. and Elie, J. E. (2014). Neural processing of natural sounds. *Nature Reviews Neuroscience*, 15(6):355–366.

Torkildsen, J. v. K., Friis Hansen, H., Svangstu, J. M., Smith, L., Simonsen, H. G., Moen, I., and Lindgren, M. (2009). Brain dynamics of word familiarization in 20-month-olds: Effects of productive vocabulary size. *Brain and Language*, 108(2):73–88.

Traub, R. D., Miles, R., and Buzsáki, G. (1992). Computer simulation of carbachol-driven rhythmic population oscillations in the CA3 region of the in vitro rat hippocampus. *The Journal of Physiology*, 451:653–672.

Turk-Browne, N. B., Scholl, B. J., and Chun, M. M. (2008). Babies and Brains: Habituation in Infant Cognition and Functional Neuroimaging. *Frontiers in Human Neuroscience*, 2.

Urakawa, S., Takamoto, K., Ishikawa, A., Ono, T., and Nishijo, H. (2015). Selective Medial Prefrontal Cortex Responses During Live Mutual Gaze Interactions in Human Infants: An fNIRS Study. *Brain Topography*, 28(5):691–701.

Vagharchakian, L., Dehaene-Lambertz, G., Pallier, C., and Dehaene, S. (2012). A Temporal Bottleneck in the Language Comprehension Network. *The Journal of Neuroscience*, 32(26):9089–9102.

Vaissière, J. (1983). Language-Independent Prosodic Features. In Cutler, A. and Ladd, R. D., editors, *Prosody: Models and Measurements*, number 14 in Springer Series in Language and Communication, pages 53–66. Springer Berlin Heidelberg.

Varnet, L., Ortiz-Barajas, M. C., Erra, R. G., Gervain, J., and Lorenzi, C. (2017). A cross-linguistic study of speech modulation spectra. *The Journal of the Acoustical Society of America*, 142(4):1976–1989.

Von Holzen, K., Nishibayashi, L.-L., and Nazzi, T. (2018). Consonant and Vowel Processing in Word Form Segmentation: An Infant ERP Study. *Brain Sciences*, 8(2):24.

von Kriegstein, K., Smith, D. R. R., Patterson, R. D., Ives, D. T., and Griffiths, T. D. (2007). Neural Representation of Auditory Size in the Human Voice and in Sounds from Other Resonant Sources. *Current Biology*, 17(13):1123–1128.

von Kriegstein, K., Smith, D. R. R., Patterson, R. D., Kiebel, S. J., and Griffiths, T. D. (2010). How the Human Brain Recognizes Speech in the Context of Changing Speakers. *Journal of Neuroscience*, 30(2):629–638.

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

Wagner, J. B., Fox, S. E., Tager-Flusberg, H., and Nelson, C. A. (2011). Neural Processing of Repetition and Non-Repetition Grammars in 7- and 9-Month-Old Infants. *Frontiers in Psychology*, 2.

Wallois, F., Patil, A., Héberlé, C., and Grebe, R. (2010). EEG-NIRS in epilepsy in children and neonates. *Neurophysiologie Clinique/Clinical Neurophysiology*, 40(5-6):281–292.

Wang, X.-J. (2010). Neurophysiological and Computational Principles of Cortical Rhythms in Cognition. *Physiological Reviews*, 90(3):1195–1268.

Watanabe, H., Homae, F., Nakano, T., and Taga, G. (2008). Functional activation in diverse regions of the developing brain of human infants. *NeuroImage*, 43(2):346–357.

Watanabe, H., Homae, F., and Taga, G. (2012). Activation and deactivation in response to visual stimulation in the occipital cortex of 6-month-old human infants. *Developmental Psychobiology*, 54(1):1–15.

Webb, A. R., Heller, H. T., Benson, C. B., and Lahav, A. (2015). Mother's voice and heartbeat sounds elicit auditory plasticity in the human brain before full gestation. *Proceedings of the National Academy of Sciences*, 112(10):3152–3157.

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., and Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences*, page pnas.0809035106.

Wolf, M., Ferrari, M., and Quaresima, V. (2007). Progress of near-infrared spectroscopy and topography for brain and muscle clinical applications. *Journal of Biomedical Optics*, 12(6):062104.

Woolley, S. M. N., Fremouw, T. E., Hsu, A., and Theunissen, F. E. (2005). Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nature Neuroscience*, 8(10):1371–1379.

Zhang, L. I., Bao, S., and Merzenich, M. M. (2001). Persistent and specific influences of early acoustic environments on primary auditory cortex. *Nature Neuroscience*, 4(11):1123–1130.

Appendix 1: Variability of the hemodynamic response shape in infants

This thesis reports several NIRS experiments, some of them presenting inverted hemodynamic responses (e.g. for 30%-compressed speech in experiment 1). These responses were totally unexpected, and we remained puzzled about them. As we were looking for an explanatory hypothesis in the literature, we noticed that this kind of responses had been repeatedly observed in infants, but rarely commented in scientific journals. Furthermore, no explanatory model had been proposed to account for these inverted hemodynamic responses. We therefore decided to fill this gap. The following discussion is adapted from the resulting publication (Issard and Gervain, 2018).

The hemodynamic response in infants is often reported to be slower to peak, delayed and/or slower to go back to baseline or different in shape than the adult hemodynamic response, although no systematic longitudinal investigation of the infant and child hemodynamic response has been undertaken to date. The current section aims to address this gap by reviewing the existing literature in order to describe the variations in the shape and latency of the hemodynamic response found in studies with infants and young children, and propose a framework to explain these variations.

.1 Variability of the hemodynamic response reported in the infant literature

The hemodynamic responses reported in infant fNIRS literature show important variation within and across studies. The main goal of this review is to discuss some of the factors underlying this variability. Since NIRS provides an indirect, metabolic measure of brain activity, these factors are crucial when designing fNIRS experiments and interpreting their results. The factors we will discuss below include (i) developmental changes of the hemodynamic response, (ii) the development of physiological and cognitive processes, (iii) stimulus complexity and (iv) experimental design.

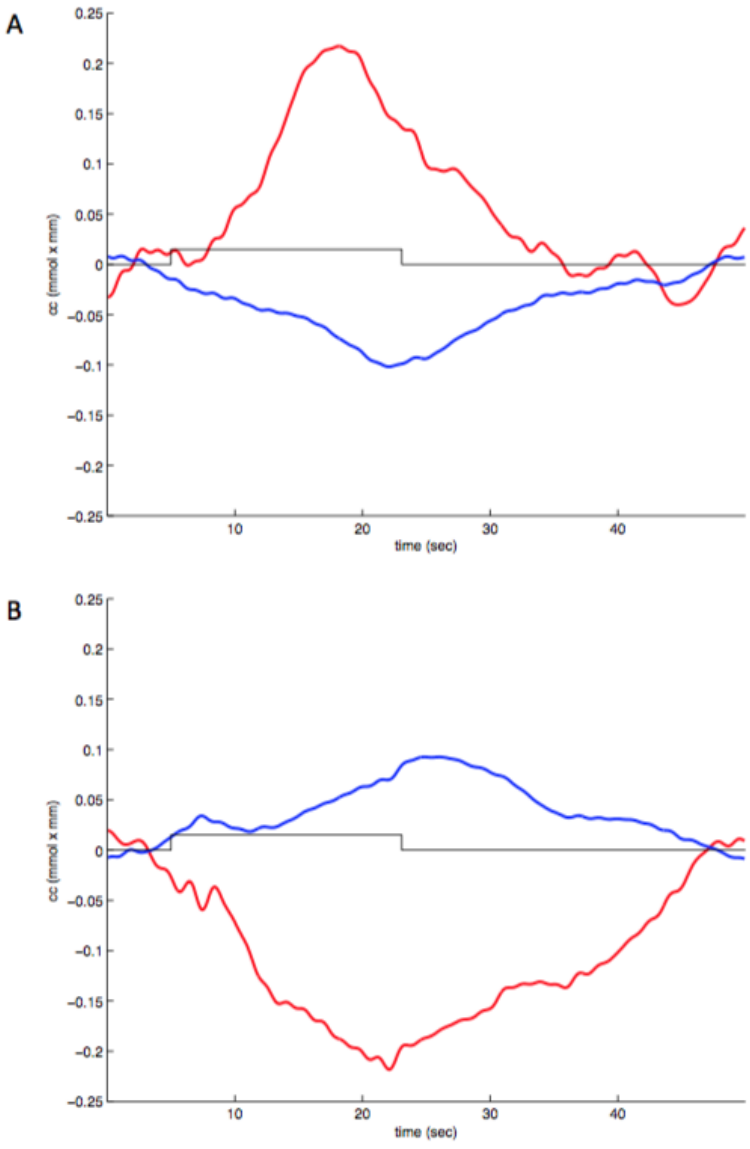
Despite their importance, these factors are rarely discussed explicitly when interpreting results, and have not yet been systematically investigated across development. While we are not offering a systematic empirical investigation either, reviewing the existing developmental NIRS literature nevertheless allows us to draw certain conclusions about how these factors impact the hemodynamic response. Importantly, however, the factors are inherently linked in any single study, it is thus impossible to identify and discuss the contribution of any factor in isolation. For the same reason, we are not providing an exhaustive review of the infant NIRS literature (for more general reviews and meta-analyses of the infant NIRS literature, we refer the reader to Cristia et al., 2013; Lloyd-Fox et al., 2010). Rather, we are focusing on studies that are comparable or matched along at least some relevant dimensions (e.g. testing the same task at different ages or testing the same age using stimuli that systematically vary), such that it becomes more apparent how factors in which they differ impact the hemodynamic response. In the sections below, we will first discuss the variability of the shape of the hemodynamic response observed at different ages in the different sensory modalities and cognitive functions. We will then describe how stimulus complexity, physiological and cognitive development and the experimental design might impact the hemodynamic response.

.1.1 Variation in the shape of the hemodynamic response reported in the developmental literature between cortical regions

The infant fNIRS and fMRI literature shows that the shape of the hemodynamic response changes with age in different brain areas (and perceptual or cognitive tasks). Canonical (i.e. statistically significant3 increase in HbO and decrease in HbR as compared to baseline, as depicted in Figure .1.1A) and inverted (statistically significant decrease in HbO and increase in HbR as compared to baseline, Figure .1.1B) reponses have both been reported in the literature. Some studies also report statistically significant HbO and HbR changes in the same direction.

Non-canonical responses, especially the inverted response, are not straightforward to interpret, because when absolute measures are taken, a decrease in oxygenation beyond a critical point is a signature of stroke, ischemia or other vascular problems. Most systems commonly used for research purposes only measure relative concentrations, i.e. concentration change. It is, therefore, difficult to understand the physiological and functional meaning of a relative decrease in oxygenation. Such a response pattern arises as a result of decreasing activation compared to a previously higher level of brain activity. This might occur for several possible reasons, e.g. when an inappropriately chosen baseline triggers more activation than the actual experimental stimulus, when habituation occurs, when a region is inhibited (Mullinger et al., 2014; Shmuel et al., 2002), due to reduced compensatory vascular mechanisms or indeed trauma. Do these factors also play a role in atypical hemodynamic responses observed in infants?

Figure 1: Canonical (A) and inverted (B) responses as observed in newborn infants in our laboratory. Red: HbO, blue: HbR.



The inverted response is often observed in young infants and newborns. Peak latency (i.e. time to reach the maximum of the hemodynamic response) has also been reported to decrease through infancy (Lloyd-Fox et al., 2017). The hemodynamic response also varies with the measured cortical region, and thus the tested sensory modality or cognitive function, and the arousal state of the participants (i.e. asleep vs awake). We provide examples of this variability across development and sensory modalities in Appendix A in Supplementary materials. For each age and sensory modality/brain region, we report the shape of the hemodynamic response measured along with the stimuli and experimental design used, focusing on studies with unsedated infants, as sedation may impact neurovascular coupling. In the section below, we discuss in detail the potential role of these factors on the variability of the hemodynamic response, in sensory processing as well as in higher-level cognitive functions, such as language processing and social cognition.

We argue that with the maturation of the brain, the hemodynamic response increasingly often takes on a canonical shape. The hemodynamic response relies on a complex interaction between the vascular system, neurons and glial cells, all of which undergo considerable maturation throughout infancy and childhood. However, because brain maturation is not homogenous between cortical regions, the hemodynamic response may vary from one brain area to another. Below (and in Appendix 1 in Supplementary materials), we thus review hemodynamic response variability separately for each sensory/cognitive function and its corresponding cortical areas.

In the temporal cortex, infant studies investigating auditory perception have demonstrated both inverted and canonical responses (Telkemeyer et al., 2009). At birth, both canonical and inverted responses were observed in the temporal cortex in response to speech and non-speech sounds between participants within the same condition (Sakatani et al., 1999), and within participants between conditions (Abboub et al., 2016; Issard and Gervain, 2017; Telkemeyer et al., 2009). Later in infancy, from 3 months onwards, infants' temporal cortex increasingly often shows a canonical response, even in infants born preterm (Emberson et al., 2017a,b), although inverted responses, with increasing HbR have also been reported as late as 7–9 months in response to complex speech stimuli (Wagner et al., 2011). It is possible to obtain a canonical hemodynamic response in the temporal cortex to various auditory stimuli (tones, vocal and non-vocal sounds) in different states of alertness, including sleep, as canonical responses have been reported in sleeping newborns as well as sleeping (non-sedated) 3-month-old infants (Taga et al., 2011).

In the posterior temporo-parietal region, the hemodynamic response to social stimuli (i.e faces and ostensive, infant-directed speech) has also been intensively investigated across development. Response to the same stimuli can vary with age. At 4 months of age, infants display a delayed but canonical response, and this response becomes faster during the first two years of life, as depicted in Figure .1.1 (Lloyd-Fox et al., 2017). But inverted responses have also been observed at similar ages in similar ROIs. At 5–6 and 7–8 months, a series of different faces presented from different points of view evoked a canonical response, whereas a

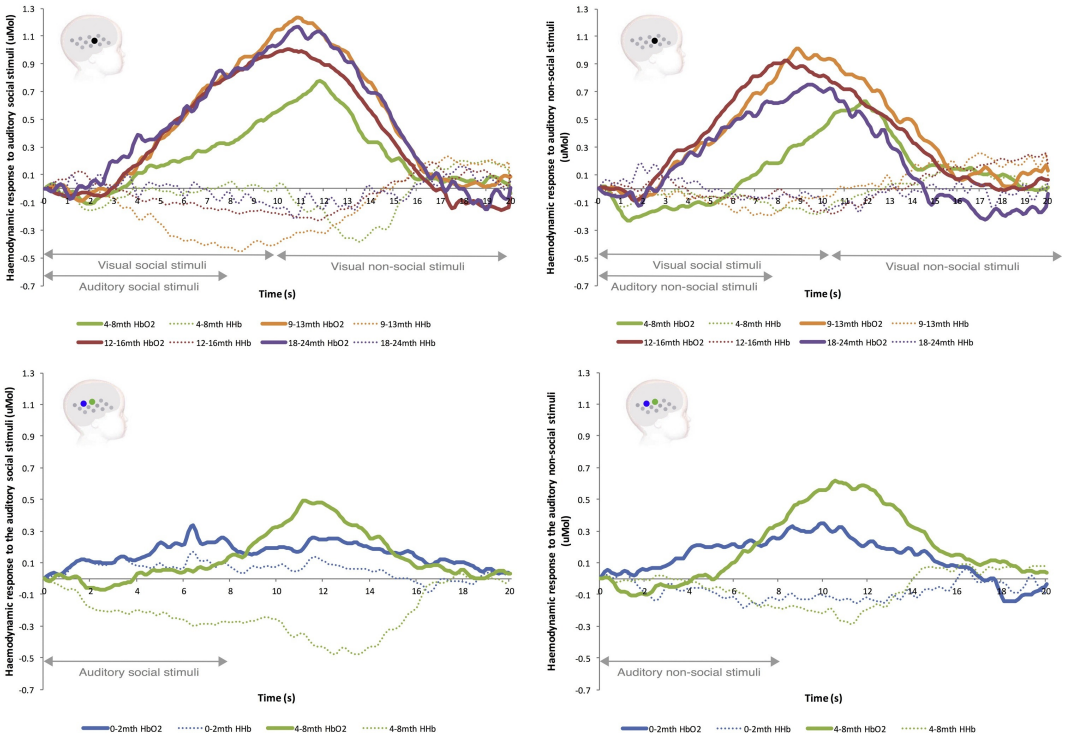


Figure 2: Hemodyamic response in the temporo-parietal junction as a function of age. Adapted from Lloyd-Fox et al. (2018).

series of the same face presented from different points of view evoked an inverted response (Kobayashi et al., 2011).

Inverted hemodynamic responses have also been reported in the occipital lobe at specific time points in development. In sleeping newborns, a flickering light (projected through the eyelids) evoked a canonical response. Between 0–3 months, a checkerboard reversal evoked an increase both in HbO and HbR in awake infants (Meek et al., 1998). Later in development, between 2 and 4 months, awake infants showed a canonical response to a checkerboard reversal, but an inverted response to a face-like pattern with blinking eyes (Taga et al., 2003b,a). Inverted responses in the occipital cortex were also observed in 4-month-old infants with pictures of faces compared to unstructured images, but in this study unstructured images also evoked inverted responses (Csibra et al., 2004). At 6 months, checkerboard pattern reversals evoked a canonical response, but unpatterned visual stimuli evoked an inverted response (Watanabe et al., 2012). The developmental trajectory of the hemodynamic response in the occipital cortex thus seems to be non-linear, varying with stimulus category.

Finally, the shape of the hemodynamic response in the frontal lobe depends on the sensory modality of the stimuli, the role the frontal lobe plays in their processing and the age of participants. For speech stimuli, the frontal lobe is typically recruited as it is part of the speech/language network (e.g. Broca's area),

whereas for visual perception it often plays a more general role related to attention. At birth, canonical responses to structured speech sounds were observed in left and right frontal cortices (Ferry et al., 2016; Gervain et al., 2008). At 3 months, both awake and sleeping infants showed a canonical hemodynamic response to speech sounds in bilateral frontal cortices (Homae et al., 2006, 2014). These results were replicated at 10 months (Homae et al., 2007). In the visual domain, a canonical response was measured to dynamic faces with mutual or adverted gaze at 4 months of age (Grossmann et al., 2008). In 5–7-month-old infants, a canonical response was observed to static faces, but only when the same face was repeatedly displayed during trials (Emberson et al., 2017a,b). By contrast, images of fruits didn't elicit significant responses in the same frontal area. Finally, an increase in HbO was observed in the orbito-frontal cortex in response to the smell of colostrum and vanilla in newborns (Bartocci et al., 2000).

On the basis of these examples, a great diversity of hemodynamic response shapes can be observed across development, showing an interaction between the age of participants and the cortical areas measured. The temporal cortex seems to present canonical responses earlier than the occipital and frontal cortices, and follows a more linear developmental trajectory than the occipital cortex. This latter shows a canonical response at birth, but an inverted response later in infancy. Finally, the frontal cortex shows more variable responses, depending on stimulus complexity and age of participants. Social stimuli, such as speech and faces, elicit canonical responses earlier than non-social stimuli (such as fruits or flashing lights).

.1.2 The factors influencing the hemodynamic response

In addition to the variability over developmental time and brain areas, the hemodynamic response seems to be further influenced by the complexity of the stimuli used, the development of the cognitive processes tested, and the experimental design, even within the same age or brain area.

Using blood oxygenation as a measure of neuronal activity relies on the assumption that the amount of change in oxygenation reflects the amount of neuronal activity. In cognitive neuroscience paradigms, changes in oxygenation in response to experimental conditions are typically compared between one another or to a baseline. The usual interpretation of this difference in hemodynamic response amplitude (or localization) is that it reflects differences in processing effort. It is, therefore, relevant to compare NIRS results, at least briefly, to models of processing effort based on behavioral measures (for a more detailed and not NIRS-specific comparison of behavioral and brain imaging studies in infants, see Turk-Browne et al., 2008). In this perspective, it can readily be explained that factors such as stimulus complexity, familiarity or the infants' developmental state all shape the hemodynamic response.

In developmental psychology, the effect of stimulus complexity/familiarity on infants' behavior, such as looking time, has received much attention (Goldowsky and Newport, 1993; Kidd et al., 2012; Hunter and Ames, 1988). Although the

factors influencing infants' looking behavior are not fully understood, one model (Kidd et al., 2012) holds that the probability for infants to look away from a stimulus is a U-shaped function of stimulus complexity. If stimuli are too simple or too complex, infants disengage (Kidd et al., 2012). While a systematic investigation of whether such a U-shaped function may also characterize NIRS responses when stimulus complexity is systematically varied is still lacking, the review of the NIRS literature we presented above suggests that parallels may exist. NIRS responses are often canonical in shape and largest in amplitude when stimuli are in the middle range of complexity, while overly complex or too simple stimuli more often evoke atypical responses.

Another model (Hunter and Ames, 1988), focusing on the impact of familiarity, predicts that when initially exposed to some stimuli, infants will show a preference for them for a period of time, as they are still in the process of exploring, encoding or learning them. Once encoding, memorization or learning is completed, infants' interest in the familiar stimulus decreases, and they readily show interest in novel stimuli. During the transition between these two periods, it is hypothesized that infants show no preference as they are equally attracted by the familiar and the novel stimuli (Hunter and Ames, 1988). This behavioral familiarization/habituation pattern also shows some similarities with the hemodynamic response, especially when experimental designs based on repeated exposure to the same stimuli are used (neural habituation). We discuss studies based on such experimental designs below.

Another relevant question is how different degrees of complexity or familiarity may change the type of processing mechanisms triggered. One relevant developmental model is the 'less is more' hypothesis (Goldowsky and Newport, 1993), which states that learners with limited memory capacity, such as infants, may use generalization, rule learning or categorization mechanisms to reduce the memory demand of highly variable or highly complex input, whereas learners with greater memory capacity tend to memorize or rote learn the input set. A change in stimulus complexity might thus trigger a qualitative rather than a quantitative change in cognitive processing, resulting in an entirely different pattern of responses in the NIRS data. The link between qualitatively different cognitive processing mechanisms and their neural signatures is not straightforward and remains unknown for most cognitive tasks, especially in development.

To the extent that the fNIRS response is indeed a measure of cognitive effort, it is important to take into account factors such as stimulus complexity, cognitive and physiological development and experimental design, when interpreting NIRS results, especially non-canonical ones. In the next sections, we discuss how these three factors modify cognitive effort and the hemodynamic response.

.2 Modulation of the hemodynamic response by complexity

.2.1 Variation due to stimulus complexity

The complexity and the familiarity of the stimuli are two potential factors that impact the hemodynamic response. Below, we provide examples for three different patterns of results that infant studies have found between their experimental conditions, and discuss to what extent these differences are attributable to stimulus complexity and/or familiarity.

First, some studies found canonical responses in all conditions, with a significant difference in amplitude across conditions. Familiarity plays a clear role in such differences. For instance, in the auditory domain, monolingual and bilingual newborns were presented with tone pairs in which the members differed along an acoustic dimension (pitch, duration or intensity) that either matched or did not match the acoustic cues carrying prosodic prominence in the language(s) the infants heard in utero (Abboub et al., 2016). The newborns' left temporo-parietal and right temporal cortices responded more to sound patterns that were inconsistent with the patterns found in the infants' native language, i.e. were unfamiliar, as compared to patterns that were consistent with it, i.e. were familiar, suggesting that processing unfamiliar melodic patterns often requires extra effort, resulting in a larger hemodynamic response. In the olfactory modality, vanilla smell evoked a larger response than the mother's colostrum in newborns' frontal cortex (Bartocci et al., 2000). Actually, responses to the mother's colostrum negatively correlated with the infants' age ($0 - 200$ h after birth). In older infants, the frontal region responded increasingly less to colostrum, probably due to the gradual replacement of colostrum with breastmilk and the decreasing relevance/familiarity of its smell. By contrast, vanilla was a new odor, which infants didn't encounter outside the testing sessions and to which all infants responded with a large hemodynamic response independently of age. Stimulus complexity also modulates the amplitude of the hemodynamic response. In an artificial grammar learning study with newborns, for example, trisyllabic sequences containing a repetition (e.g. "mulele") evoked a larger hemodynamic response than sequences containing no repetition (e.g. "mulevi") in the frontal and temporal regions (Gervain et al., 2012, 2008). In this case, the presence of a repetition created an underlying abstract structure, absent from the random sequences, and triggering a larger hemodynamic response. As another example of the role of complexity, at 3 months, both natural and sine wave speech induced a greater response than complex tones in the left posterior temporal and the right temporo-parietal cortex and the left frontal cortex (Homae et al., 2014). Natural speech and its sine wave equivalents are both acoustically more complex than tones, leading to a larger hemodynamic response. Differences in stimulus complexity also produce canonical responses of different amplitudes in the visual domain. In 3-month-old infants, for example, videos of mobile objects elicited a significantly greater hemodynamic response than checkerboard pattern

reversal in both frontal and occipital regions (Watanabe et al., 2008). Mobile objects were visually more complex than the checkerboard, containing more colors and a motion pattern more sophisticated than the reversal. Similarly, during live social interaction, the frontal cortex of 7-month-old infants showed larger responses to direct than to averted gazes, direct gazes being more engaging than averted gazes (Urakawa et al., 2015). When stimuli evoke canonical responses, complexity produces larger responses both in the visual and auditory domains.

Second, another pattern of results often observed in infant studies is a significant hemodynamic response in one condition, compared with a null response in some other condition. Differences in familiarity between conditions have been observed to play a role. For instance, comparing the responses to forward (FW) speech, backward (BW) speech, and silence in neonates, forward speech produced a classical canonical hemodynamic response in the left temporal cortex, but no significant response was measured during silence and backward speech (Pena et al., 2003). FW and BW speech are of the same complexity, as BW speech is the time-reversed version of FW speech, but BW speech is less familiar than FW speech, and may thus not be processed by the speech/language network of the newborn brain. Importantly, the language used in this study was the one the infants heard prenatally (Italian). Follow-up studies the contrasted responses to the native language with an unfamiliar language, both played forward and backward. Interestingly, one study found a larger response to forward speech and no significant response for silence and backward speech in the left temporo-parietal cortex of neonates, but only for the neonates' native language, Japanese. For English, the unfamiliar language, a bilateral temporal response was found with no difference between FW and BW speech (Sato et al., 2012). Comparing responses to FW and BW English and FW and BW Spanish in English-exposed newborns, the same unfamiliar Spanish stimuli evoked diminished non-significant responses in the fronto-temporal region when contrasted with English, the native language, but a larger FW-selective response when paired with Silbo Gomero, a Spanish based whistled language, therefore a non-speech-like but linguistic sound (May et al., 2018). These results suggest that familiarity shapes the observed hemodynamic response, but also that familiarity itself may be modulated by contextual effects. The unfamiliarity of the stimuli has been observed to trigger the absence of a significant hemodynamic response at even more fine-grained levels in speech processing. Phonologically possible words, allowed by language-universal constraints on syllable structure, evoked a significant response, but no significant response was found for words violating these constraints (Gómez et al., 2014). In the social visual domain, the occipital cortex of 4-month-old infants exhibited an inverted response to pictures of faces, but no response to unstructured images constructed from the same spatial frequencies and color distribution (Csibra et al., 2004).

Third, yet other studies observed canonical responses in some conditions and inverted ones in others. Comparing the native language with a non-native language in a paradigm similar to those used in the previously mentioned studies, but using low-pass filtered, thus less complex speech stimuli, a canonical, direction-

insensitive hemodynamic response to the native language, English, was observed, as well as to the unfamiliar language, Tagalog, played backwards, but an inverted hemodynamic response was observed to FW Tagalog (May et al., 2011). These results suggest that stimulus complexity interacts with familiarity when modulating the hemodynamic response. Canonical and inverted responses can also be observed within the same study when complexity alone is manipulated. At birth, for instance, canonical responses were observed to normal and moderately time-compressed speech (Issard and Gervain, 2017), but highly time-compressed speech produced an inverted response. For visual stimuli, a checkerboard reversal pattern evoked a canonical response in the occipital cortex of 3-month-olds, but a blinking face evoked an inverted response (Taga et al., 2003b,a). Similarly, at 6 months, unpatterned visual stimuli evoked an inverted hemodynamic response, while checkerboard pattern reversals evoked a canonical hemodynamic response (Watanabe et al., 2012). For social stimuli, at 5 months faces with direct gaze and the participant's own name evoked a canonical response in the left frontal cortex, but faces with averted gaze or another name evoked inverted hemodynamic responses (Grossmann et al., 2010a,b).

Together, these results suggest that the hemodynamic response is highly modulated by the presented stimuli within the same sensory modality, cognitive ability and age group. In line with the idea that the hemodynamic response is a reflection of cognitive effort, more complex and thus more demanding stimuli often elicit a larger hemodynamic response than simpler stimuli. However, the effects of stimulus complexity and familiarity are non-linear: when stimuli become overly complex or demanding, null or inverted responses can appear.

.2.2 Variation related to developmental changes

Brain maturation and the developmental stage of the tested cognitive function also modulate how the experimental stimuli are responded to.

At the neurobiological level, the immaturity of the vascular system in infants has been argued to play a role in the appearance of inverted responses. Indeed, they could be the result of insufficient cerebral blood flow (Meek et al., 1998). In young infants, the cerebral blood flow may sometimes be insufficient. In this case, HbR would not be fully eliminated, leading to an increase in HbR and a decrease in HbO. Infants' immature vascular system may insufficiently respond to the metabolic demand of the neural population. Indeed, in the somato-sensory cortex of preterm neonates, blood flow has been shown to increase immediately after stimulus onset and to return to baseline prior to HbR and HbO. Arterial to venous transit time was found to be longer than in adults (Roche-Labarbe et al., 2014). This increased transit time is consistent with an immaturity of the brain vasculature in infancy (Norman and O'Kusky, 1986). The capillary bed increases during the first months of life followed by a process of remodeling the vascular network as a function of local neural activity (Kozberg and Hillman, 2016).

The hypothesis of an occasional decoupling between the vascular and the

neural part of the neuro-vascular system is further confirmed by EEG-NIRS co-registration studies in infants. In highly premature newborns, a change in a single consonant in French syllables elicited larger electrophysiological and hemodynamic responses than no change, whereas a change of voice from male to female elicited a larger electrophysiological response, but a smaller hemodynamic response than no change (Mahmoudzadeh et al., 2013, 2017). This confirms that the relationship between neural and vascular activity is not linear in the developing brain.

In parallel with biological maturation, perceptual, cognitive and learning skills also develop, and their developmental stage impacts how stimuli are processed. In social cognition, for instance, 4 different age groups ranging from 4 to 24 months all showed a canonical hemodynamic response to social stimuli such as peek-a-boo videos and social sounds, with the peak latency decreasing over developmental time (Lloyd-Fox et al., 2017). This may reflect infants' increasing expertise in social stimuli during the first two years of life. Similarly, when comparing activity evoked by social auditory stimuli in 0–2-month-olds and 4–8–month-olds, the same authors reported a larger and more typical response in the older group. Interestingly, it has been shown that specialization for the human voice occurs between 4 and 7 months: the left and right superior temporal cortices showed increased responses to the human voice when compared to non-vocal sounds at 7, but not at 4 months (Grossmann et al., 2010a,b). This supports the hypothesis that infants present a larger response when they master the targeted cognitive skill, for instance when they specialize for the processing of the human voice as the most important social sound. Similarly in visual social stimuli, 5–6 and 7–8-month-old infants presented with pictures of the same or different faces showed a decrease in HbO in the right temporal area for different faces at 5–6 months, but an increase in HbO for the same stimuli at 7–8 months. This confirms the hypothesis that mastery of a cognitive skill leads to greater responses.

Similar developmental trends can be observed in the speech perception domain. For instance, a developmental change was found in response to the native dialect vs. another dialect: 3-month-old Parisian infants showed a similar canonical response to both Parisian and Quebecois French talking faces, while 5-month-olds showed an advantage for the Parisian stimuli in the left temporal areas (Cristia et al., 2013). Similarly, in a pitch accent discrimination task, Japanese 4-month-olds showed canonical activation bilaterally, while 10-month-olds already exhibited left-lateralized, i.e. language-specific responses to the same stimuli, i.e. a null response in the right hemisphere (Sato et al., 2010). In the same vein, 6–7-month-olds present a greater hemodynamic response to across-category than within-category phonemic changes, mirroring behavioral data. By contrast, the responses to the two conditions are similar at 10–11 months, as the phonological system undergoes reorganization, attuning to the native language. Then the response to across-category phonemic change becomes larger again with a left hemisphere advantage at 13–14 and 25–28 months (Minagawa-Kawai et al., 2007). Perceptual narrowing, i.e. attunement to the native language, brings about a reorganization, whereby infants switch from an auditory to a linguistic processing of speech sounds and

shape their phonological space to better match the phoneme contrasts found in the native language during the second half of the first year of life. This changes the nature and efficiency of the underlying cognitive processing, and therefore its metabolic costs.

Together, these data suggest that development has a considerable impact on how the infant brain processes the same stimuli. As sensory and cognitive functions develop, the newly emerging cognitive or perceptual skills often lead to additional, higher level steps of processing, for instance because a novel, more abstract representation is formed or because a category is learned or reshaped, and might thus evoke a larger hemodynamic response. By contrast, for cognitive functions where processing becomes more automatic and/or faster with development, a smaller hemodynamic response is observed. Therefore, similarly to the U-shaped developmental curves observed in several perceptual and cognitive domains at the behavioral level, developmental change may not be linear in direction at the neural level either, with increases in hemodynamic response amplitude early in development and decreases later.

.3 Modulation of the hemodynamic response by experimental design

Specific experimental designs can have an effect on the hemodynamic response, as they modify the context of stimulus presentation by adding another level, that of local and global stimulus arrangement/context. Below we discuss how this can influence the shape of the hemodynamic response.

.3.1 Simple event-related or block designs

The simplest and most commonly used design in infant NIRS studies is the presentation of stimuli belonging to different experimental conditions in an interleaved, random order, where one long (in event-related designs) or several shorter (in block designs) stimuli from the same condition are presented (for a duration of several seconds to allow for the hemodynamic response to build up). Both canonical (Homae et al., 2007; Taga et al., 2003b,a) and inverted (Telkemeyer et al., 2009; Watanabe et al., 2012) responses have been observed in such designs, as shown in many of the studies reviewed above.

In this design, the order of stimuli or blocks is randomized. Any interaction between consecutive stimuli is, therefore, random, and should not influence the hemodynamic response in a systematic way. Rather, the measured hemodynamic response is expected to be influenced only by the properties of the stimuli.

.3.2 Repetition effects

Repetition effects might arise when stimuli or blocks from the different conditions are not intermixed, but rather grouped together by condition, resulting in a large number of repetitions of the same stimulus type.4 Repetitions can influence the neural and hence the hemodynamic response by producing either enhancement, or suppression. Repetition suppression can be defined as a decrease of neural activity in response to the repeated presentation of the same stimulus or stimulus category. Initially shown in non-human primates at the single cell level, this effect was later demonstrated at the level of the hemodynamic response in humans, first in adults (Buckner and Koutstaal, 1998) and more recently in infants (Nakano et al., 2009). Repetition enhancement can be defined as an increase of neural activity in response to repeated presentation of the same stimulus or stimulus category. Both repetition enhancement and suppression effects have been found in adults depending on experimental parameters, especially stimulus complexity and quality. Suppression is more common. Repetition enhancement is mainly assumed to occurs when stimuli are degraded, masked or when participants do not have a stable memory representation of the stimuli (Henson and Rugg, 2003; Henson et al., 2002, 2000; Naccache and Dehaene, 2001; Segaert et al., 2013).

Similarly to adults, both repetition suppression and enhancement have been found in developmental populations. Indeed, several studies have found an increase of the hemodynamic response when a condition or a stimulus was repeated. Bouchon et al. (2015) presented newborns with trisyllabic sequences following either an ABB (e.g. “mulele”) or an ABC pattern (e.g. “mulevi”), following-up on Gervain et al. (2008). In contrast to Gervain et al. (2008), however, who presented blocks in an interleaved, random order, Bouchon et al. (2015) grouped together all blocks of the same condition. Gervain et al. (2008) found an increasingly greater response over the course of the experiment for the ABB patterns as compared to the ABC ones, whereas Bouchon et al. (2015) showed a repetition enhancement effect during the time course of the ABC condition. The grouped vs. interleaved presentation modes may have contributed to these differences, although it needs to be noted that the two studies also differed in stimulus complexity, as Gervain et al. (2008) used 280 unique trisyllabic sequences, each of which occurred only once, whereas Bouchon et al. (2015) used 24 different sequences, each repeated 6 times.

Repetition suppression effects were also found in infants. At 3–4 months, the frontal cortex showed reduced hemodynamic response over time to the repetition of the same syllable 15 times (Nakano et al., 2009). Similarly, 5–7-month-olds showed a reduced hemodynamic response in the frontal cortex when hearing the same word 8 times in a block as compared to blocks containing 8 different words (Emberson et al., 2017b,a).

What factors determine whether enhancement or suppression is observed? The two studies that showed a repetition suppression effect both repeated the exact same physical stimuli within each condition (Emberson et al., 2017b; Nakano et al., 2009), whereas in the studies that found enhancement, repetition occurred at

a more abstract, structural level (e.g. in Gervain et al., 2008; Bouchon et al., 2015, different trisyllabic sequences were presented, but they all shared a common ABB pattern). Repetition enhancement might thus reflect the ongoing effort to build a representation of the stimulus category, whereas repetition suppression may typically occur when the category is encoded in a more stable way (see Nordt et al., 2016, for a discussion).

The repetition suppression effect serves as a basis to create powerful experimental designs for functional studies of the cortex, namely the habituation/dishabituation design, and it has been suggested that the neural mechanisms might be analogous to the cognitive mechanisms underpinning behavioral habituation paradigms (Turk-Browne et al., 2008). In these experiments, a stimulus is repeated several times to produce neural suppression, followed by the presentation of a test stimulus, different from the repeated stimulus along the tested dimension (Grill-Spector and Malach, 2001). This design provides a powerful tool to show differential encoding of two stimulus categories. Specifically, a rebound of activity is observed when a novel test stimulus is presented, as the suppression effect subsides. If the two types of stimuli were presented in a simple randomized fashion, activity could be similar for both. This design is widely used in adult fMRI as well as in adult and infant EEG studies. In infant EEG studies, the habituation effect is revealed by a response suppression when the test stimulus is preceded by stimuli from the same category as compared to stimuli preceded by a different-category item (e.g. Friedrich and Friederici, 2004; Gliga and Dehaene-Lambertz, 2007; Jeschonek et al., 2010; Torkildsen et al., 2009).

Habituation designs have recently been adapted for infant NIRS studies to reveal fine-grained discrimination abilities. For instance, newborns were familiarized with repeated words and the hemodynamic response to a change in consonant or vowel was measured in the test phase. A rebound of the hemodynamic response was observed in the right frontal cortex in response to a consonant change (Benavides-Varela et al., 2012). Here, the suppression effect reveals a subtle discrimination ability, namely that of the consonant/vowel distinction.

.3.3 Alternating presentation

Alternating/non-alternating designs are often used in developmental NIRS studies to show subtle discrimination. This design was originally developed for behavioral paradigms to "offer a more sensitive measure than the between-trial comparisons" (Best and Jones, 1998). Two types of blocks are presented: homogeneous or non-alternating blocks composed of stimuli from the same experimental condition, and alternating blocks composed of stimuli from two different conditions in alternation.

One of the first NIRS studies using this design presented 4- and 10-month-old infants with disyllabic words containing either a high-low or a low-high pitch accent, typical of Japanese, or pure tone pairs corresponding to the pitch contours of the words. In each of these two conditions, half of the blocks contained only one of the two contours, the other half contained an alternation between the two. The

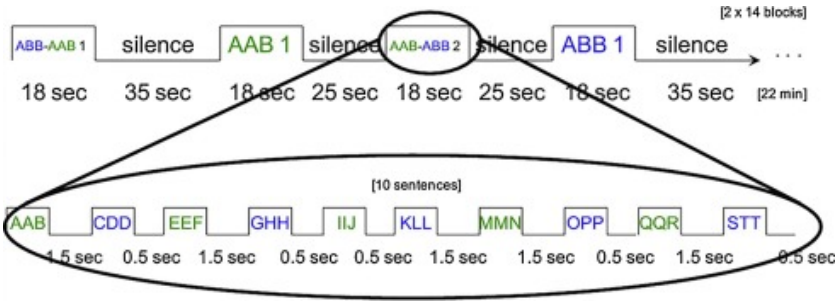


Figure 3: Alternating/non-alternating design with numerous variable stimuli within conditions. Adapted from Gervain et al. (2012).

alternating blocks elicited a larger response in the word condition than in the pure tone condition in the 10-month-old infants (Sato et al., 2010). Another example for the use of alternating/non-alternating designs comes from the field of number perception. 6-month-old infants were presented with blocks of 10 different images containing either 8 or 16 dots varying in size, color and spatial position. Half of the blocks contained images of 8 dots only, the other half contained alternating images of 8 and 16 dots. Alternating blocks elicited a larger response than non-alternating blocks in the right parietal region (Edwards et al., 2016). In newborns, experiments that used an alternating/non-alternating design have sometimes found a greater response to non-alternating than to alternating blocks (Gervain et al., 2012, 2016; Issard and Gervain, 2017). In newborns and infants, auditory short-term memory is less developed than in adults or children, estimated between 800 and 1500 ms for newborns (Cheour et al., 2002). This estimation is much shorter than the interblock interval required for the hemodynamic response to decline, but is similar to the delay typically used between items within blocks. If the two conditions are presented side by side as is the case within alternating blocks, then it may be easier for newborns to remember the previous stimulus, compare it to the current stimulus, and potentially discriminate between the two.

.4 Conclusions: the hemodynamic response as a reflection of cognitive effort

In this chapter, we reviewed a wide variety of studies in which fNIRS is used with typical and clinical developmental populations of various ages in different perceptual and cognitive tasks. Experimental results suggest that in addition to the variability over developmental time and brain areas, the hemodynamic response seems to be further influenced by the complexity of the stimuli used, the development of the cognitive processes tested, and the experimental design, even within the same age or brain area. Each of these factors has a different impact on the hemodynamic response, and they interact with one another.

Stimulus complexity seems to influence the hemodynamic response in a non-linear fashion. Based on the existing literature, we have argued that when stimuli are too simple, they evoke smaller responses than when stimuli are of middle-range complexity. Overly complex stimuli lead to null or inverted responses.

Cognitive development can modulate the hemodynamic response within the same cortical area. When a cognitive function is more developed, a canonical response is more likely to be observed. In experiments testing a less developed function (i.e. younger participants), a larger response may be interpreted as a signature of more efforts to process the stimuli. When the stimuli are particularly difficult for the tested developmental stage, inverted responses can more often be observed.

Experimental design can also influence the hemodynamic response: a repeated presentation of identical or similar stimuli might lead to a reduction of the hemodynamic response (i.e. habituation), while designs in which conditions alternate might evoke different hemodynamic response amplitudes between alternating and non-alternating blocks.

The influence of these modifying factors may be best understood in a common framework combining behavioral and neural mechanisms in which the hemodynamic response is interpreted as processing effort. For an a priori prediction of expected effects and an appropriate interpretation of NIRS results, it is essential to have an operational model, behavioral and/or neural, of the processing mechanisms involved, and how the different factors manipulated in a given study might modulate processing effort.

Appendix 2: Table summarizing the shape of hemodynamic responses observed in the infant NIRS literature

The following table synthesizes the NIRS literature discussed in Appendix 1 (Issard and Gervain, 2018). It corresponds to the supplementary material of the published paper.

Sensory modality	Cortical area	Age of participants	Awake / asleep	Stimuli	Experimental design	Hemodynamic response	Reference
auditory	right fronto-temporal	Full-term newborns	asleep	Pure tones following or violating the prosodic structure of the native language	Simple blocks	Canonical with the native prosodic structure, inverted with the wrong structure	Abboub et al. (2016)
auditory	Right frontal, left temporal, right parietal	Full-term newborns	asleep	words with consonantal vs. vocalic change in test blocks	Habituation-test	Canonical, test>habituation for vocalic change Inverted in test blocks for consonantal change	Benavides-Varela et al. (2012)
auditory	Right temporal, right parietal, left fronto-temporal	Full-term newborns	asleep	12 trisyllabic sequences containing a repetition (ABB) or no internal structure (ABC)	Habituation-test	Canonical, repetition>no-structure in the right temporal area. First inverted, with gradual increase over time in no-structure blocks in the left fronto-temporal	Bouchon et al. (2015)
auditory	Fronto-	Full-term	asleep	Sequences of 6 consonant-vowel	Habituation-test	canonical	Ferry et al.

	temporal	newborns		syllables			(2015)
auditory	Temporal & frontal	Full-term newborns	asleep	280 trisyllabic sequences	Simple blocks	canonical	Gervain et al. (2008)
auditory	Temporal & frontal	Full-term newborns	asleep	140 trisyllabic sequences containing a repetition or no internal structure	Simple blocks	Canonical, repetition > unstructured	Gervain et al. (2012)
auditory	Temporal & frontal	Full-term newborns	asleep	Scale-invariant vs. variable scale sounds	Alternating/non-alternating	non-alternating > alternating	Gervain et al. (2016)
auditory		Full-term newborns	asleep	phonologically probable and improbable words,	Simple blocks	Inverted for probable words	Gómez et al. (2014)
auditory	Right temporal, right frontal, & left temporo-parietal	Full-term newborns	asleep	120 French sentences at 100, 60, or 30% of their duration	Alternating/non-alternating	Canonical for normal and 60% compressed, inverted for 30% compressed	Issard & Gervain (2017)
auditory	Left & right temporal,	Full-term newborns	asleep	six English sentences and six Tagalog sentences	Simple blocks	Canonical for backward tagalog and	May et al. (2011)

	left frontal, & right temporo-parietal			played either forward or backward		forward English in the temporal region inverted for forward tagalog in the frontal region canonical for backward, inverted for forward in the temporo-parietal region backward>forward	
auditory	Temporal, frontal & fronto-temporal	Full-term newborns	asleep	English (native, spoken) vs spanish (non-native, spoken) vs. silbo sentences (non-native, whistled) played either forward or backward	Simple blocks	Canonical, forward> backward in English and Spanish, Spanish > Silbo Gomero	May et al. (2017)
auditory		Full-term newborns	asleep	Italian sentences played forward vs. backward vs. silence	Simple blocks	Canonical for forward speech	Peña et al. (2003)

auditory	frontal	Pre-term and full-term newborns		Popular piano music	Simple blocks	Canonical or inverted depending on participant	Sakatani et al. (1999)
auditory	Temporal & temporo-parietal	Full-term newborns	asleep	Japanese vs. English, played either forward or backward vs silence	Simple blocks	Canonical, Japanese forward > japanese backward > silence, Japanese forward > English forward	Sato et al. (2012)
auditory	Temporo-parietal	Full term newborns	unspecified	Noise segments of 4 different durations	Simple blocks	Canonical for the longer durations, inverted with the shortest duration	Telkemeyer et al. (2009)
auditory	temporal, frontal, temporo-parietal	3 months old	asleep	36 japanese sentences with normal vs flat prosody	Event-related	Canonical, normal > flattened	Homae et al. (2006)
auditory	temporal, frontal, temporo-parietal	3 months old	awake	12 japanese sentences with normal vs sine-wave speech (SWS) vs pure tones	Event-related	Canonical, natural & SWS > tones	Homae et al. (2014)

auditory	Frontal & temporal	3 months old	asleep	2 syllables: one for habituation and one for test	Habitution/test	Canonical, repetition suppression effect in frontal region	Nakano et al. (2009)
auditory	Temporal, parietal, prefrontal, and occipital	3 months old	asleep	random sequence of 25 pure tones	Simple blocks	canonical	Taga et al. (2011)
Audio-visual	temporal	3 months old	awake	78 videos of passages read with Parisian or quebois accent	Alternating/non-alternating	canonical	Cristia et al. (2014)
		5 months old				Canonical, alternating>non-alternating	
Audio-visual	Frontal Fronto-temporal Temporal Temporo-parietal	6 months old	awake	36 pseudowords, legal vs illegal in the participants' native language trained vs untrained to be matched with 18 pictures of complex, colorful pseudo-objects	Event-related	Inverted pre-training Canonical post-training	Obrig et al. (2017)

auditory	temporal	4 months old	awake	Vocal vs. non-vocal sounds	Simple blocks	Canonical, non-vocal>vocal	Grossman et al. (2010)
		7 months old				Canonical, vocal>non-vocal	
auditory	Right temporal	7 months old	awake	74 semantically neutral German words produced with happy, angry, and neutral prosody	Simple blocks	Canonical, angry > happy & neutral in posterior temporal cortex, happy > angry & neutral	
auditory	Temporal & frontal	5-7 months old	awake	8 english words	Blocks with either the same word repeated or presented once per block.	Canonical, variable>repeated	Emberson et al. (2017b)
auditory	temporal, frontal, temporo-parietal	7-9 months old	awake	280 trisyllabic sequences	Simple blocks	inverted	Wagner et al. (2011)
auditory	temporal, frontal, temporo-parietal	10 months old	asleep	Normal vs flat prosody speech	Event-related	canonical	Homae et al. (2007)

auditory	temporal	4 months old	awake	Pitch pattern within words or tone sequences	Alternating/non-alternating	canonical	Sato et al. (2010)
		10 months old				Canonical, words> pure tones in the left temporal region	
auditory	Temporal areas	3-4, 6-7, 10-11, 13-15, & 25-28 months old	awake	4 pseudo-words with within vs. across category phonemic change	Alternating/non-alternating	Canonical broadly distributed in 3 to 15 m.o. Canonical, restricted area in 25-28 m.o. Across > within phonemic change in 6-7, 13-14, and 25-28 m.o.	Minagawa-Kawaï et al. (2007)
visual	occipital and frontal	newborns	asleep	stroboscopic light	Simple blocks	canonical	Taga et al. (2003a)
visual	occipital	0-3 months old	awake	black and white checkerboard with 5-Hz pattern reversal	Simple blocks	canonical	Meek et al. (1998)

visual	Occipital & frontal	3 months old	awake	Videos of mobile objects vs. checkerboards pattern reversal	Event-related	Canonical, mobile > checkerboards	Watanabe et al. (2008)
visual	occipital and frontal	2-4 months old	awake	Checkerboard pattern reversal vs face-like pattern with blinking eyes	Event-related	Canonical with checkerboards, inverted with face-like patterns	Taga et al. (2003b)
visual	occipital	4 months old	awake	Faces vs unstructured images	Event-related	Inverted for faces	Csibra et al. (2004)
visual	Occipital and parietal	6 month-olds	awake	Images with variable number of dots	Alternating/non-alternating	Canonical, alternating > non-alternating	Edwards et al. (2015)
visual	Occipital and frontal	5-7 months-old	awake	Pictures of 8 different faces or fruits	Blocks with either the same picture presented repeatedly or 8 pictures presented once per block	canonical	Emberson et al. (2017)
Audio-visual	occipital	6 month-old full-term and preterm infants	awake	Predicitve sound followed by a predicted visual stimulus (80% of trials) or an	Event-related	Canonical in both conditions for full-terms; canonical in predicted and	Emberson et al. (2017)

				ommission of the visual stimulus (20% of trials)		inverted in visual omissions for pre-terms.	
visual	Occipital and frontal	6 months old	awake	unpatterned screens vs black-and-white checkerboard pattern reversals	Event-related	Canonical for checkerboard, inverted for unpatterned in the occipital cortex. Canonical in the frontal cortex.	Watanabe et al. (2012)
Audio-visual	right temporal	0 to 24 months old	Newborns asleep, 2 to 24 month-old awake	Video clips of peek-a-boo games (only audio for newborns)	Simple blocks	Canonical, increasingly shorter with development	Lloyd-Fox et al (in press)
visual	Right temporal & right frontal	4 months old	awake	Animated face with mutual or averted gaze	Event-related	Canonical, mutual>averted	Grossman et al. (2008)
visual	Left frontal	5 months old	awake	Pictures of a smiling face with mutual or averted gaze	Event-related	Canonical for mutual, inverted for averted gaze	Grossman et al. (2010)

auditory				Own name vs other name		Canonical for own name, inverted for other name	
Audio-visual	Temporal and temporo-parietal	4-6 months old	awake	Live nursery rhymes in IDS with hand movements	Event-related	canonical	Lloyd-Fox et al. (2015)
visual	Temporal	5-6 months old	awake	Pictures of 5 female faces	Blocks with the same face presented repeatedly, or different faces presented once per block	Inverted for both same and different faces	Kobayashi et al. (2011)
		7-8 months old				Canonical for different faces	
visual	temporal	7-8 months old	awake	Facial movements of point-light displays, upright or inverted	Event-related	Canonical for upright, flat for inverted	Ichikawa, et al. (2010)

visual	temporal	7-8 months old	Awake	Photos of mother's face or strangers' face	Blocks with the mother's face presented repeatedly or strangers' face presented once per block.	canonical	Nakato et al. (2011)
visual	Temporo-parietal	7-8 months old	Awake	Canonical faces	Simple blocks	Canonical	Honda et al. (2010)
				Scrambled faces		Canonical in the left hemisphere, inverted in the right hemisphere	
smell	frontal	Preterm newborns	asleep	Detergent smell	Simple blocks	Inverted	Bartocci et al. (2001)
smell	frontal	newborns	asleep	own mother's colostrum vs. vanilla smell	Simple blocks	Canonical for both	Bartocci et al. (2000)
tactile	left fronto-central	Preterm newborns	asleep	gentle strokes the right hand with a toothbrush	Simple blocks	canonical	Roche-Labarbe et al. (2014)
tactile	prefrontal	3,6, and 10 months old	awake	moving a column-shaped, end-rounded piece of	Simple blocks	Increase in HbO at 10 months old	Kida et al. (2013)

				wood vs a velvet-wrapped wood packed with cotton over the left palm			

Otherwise specified, results apply to all cited cortical regions.

Appendix 3: Non-parametric permutation tests for infant NIRS data

Near-infrared spectroscopy is a rapidly growing technique whose methodology is rapidly evolving. In cognitive neuroscience, NIRS is mainly used in developmental research, and infant data may require specific statistical approaches. In particular, infant data often do not meet the assumptions for parametric approaches. Another problem of infant data is their lack of statistical power. Other statistical procedures than the classical corrected parametric ones must thus be applied. We therefore proposed a procedure inspired by the one proposed by Maris and Oostenveld (2007) for MEEG data, but adapted to the specificity of NIRS. This poster was presented at the biennial meeting of the Society for functional Near-Infrared Spectroscopy (Paris, October 2016). The strategy used has partly evolved following the interesting discussions that took place around this poster. In particular, a procedure similar to the one proposed by Maris and Oostenveld (2007) for MEEG data has finally been implemented to make use of the sampling rate of NIRS and study spreads of activity over the cortex. The corresponding MATLAB codes can be downloaded at github.com/cissard/NIRS.

Issard, C. & Gervain, J. Parametric vs. permutations to analyze infant fNIRS data: analyzing the same dataset in three different ways. Biennial meeting of the Society for functional Near-Infrared Spectroscopy. Paris, France. October 2018.

Parametric vs permutation tests to analyze newborns fNIRS data: Analyzing the same dataset in three different ways

Cécile Issard[1] & Judit Gervain[1,2]

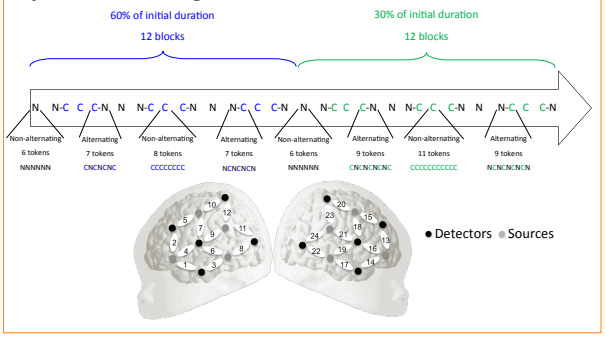
1: Laboratoire Psychologie de la Perception, Université Paris Descartes, Paris, France. (cecile.issard@etu.parisdescartes.fr)

2: Laboratoire Psychologie de la Perception, UMR 8242 Centre National de la Recherche Scientifique, Paris, France

Motivation

- GLM or ANOVAs/t-tests on mean amplitude in multiple channels: increase in Type I error →need to be corrected for multiple comparisons
- Bonferroni and FDR (Benjamini & Yekutieli, 2001) corrections may reduce statistical power
- Non-parametric permutation tests deal with the Multiple Comparisons Problem (MCP) and have only minimal assumptions (Holmes et al., 1996)
- Adaptation to find spatio-temporal clusters in MEEG data (Maris & Oostenveld, 2007) used in NIRS (Abboub et al., 2016; Edwards et al., 2016; Ferry et al., 2015).
- HRF has a precise shape and duration →time is not a relevant dimension to analyze statistically in NIRS.

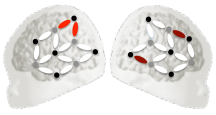
Experimental design



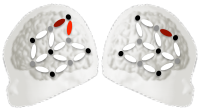
Parametric analysis

(t-tests)…

30% of initial duration vs. baseline



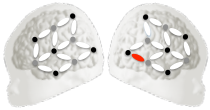
Normal duration vs. 30%



60% vs. 30% of initial duration

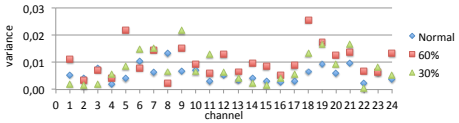
Corrected (FDR)

30% of initial duration vs. baseline

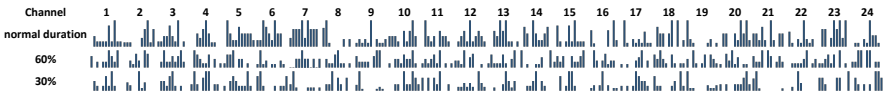


Assumptions ?

Homoscedasticity



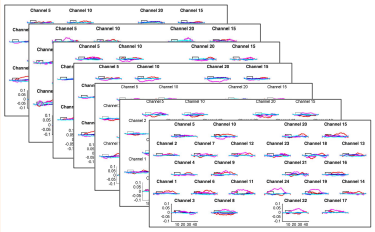
Normality



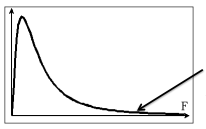
Ditributions of OxyHb concentration grouped by intervals of 1/10 of range

Single-threshold permutation test

Method



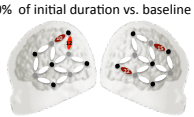
Fmax=3,25
Fmax=5,68
Fmax=1,27
Fmax=12,85
Fmax=0,12
Fmax=2,4
Fmax=3,12
.
.
.



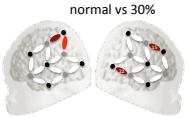
Pvalue : Proportion of Fmax higher than the actual F

Results

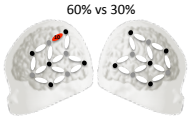
30% of initial duration vs. baseline



normal vs 30%



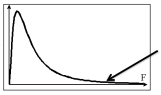
60% vs 30%



Cluster-based permutation test

Methods

- Select channels with F/t-value above a predefined threshold
- Sum up the adjacent F/t-values within each cluster: cluster statistic.
- Repeat n times (typically 1000) shuffling the labels.
- Note the cluster statistics you obtain under random permutations



P value : Proportion of clusters bigger than the actual ones

Results

Observed clusters :
- Channels 10 and 12
- Channel 15
- Channel 22

Permutations clusters :
- Critical Size (p=0,05) = 5 ch.

No significant cluster

Summary & Discussion

- Infants' (untransformed) NIRS data violate parametric tests assumptions within channels.
- FDR is too conservative for infant studies that have a small number of trials and a lower S/N ratio than adults studies.
- Single-threshold permutation tests are a powerful alternative that deals with the multiple comparison problem.
- Cluster-based permutation tests provide a data-driven alternative to manually-defined ROIs, but the spatial resolution of NIRS in newborns makes difficult to yield significant clusters.
- Recommendation: Single-threshold permutation test at channel resolution and ANOVA with ROIs defined by the channel-level permutation test and transformed data.

References

Abboub, N., et al. (2016). *Brain and Language 162* 46-59.

Benjamini, Y. & Yekutieli, D. (2001). *The Annals of Statistics 29*(4) 1165-1188.

Edwards, L. A., et al. (2016). *Developmental Science* 19(5), 757–769.

Ferry, A. L., et al. (2015). *Developmental Science 19*(3) 488–503.

Holmes, A. P.et al. (1996). *Journal of Cerebral Blood Flow and Metabolism 16*(1) 7-22.

Maris, E., Oostenveld, R. (2007). *Journal of Neuroscience Methods 164*(1) 177-190.

Title: Dealing with acoustical variability in speech at birth

Abstract: Speech probably represents the most important natural sound for humans. As speech sounds are variable, responding preferentially to speech sounds implies responding to the range of acoustical parameters that they can take. Indeed, we all have a different voice, we speak with different melodies, and we might have a foreign accent, but those who listen to us still all perceive the same words and phrases. This implies that we have extracted invariant linguistic units from variable sounds. Similarly, infants learn their native language from various speakers who speak with different speech rates and voice qualities from moment to moment. This means that, from the beginning of their life, humans are able to directly form invariant objects from the raw, variable sound. This implies that the auditory code should be flexible towards the broad range parameters than speech sounds can take, consistent with the idea that the higher stations of the auditory system are sensitive to the presence of abstract auditory entities (in our case speech), rather than absolute spectro-temporal parameters. Therefore a key question is how humans manage to extract these invariant representations of speech sounds from the beginning of their lives. The present thesis aims to uncover how these invariant representations of speech are built in human newborns by measuring newborns' hemodynamic and electrophysiological responses to natural speech, and temporally or spectrally modified speech.

In a first experiment, we presented normal speech as well as moderately (60% of initial duration) or highly time-compressed (30% of its initial duration) speech in the participants' native language (French). We recorded the hemodynamic response to these stimuli over the frontal, temporal and parietal cortices using NIRS. The results show no difference between normal and 60%-compressed speech, but differential responses between normal and 30%-compressed speech as well as between 60%- and 30%-compressed speech in a set of frontal, temporal, and temporo-parietal regions, similarly to the adult brain. This provides evidence that the newborn brain responds to speech in a stable manner over a range of time-scales that is similar to previous findings in adults. In a second set of experiments, we asked whether this ability relies on prenatal experience with the native language's rhythmic structure. We replicated the same experiment in two unfamiliar languages, one that is rhythmically similar to the native language (Spanish), and one that is rhythmically different (English). In English, only 30%-compressed speech evoked significant responses in a temporo-parietal region also activated for French, but the exact pattern of activations was different from those for French. This confirms that 30%-compressed speech is processed differently than normal and 60%-compressed speech. This also shows that prenatal experience shapes speech processing at birth. In particular, prenatal experience with the prosodic or phonological structure of the language might help infants encode speech in a stable way by providing auditory landmarks in the signal.

To conclude, the results presented in this thesis support the idea that speech is encoded as an abstract auditory object from the first stages of auditory processing. This auditory code is further modulated by higher level linguistic processing, integrating knowledge of the listener's native language. This knowledge is likely acquired from intra-uterine life, enabling a stable encoding of speech, adapted to the listener's linguistic environment from birth.

Keywords: Near-infrared spectroscopy, electro-encephalography, newborn, speech perception, acoustical variability

Résumé : La parole représente probablement le son naturel le plus important pour l'homme. Les sons de parole étant variables, réagir préférentiellement aux sons de parole implique de répondre à la gamme de paramètres acoustiques qu'ils peuvent prendre. En effet, nous avons tous une voix différente, nous parlons avec des mélodies différentes et nous avons peut-être un accent étranger, mais ceux qui nous écoutent perçoivent toujours les mêmes mots et les mêmes phrases. Cela implique que nous avons extrait des unités linguistiques invariantes à partir de sons variables. De même, les enfants apprennent leur langue maternelle à partir de différents locuteurs qui parlent avec des rythmes différents et des intonations différentes d'un moment à l'autre. Cela signifie que, dès le début de leur vie, les humains sont capables de former directement des objets invariants à partir du son variable. Pour cela le code neural doit être flexible envers les paramètres acoustiques que les sons de la parole peuvent prendre, conformément à l'idée que les étages supérieurs du système auditif sont sensibles à la présence d'entités auditives abstraites (ici la parole), plutôt qu'à des paramètres spectro-temporels absolus. Par conséquent, une question clé est de savoir comment les humains parviennent à extraire ces représentations invariantes des sons de parole dès le début de leur vie. Cette thèse vise à découvrir comment ces représentations invariantes de la parole sont construites chez le nouveau-né humain en mesurant les réponses hémodynamiques et électrophysiologiques du nouveau-né à la parole naturelle et à la parole modifiée temporellement ou spectralement.

Lors d'une première expérience, nous avons présenté de la parole normale ainsi que de la parole modérément compressée (60% de la durée initiale) ou fortement compressé dans le temps (30% de sa durée initiale) dans la langue maternelle des participants. Nous avons enregistré la réponse hémodynamique à ces stimuli sur les cortex frontal, temporal et pariétal à l'aide de NIRS. Les résultats ne montrent aucune différence entre la parole normale et la parole compressée à 60 %, mais des réponses différentielles entre la parole normale et la parole compressée à 30 % ainsi qu'entre la parole compressée à 60 % et la parole compressée à 30 % dans un ensemble de régions frontales, temporales, et temporo-pariétales, de la même manière que le cerveau adulte. Ceci montre que le cerveau du nouveau-né répond à la parole de manière stable sur une gamme d'échelles de temps similaire à celle observée précédemment chez l'adulte. Dans une deuxième série d'expériences, nous nous sommes demandé si cette capacité s'appuie sur l'expérience prénatale de la structure rythmique de la langue maternelle. Nous avons reproduit la même expérience dans deux langues inconnues, l'une rythmiquement similaire à la langue maternelle (l'espagnol) et l'autre rythmiquement différente (anglais). En anglais, seule la parole compressée à 30% évoque des réponses significatives dans une région temporo-pariétale également activée pour le français, mais le schéma exact d'activations est différent de celui du français. Cela confirme que la parole compressée à 30% est traitée différemment de la parole normale et de la parole compressée à 60%. Cela montre également que l'expérience prénatale façonne le traitement de la parole à la naissance. En particulier, une expérience prénatale de la structure prosodique ou phonologique de la langue pourrait aider les nourrissons à coder la parole de manière stable en fournissant des repères auditifs dans le signal.

En conclusion, les résultats présentés dans cette thèse soutiennet l'idée que la parole est codée comme un objet auditif abstrait dès les premières étapes du traitement auditif. Ce code auditif est en outre modulé par un traitement linguistique de plus haut niveau, intégrant la connaissance de la langue maternelle de l'auditeur. Ces connaissances sont probablement acquises à partir de la vie intra-utérine, ce qui permet un codage stable

de la parole, adapté à l'environnement linguistique de l'auditeur dès la naissance.

Mots-clés (français) : Spectroscopie de proche infrarouge, électro-encéphalographie, nouveau-né, perception de la parole, variabilité acoustique